

# Panel Methods (5): Modern DID

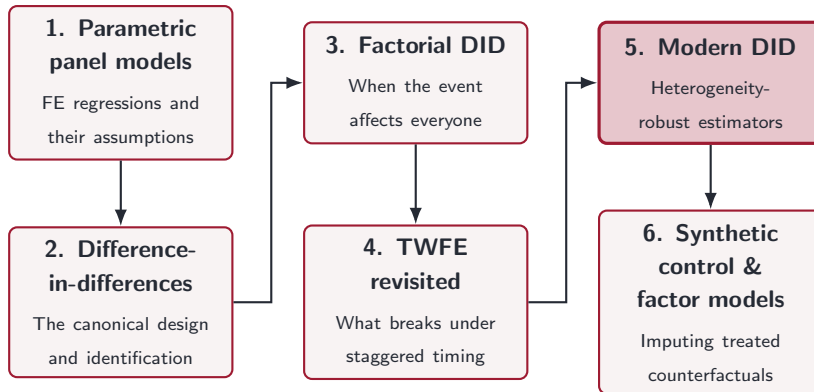
Lectures on Causal Panel Analysis

---

Yiqing Xu  
Stanford University

September 2026

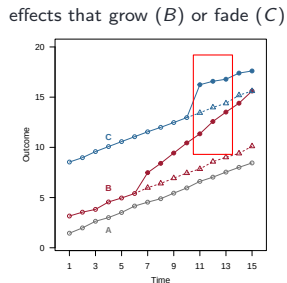
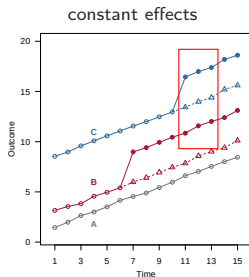
Feel free to use or adapt the slides with attribution (CC BY 4.0).



- Lecture 4 diagnosed the problem: TWFE on staggered panels mixes clean and contaminated comparisons, and parallel trends failures are common
- This lecture is about the remedies:
  - A tour of the **HTE-robust estimators**: what each one compares, and when each applies
  - The **diagnostics** that come with the imputation approach
  - Three worked examples, and what we learned from **reanalyzing 49 published studies** with these tools
- Lecture 6 introduces low-rank methods, synthetic control and factor models, that can address some parallel trends violations

## Review: What Goes Wrong under Heterogeneous Effects

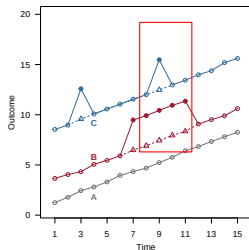
- Three units:  $A$  never treated,  $B$  treated from period 7,  $C$  from period 11; solid = observed, dashed = the untreated trajectory  $Y(0)$
- **Result** (Goodman-Bacon 2021 and others): with heterogeneous effects, TWFE need not recover any convex combination of the individual effects, even when parallel trends holds
- **Intuition**: in the boxed periods,  $B$ 's treated observations serve as the comparison for  $C$ , the “forbidden comparison”; harmless while  $B$ 's effect is constant, biased once it keeps growing
- The estimators in this lecture avoid such comparisons by construction



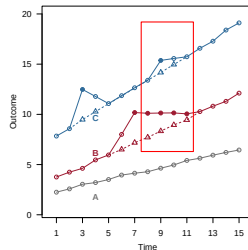
## ...and It Compounds with Reversal and Carryover

- Most political science panels have **treatment reversal**: here  $B$  is treated in periods 7–11 and then exits, and  $C$  switches on twice
- In the boxed window,  $B$  and  $C$  serve as comparisons for each other while both are treated
- Effects that build up and **linger after exit** (carryover) make it worse: the post-exit periods are no longer clean comparisons either
- Add parallel trends violations and anticipation, and the single-coefficient TWFE regression stands little chance; we need estimators that say **which cells compare with which**

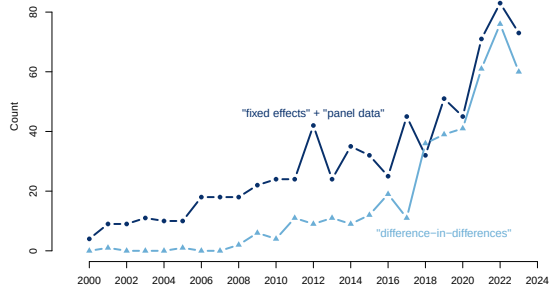
effects switch off with the treatment



effects build up and linger after exit



# The Menu Has Grown



- In our review of top political science journals, 77% of panel papers use FE models, mostly TWFE
- The econometric critiques produced many new estimators in a short span
- This lecture: a map of that menu, organized by which cells each estimator compares

## Preview: What the Reanalyses Will Show

### Common practice

- Fixed effects models in 77% of panel papers, TWFE in 58%
- Cluster-robust SEs in 98% of the reanalyzed studies; a cluster bootstrap in 16%
- Some plot of the raw data in 65%



### Do the results hold up? Yes and no

- Yes: HTE-robust estimators **rarely** flip signs
- No: parallel trends violations are still common
- No: power runs out once the estimator is HTE-robust
- No: few findings survive a **mild** sensitivity analysis

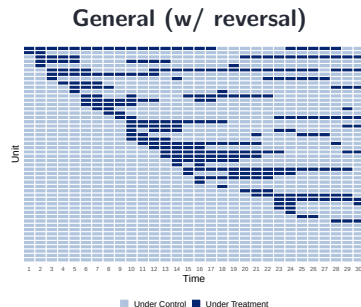
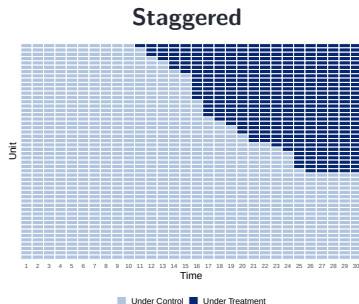
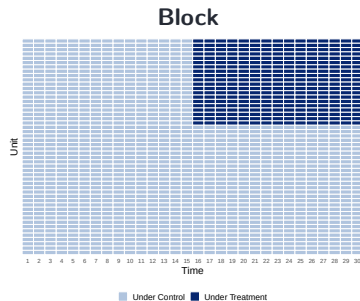
Strong empirical support for fewer than a third of the findings



### Takeaways

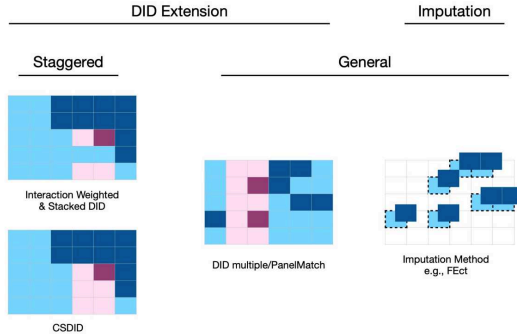
- Parallel trends, and the research design behind it, is the first-order issue
- The concern over heterogeneous effects is valid but second-order
- Validation is the key: an event-study plot is the minimum; sensitivity analysis helps
- “Robust” DID needs **a strong design** and **a lot of power**

# Three Data Settings



- Which estimators are available depends on the setting; treatment reversal (present in the majority of political science studies) rules out the staggered-only tools

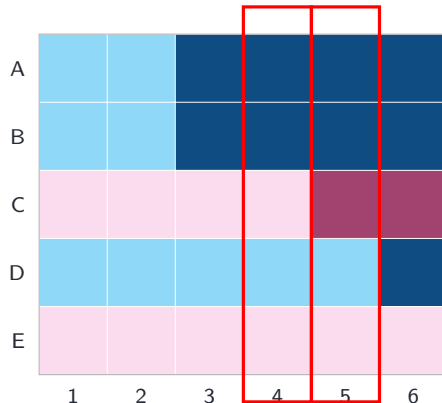
# The Organizing Principle



- **DID extensions** pick explicit  $2 \times 2$  comparisons; **imputation methods** fit an outcome model on untreated cells only

## Sun and Abraham (2021): Interaction Weighted

- Comparison group: **never-treated**
- Estimate a cohort ATT (CATT) by  $2 \times 2$  DID for each cohort  $g$  and period since treatment  $l$
- $ATT = \text{average CATT, weighted by cohort size}$
- The same construction for every cohort; shown here: cohort  $C$  at  $l = 0$ , the red window pairing the reference period with the first treated period



- The building block behind these estimators is the **group-time ATT**:

$$ATT(g, t) = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid G_i = g], \quad t \geq g$$

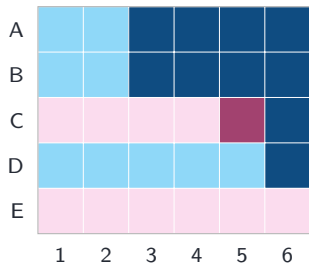
the effect for the cohort first treated at  $g$ , in calendar period  $t$

- One number per cohort-period pair is too much detail; the art is in the **aggregation**:
  - By relative time  $l = t - g$ : the **event-study** plot
  - By calendar time  $t$ : does the effect grow or fade over the sample period?
  - Overall: a weighted average across all treated cohort-periods
- Different papers need different aggregations; report the one your theory speaks to, not just the overall number

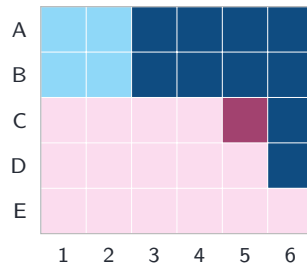
## Callaway and Sant'Anna (2021)

- Comparison group: **not-yet-treated** (in addition to never-treated)
- “Doubly robust” with covariates: combines outcome regression and propensity-score weighting

Sun and Abraham (2021)



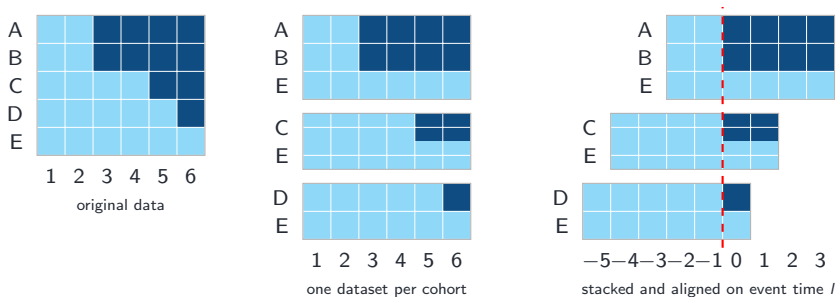
Callaway and Sant'Anna (2021)



- More comparison cells  $\Rightarrow$  more power, at the price of assuming parallel trends also holds for the not-yet-treated

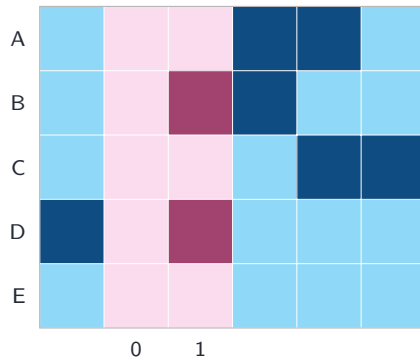
## Stacked DID (Cengiz et al. 2019)

- Build one dataset per cohort: the cohort's units plus a copy of the pure comparison group, the never-treated  $E$
- “Stack” the cohort datasets and re-index time relative to each cohort's onset, so that the switches line up
- One saturated regression on the stack, with unit and period effects specific to each dataset; similar to interaction weighting, but with disproportionate weights across cohorts

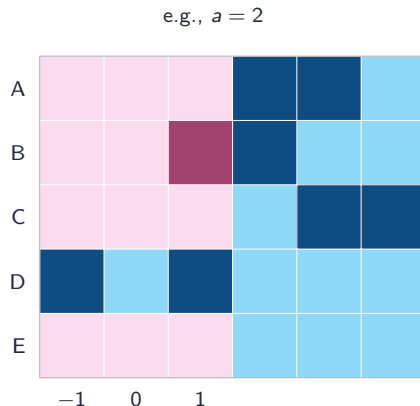


In practice a unit serves as a comparison for a cohort only while it stays untreated inside that cohort's event window (“clean controls”).

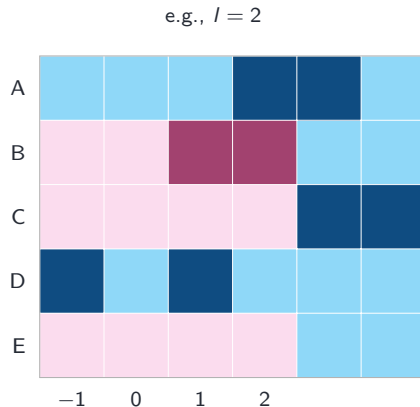
- No cohorts: one average effect **for switchers**, not the ATT; works with treatment reversal
- Switchers:  $(i, t)$  with  $D_{it} \neq D_{i,t-1}$ ; here *B* and *D* at period 3
- Stable group:  $(i, t)$  with  $D_{it} = D_{i,t-1}$ , sharing the switchers' status in the previous period; here *A*, *C*, *E*
- DID<sub>M</sub>: a DID from period 2 to 3, switchers against the stable group, estimates the **contemporaneous** effect at the period of the switch; leavers ( $1 \rightarrow 0$ ) are compared with stable treated units in the same way, with the sign reversed
- Same data as PanelMatch on the next page: DID<sub>M</sub> is PanelMatch with  $a = l = 1$  and no refinement



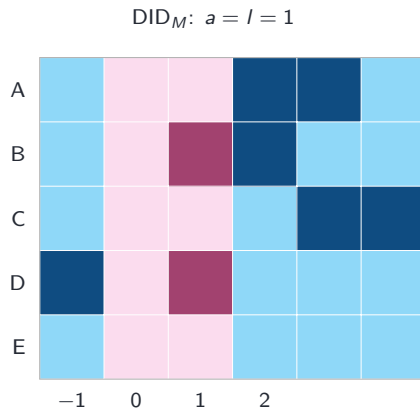
- General setting: match each treated  $(i, t)$  to units with the **same treatment history** over the previous  $a$  periods
- $a = 2$ : for  $B$  at period 3 the matched set is  $\{A, C, E\}$ , untreated in periods 1–2 like  $B$ ;  $D$  (treated in period 1) is **excluded**
- A DID from period 2 to 3 within the matched set: the effect at  $l = 1$
- Later periods  $l = 2, 3, \dots$  (up to reversal): units that switch inside the window drop out; here  $A$  leaves at  $l = 2$
- Refine the matched set on covariates  $X_{it}$
- $\text{DID}_M$ : the case  $a = l = 1$  without refinement; at period 3 the switchers  $B$  and  $D$  are matched to  $A, C, E$



- General setting: match each treated  $(i, t)$  to units with the **same treatment history** over the previous  $a$  periods
- $a = 2$ : for  $B$  at period 3 the matched set is  $\{A, C, E\}$ , untreated in periods 1–2 like  $B$ ;  $D$  (treated in period 1) is **excluded**
- A DID from period 2 to 3 within the matched set: the effect at  $l = 1$
- Later periods  $l = 2, 3, \dots$  (up to reversal): units that switch inside the window drop out; here  $A$  leaves at  $l = 2$
- Refine the matched set on covariates  $X_{it}$
- $\text{DID}_M$ : the case  $a = l = 1$  without refinement; at period 3 the switchers  $B$  and  $D$  are matched to  $A, C, E$

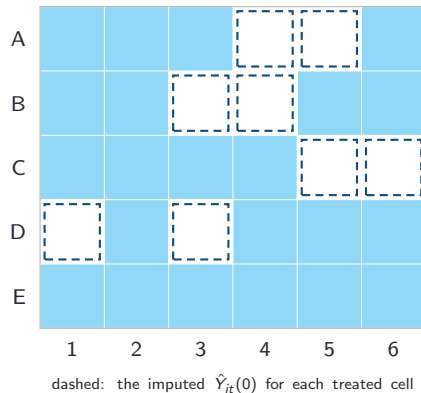


- General setting: match each treated  $(i, t)$  to units with the **same treatment history** over the previous  $a$  periods
- $a = 2$ : for  $B$  at period 3 the matched set is  $\{A, C, E\}$ , untreated in periods 1–2 like  $B$ ;  $D$  (treated in period 1) is **excluded**
- A DID from period 2 to 3 within the matched set: the effect at  $l = 1$
- Later periods  $l = 2, 3, \dots$  (up to reversal): units that switch inside the window drop out; here  $A$  leaves at  $l = 2$
- Refine the matched set on covariates  $X_{it}$
- $\text{DID}_M$ : the case  $a = l = 1$  without refinement; at period 3 the switchers  $B$  and  $D$  are matched to  $A, C, E$



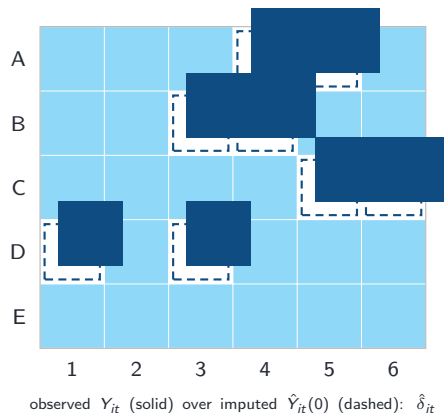
## Imputation Methods (Borusyak et al. 2024; Liu, Wang, and Xu 2024)

- Also for the general setting, with a different logic: model the untreated outcome instead of choosing comparisons
- Set the treated cells aside and fit a model for  $Y_{it}(0)$  on the **untreated cells only**; with two-way fixed effects, this is the fixed effects counterfactual estimator (FEct)
- Impute  $\hat{Y}_{it}(0)$  for every treated cell
- One estimate per treated cell:  $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$
- Summarize the  $\hat{\delta}_{it}$  however the question demands: overall ATT, dynamic effects, subgroups; no forbidden comparisons by construction; efficient under homoskedasticity (Borusyak et al. 2024)



## Imputation Methods (Borusyak et al. 2024; Liu, Wang, and Xu 2024)

- Also for the general setting, with a different logic: model the untreated outcome instead of choosing comparisons
- Set the treated cells aside and fit a model for  $Y_{it}(0)$  on the **untreated cells only**; with two-way fixed effects, this is the fixed effects counterfactual estimator (FEct)
- Impute  $\hat{Y}_{it}(0)$  for every treated cell
- One estimate per treated cell:  $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$
- Summarize the  $\hat{\delta}_{it}$  however the question demands: overall ATT, dynamic effects, subgroups; no forbidden comparisons by construction; efficient under homoskedasticity (Borusyak et al. 2024)



## Extended TWFE, or the Two-Way Mundlak Regression (Wooldridge 2025)

- Keep the regression, but let the effect differ by cohort and period: one coefficient  $\tau_{gt}$  for every cohort  $g$  in every post-treatment period  $t \geq g$ , instead of a single  $\tau$

$$Y_{it} = \alpha_i + \xi_t + \sum_g \sum_{t' \geq g} \tau_{gt'} \mathbf{1}\{G_i = g\} \mathbf{1}\{t = t'\} + \varepsilon_{it}$$

- “Mundlak”: in a balanced panel, replacing the unit effects by cohort dummies (the unit-level means of the regressors) gives numerically the same  $\hat{\tau}_{gt}$ , so the whole thing is one pooled regression
- Which cells compare: with no pre-period dummies, every pre-treatment cell stays in the untreated pool that pins down  $\alpha_i$  and  $\xi_t$ , exactly as in imputation (IW and CSDID instead compare against one reference period); in a balanced panel without covariates the  $\hat{\tau}_{gt}$  equal the imputation estimates
- Why it has caught on: one regression command with regression standard errors; covariates enter interacted with cohort and period; the same device extends to **nonlinear models** (Poisson for counts, logit for binary outcomes; Wooldridge 2023), where subtracting an imputed  $\hat{Y}(0)$  makes no sense
- Limits: staggered adoption only (a cohort is a first-treatment date, so no reversal); many coefficients, to be aggregated as with CSDID; parallel trends still does the identifying work. Software: `etwfe` (R), `jwddid` (Stata)

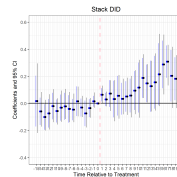
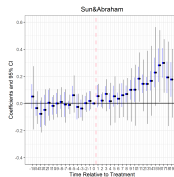
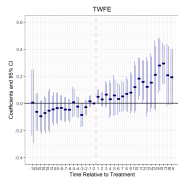
# The Estimators, Side by Side

|                    | DID extensions                 |                              | Imputation methods                                  |                                             |
|--------------------|--------------------------------|------------------------------|-----------------------------------------------------|---------------------------------------------|
|                    | (2 × 2 DID as building blocks) |                              | (outcome model w/ FE)                               |                                             |
| Setting            | Staggered                      | General                      | Staggered                                           | General                                     |
| Estimand           | ATT                            | ATT for switchers            | ATT by cohort-period                                | ATT                                         |
| Estimator          | IW, CSDID, stacked             | PanelMatch, DID <sub>M</sub> | extended TWFE<br>(two-way Mundlak)                  | BJS, FE <sub>ct</sub>                       |
| Comparisons        | never/not-yet-treated          | matched set                  | imputed counterfactual,<br>one saturated regression | imputed counterfactual                      |
| Pre-treatment info | one reference period           | the $a$ previous periods     | all untreated cells                                 | all untreated cells                         |
| Key assumption     | parallel trends                | parallel trends              | zero conditional mean<br>or parallel trends         | zero conditional mean<br>or parallel trends |

## Which Estimator Should I Use?

- **Two-period or block assignment:** canonical DID is fine; TWFE equals DID here
- **Staggered, no reversal:** the full menu applies; IW/CSDID if cohort effects are the object, imputation for flexibility and diagnostics; extended TWFE when the outcome calls for a nonlinear model
- **Treatment reversal** (most political science panels): PanelMatch,  $DID_M$ , or imputation; the staggered-only tools do not apply
- **Few treated units or one treated unit:** move to synthetic control territory (Lecture 6)
- In well-powered data the choice rarely drives the answer; choose for **diagnostics and transparency**, then show robustness across estimators

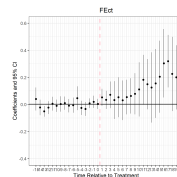
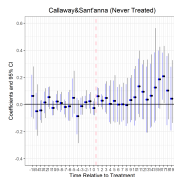
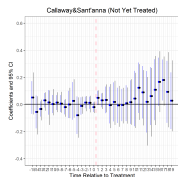
# One Staggered Application, Five Estimators



TWFE event study

Sun and Abraham

stacked DID



CSDID (not-yet-treated)

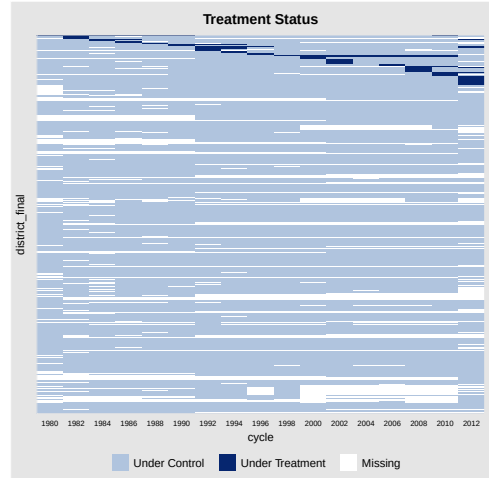
CSDID (never-treated)

imputation (FEct)

- Bischof and Wagner (2019): the estimators tell one story; when parallel trends looks plausible, agreement is the norm

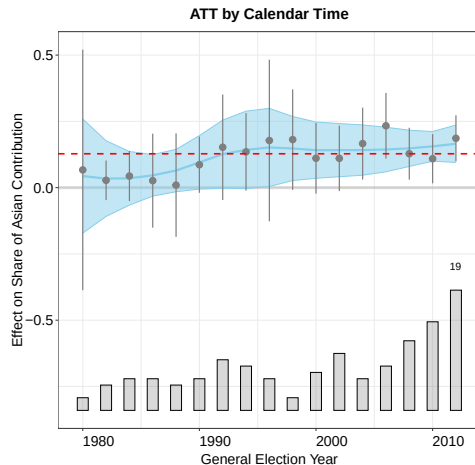
## Example 1: Coethnic Mobilization (Grumbach and Sahn 2020)

- Do ethnic minority candidates mobilize coethnic donors?
  - Unit of analysis: 498 US congressional districts, 1980–2012
  - Treatment: Asian (vs. White, Black, and Hispanic) candidates in general elections
  - Outcome: percentage of Asian donations
- The treatment matrix is unbalanced with missing data; the imputation approach handles it without reweighting gymnastics

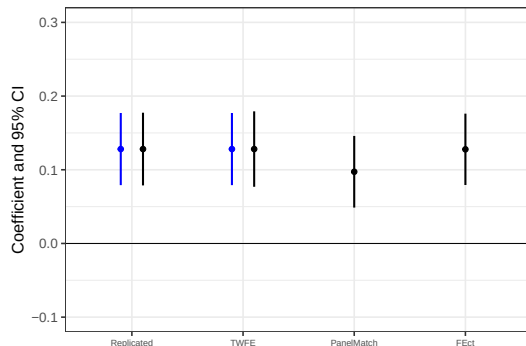


## Effects Are Heterogeneous

- Effects grow over calendar time: exactly the heterogeneity that breaks the single- $\tau$  TWFE regression
- Do the HTE-robust estimators change the answer here?



## Do the Estimators Agree in Practice?

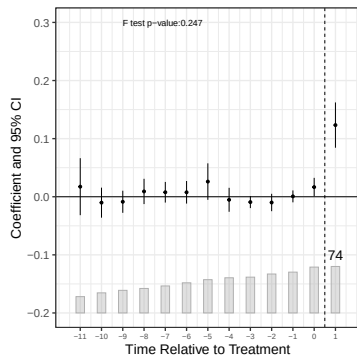


- Grumbach and Sahn (2020): TWFE, PanelMatch, and FEct give similar answers; here the design is strong and HTE is second-order

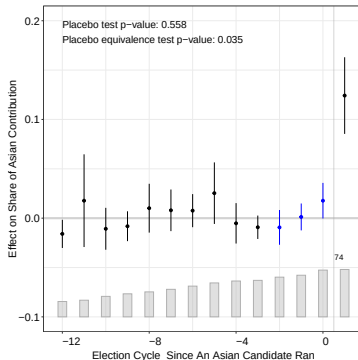
The imputation approach turns identification checks into estimable quantities:

- **Pre-trend  $F$  test:** are the average  $\hat{\delta}_{it}$  in pre-treatment periods jointly zero?
- **Placebo test:** hide the last few pre-treatment periods from estimation, then check the “effect” in the hidden periods
- **Carryover test:** hide the first few **post-exit** periods, then check for lingering effects after treatment ends
- **Equivalence (TOST) tests:** instead of failing to reject zero, reject pre-trends larger than a tolerance; more informative in well-powered settings

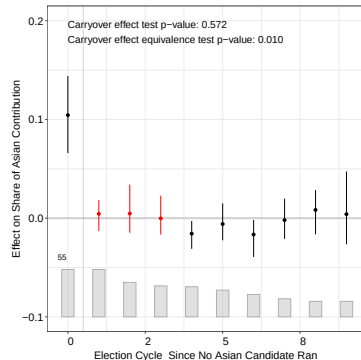
# The Suite Applied to Example 1 (Grumbach and Sahn 2020)



F test for no pre-trend



placebo test



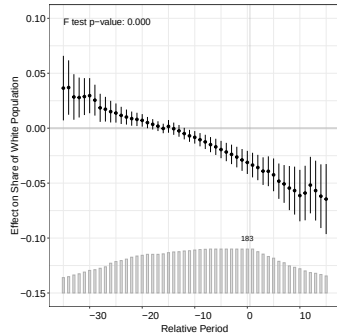
carryover test

- None of the three rejects its null here; with 74 treated observations, power is the caveat to keep in mind

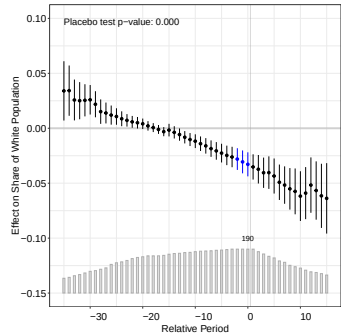
## Example 2: Lawsuits against Land-Use Restrictions (APSR 2020)

- Treatment: a Fair Housing Act lawsuit against a city's land-use restrictions; outcome: the share of the city's residents who are non-Hispanic white (California cities, 1970–2010, staggered)
- The imputation event-study plot slopes downward **before** treatment: the pre-trend  $F$  test and the placebo test both reject ( $p < 0.001$ )

dynamic effects, with the pre-trend  $F$  test

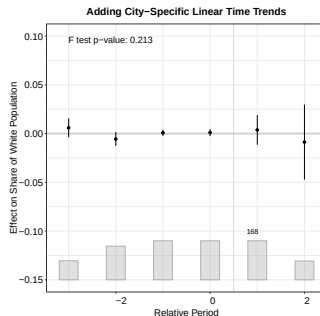
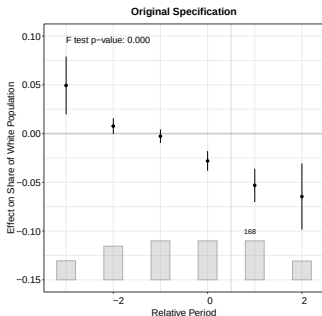


placebo test (blue: the hidden periods)

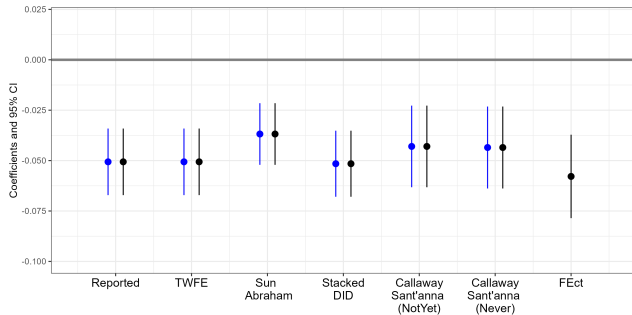


## Example 2: Adding City-Specific Linear Time Trends

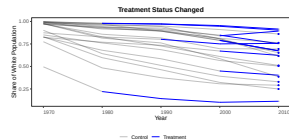
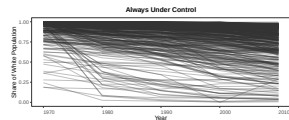
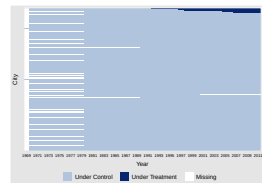
- The demographic series are mostly interpolated between Census years; dropping the interpolated years changes little
- Adding **city-specific linear time trends** removes the pre-trend ( $F$  test  $p = 0.21$ ) and the negative effect with it, on the same  $-3$  to  $+2$  window
- The downward slope predates the lawsuits; once the trends absorb it, nothing is left



## Example 2: The Estimators Still Broadly Agree

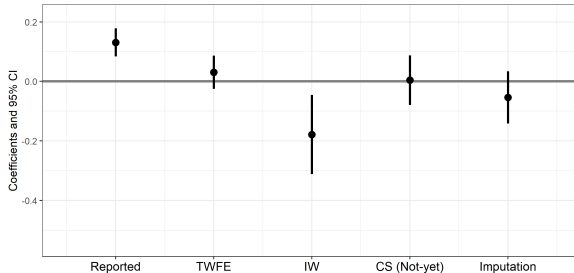


- Seven estimators, one answer, about  $-0.05$ : they all inherit the **same** violated assumption
- Agreement across estimators is not robustness; the raw-data plots and the two tests catch what comparing estimators cannot

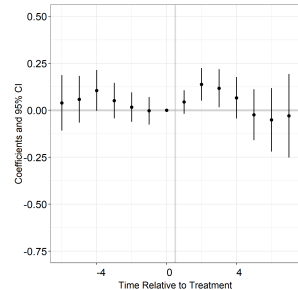


### Example 3: Updating Cadastral Maps and Property Tax Revenue (JOP 2021)

- The authors ask whether updating the cadastral map, the register of properties, raises property tax revenue in Brazilian municipalities (2004–2015, staggered adoption)
- Full-sample reanalysis: the estimators **disagree**; the reported estimate is positive, TWFE and CSDID sit near zero, interaction weighting is negative
- The paper's event-study specification, replicated on its subsample, shows the positive effect

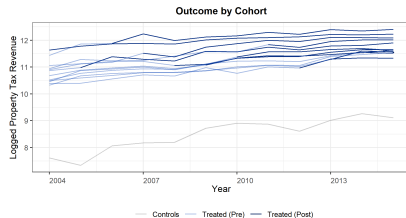
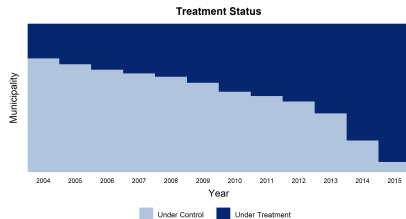


replicated on the full sample

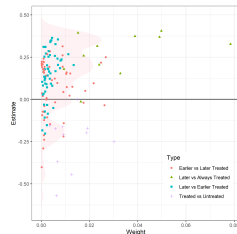


the paper's specification on its subsample

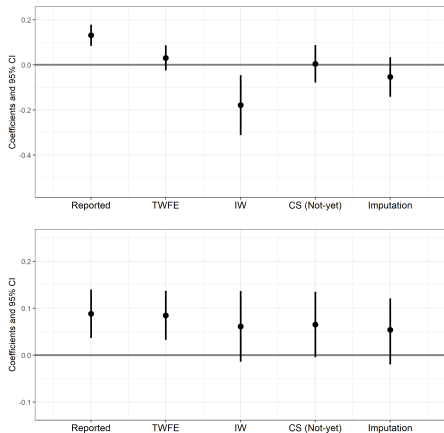
## Example 3: Look Deeper



- A large **always-treated** block sits on top of the staggered adopters; only a handful of municipalities are never treated
- Those never-treated municipalities collect far less revenue than every treated cohort: a weak comparison group
- Goodman-Bacon's decomposition of the TWFE estimate: comparisons among treated cohorts sit around  $+0.25$ , comparisons against the never-treated around  $-0.4$ ; the answer depends on **which comparisons carry weight**

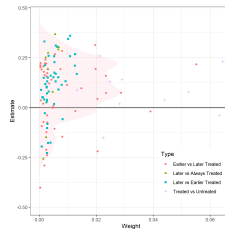


## Example 3: A Trimmed Subsample



top: full sample; bottom: trimmed subsample

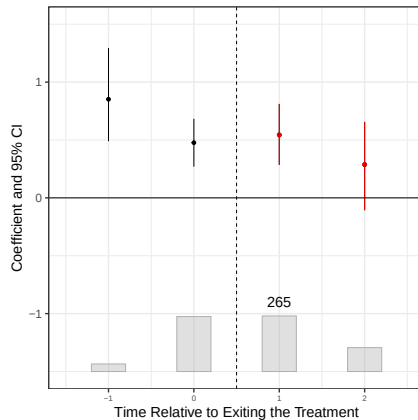
- Drop the always-treated block and the far-away never-treated units, without looking at the outcome
- On the subsample the estimators **converge**, all between 0.05 and 0.09 with overlapping intervals; the comparisons now form one cloud
- The disagreement was a symptom of a weak comparison set, not of heterogeneity; trimming to “compare like with like” is design work (Imbens and Rubin 2015)



- **Example 1, coethnic mobilization:** strong design; heterogeneity matters marginally; the estimators, TWFE included, broadly agree
- **Example 2, land-use lawsuits:** clear signs of a parallel trends violation; heterogeneity is second-order; agreement among estimators does not mean robustness; plotting the raw data and the two tests spot the problem
- **Example 3, cadastral maps:** when estimators disagree, suspect the comparison set before the estimators; design-phase trimming, done without the outcome, restores agreement

## Carryover Effects Are Common

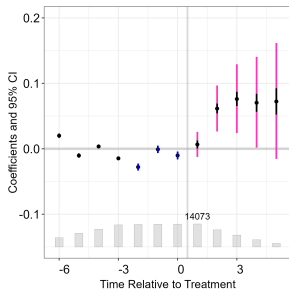
- With treatment reversal, effects that linger after exit violate the no-carryover assumption; we find this in roughly a third of the reanalyzed studies
- When carryover is plausible, estimate effects relative to treatment **entry** only, or model the exit explicitly



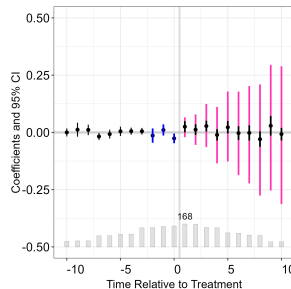
Distelhorst and Locke (2018): rejects no carryover — unlike Example 1  
(Grumbach and Sahn 2020)

## Sensitivity Analysis: Relaxing Parallel Trends

- Rambachan and Roth (2023): allow post-treatment confounding up to  $M$  times the largest change between neighboring placebo periods;  $M = 0$  means parallel trends holds exactly



Hall and Yoder (2022)

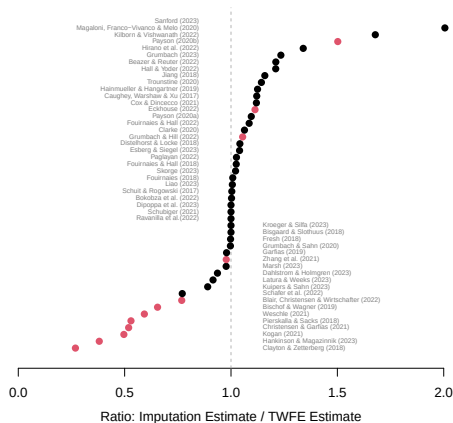


Caughey, Warshaw, and Xu (2017)

- The question shifts from “does the effect exist” to “how much confounding would it take to erase it”

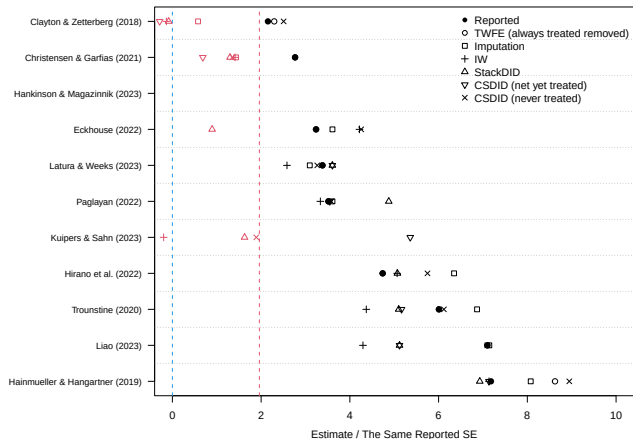
- We replicated a main result from top political science journals, 2017–2023: 102 papers use linear panel models; 64 have a “proper” TWFE design; 49 are replicable
- Each study re-estimated with the full menu: IW, CSDID, stacked, PanelMatch/DID<sub>M</sub>, and imputation, plus the diagnostic suite and sensitivity analysis
- Settings in the sample: 55% general (with reversal), 27% staggered, 12% multi-period block, 6% classic  $2 \times 2$ : the staggered-only estimators do not even apply to most studies

## Finding 1: Estimates Rarely Flip, but Variability Exists



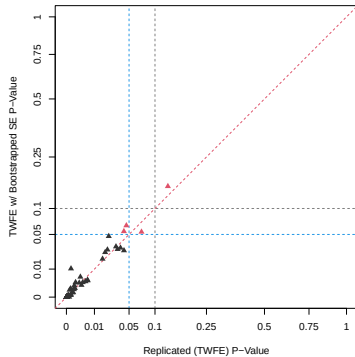
- The ratio of imputation to TWFE estimates centers on 1.0 (mean 1.02, median 1.00); large deviations concentrate in sparse-data cases

## Finding 1 (continued): The Staggered Cases

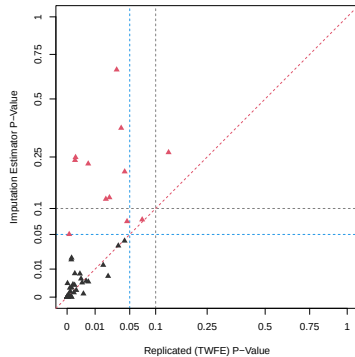


- Across the staggered studies, the five HTE-robust estimators and TWFE broadly agree on magnitudes

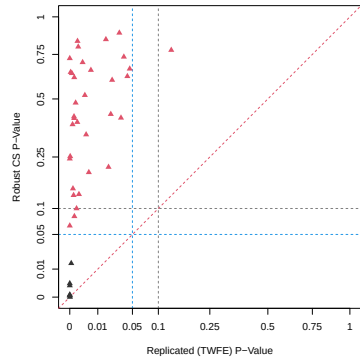
## Finding 2: “Robust” DID Is Power Hungry



bootstrapped SEs (91% survive)



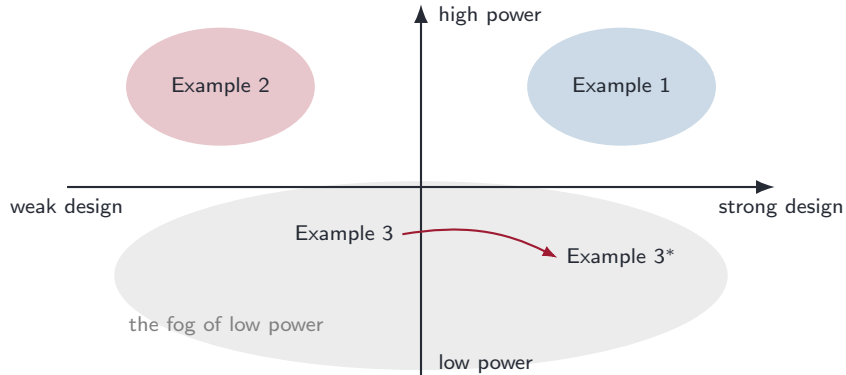
+ relaxing constant effects (75%)



+ relaxing parallel trends (23%)

- Each layer of robustness costs power; fewer than a third of the findings survive mild sensitivity analysis

# Big Picture: Design and Power



- Estimator choice barely moves a study along either axis; **design** and **power** do. Trimming moved Example 3 to the right; moving up takes more data

# The Paradox of Committing to TWFE

- Clear signs of parallel trends violations in roughly a quarter of the studies; in over half, we **cannot tell**, because there are too few pre-periods or too little power
- With enough power, you can afford an HTE-robust estimator, so use one
- Without enough power, you cannot **validate** the assumptions TWFE needs anyway
- Either way, blind TWFE is hard to defend; the binding constraints are **design and power**, not the estimator

# Recommendations

|                | Do                                                                  | Don't                                               |
|----------------|---------------------------------------------------------------------|-----------------------------------------------------|
| Design         | Start from a research design; check for feedback from past outcomes | Start by running regressions on whatever data exist |
| Raw data       | Plot treatment status and outcome trajectories first                | Run regressions without looking at the data         |
| Estimators     | Use HTE-robust estimators; always plot dynamic effects              | Choose models on belief; report coefficients only   |
| Diagnostics    | Run visual and statistical tests of the assumptions                 | Skip visualization and diagnostics                  |
| Clustering     | Cluster at the level of treatment assignment or higher              | Cluster below the assignment level                  |
| Small clusters | Cluster-bootstrap when clusters $\lesssim 50$                       | Rely on asymptotic SEs                              |

- The HTE-robust menu is organized by comparisons: DID extensions choose their  $2 \times 2$ s explicitly; imputation methods model  $Y(0)$  from untreated cells
- In 49 reanalyses, estimator choice rarely flips results; parallel trends violations and **low power** are the first-order problems
- Run the diagnostic suite (pre-trend, placebo, carryover, equivalence) and report a sensitivity analysis, not just a point estimate

- `fect` (R & Stata): imputation estimators with the full diagnostic suite
- `panelView` (R & Stata): treatment status and outcome visualization
- `did`, `PanelMatch`, `DIDmultiplegt`: the DID extensions; `didimputation`: BJS imputation; `etwfe` (R) and `jwddid` (Stata): extended TWFE
- Tutorial: <https://yiqingxu.org/packages/fect/>

# References

- Bischof, Daniel, and Markus Wagner. 2019. "Do Voters Polarize When Radical Parties Enter Parliament?" *American Journal of Political Science* 63 (4): 888–904.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91 (6): 3253–85.
- Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.
- Caughey, Devin, Christopher Warshaw, and Yiqing Xu. 2017. "Incremental Democracy: The Policy Effects of Partisan Control of State Government." *Journal of Politics* 79 (4): 1342–58.
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *Quarterly Journal of Economics* 134 (3): 1405–54.
- de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–96.
- Chiu, Albert, Xingchen Lan, Ziyi Liu, and Yiqing Xu. 2026. "Causal Panel Analysis under Parallel Trends: Lessons from a Large Reanalysis Study." *American Political Science Review* 120 (1): 245–66.
- Distelhorst, Greg, and Richard M. Locke. 2018. "Does Compliance Pay? Social Standards and Firm-Level Trade." *American Journal of Political Science* 62 (3): 695–711.
- Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.
- Grumbach, Jacob M., and Alexander Sahn. 2020. "Race and Representation in Campaign Finance." *American Political Science Review* 114 (1): 206–21.

## References (cont.)

- Hall, Andrew B., and Jesse Yoder. 2022. "Does Homeownership Influence Political Behavior? Evidence from Administrative Data." *Journal of Politics* 84 (1): 351–66.
- Imai, Kosuke, In Song Kim, and Erik H. Wang. 2023. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 67 (3): 587–605.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68 (1): 160–76.
- Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* 90 (5): 2555–91.
- Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99.
- Wooldridge, Jeffrey M. 2023. "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data." *Econometrics Journal* 26 (3): C31–C66.
- Wooldridge, Jeffrey M. 2025. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." *Empirical Economics* 69 (5): 2545–87. Working paper 2021.