Part D: Panel Data Methods

D3: Staggered-Adoption Difference-in-Differences

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

D3 outline

1. Staggered adoption: Setting and estimands
2. Traditional estimators
3. What to do instead

Staggered adoption/rollout setting

	$i = Z$	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$					
$t = 2$					
$t = 3$					
$t = 4$					
$t = 5$					
$t = 6$					

- Assume binary treatment; i gets treated at $t = E_i$ and stays treated forever: $D_{it} = \mathbf{1}\left[t \geq E_i\right]$
  - "Cohort" $=$ units with the same E_i
- May or may not have never-treated units ($E_D = \infty$), always-treated units ($E_Z = 1$)
  - Assume away always-treated that will mostly be useless

An extra source of variation

- In addition to treated vs. never-treated comparisons, we can now compare treated vs. not-yet-treated
- Sometimes researchers explicitly drop the never-treated group to focus on more comparable units
  - E.g. a panel of mothers, where $E_i =$ year of birth of first child
  - May drop women without kids, as they are not expected to be on parallel trends
- Other times, researchers leverage both

Notation and assumptions

- Fixed sample $\Omega = \{it\}$ with untreated obs Ω_0 and treated obs Ω_1
  - Ω_0 includes never-treated and not-yet-treated *it* pairs
- Causal structure: $Y_{it}(D_{it})$ (no spillovers, no anticipation effects)
- Causal effects: $\tau_{it} = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0)\right]$, varying across i, t
- Parallel trends: $\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t$

A note on modeling dynamics

- We could have defined potential outcomes as $Y_{it}(\ldots, d_{t+2}, d_{t+1}, d_t, d_{t-1}, d_{t-2}, \ldots)$
- With absorbing treatment, same as $Y_{it}(e)$ where e is date first treated (or ∞)
  - No anticipation effects: $Y_{it}(e) = Y_{it}(e') \equiv Y_{it}(\infty)$ for $e, e' > t$
- We will focus on ATT-type estimands based on $Y_{it}(E_i) - Y_{it}(\infty)$

$\Rightarrow$ Can simplify potential outcomes: $Y_{it}(1) \equiv Y_{it}(E_i) = Y_{it}, \quad Y_{it}(0) \equiv Y_{it}(\infty)$

  - Interpret $\tau_{it} = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0)\right]$ as reflecting, in part, dynamic effects

Target estimands

- Linear target estimand: $\sum_{it \in \Omega_1} w_{it} \tau_{it}$ for w_{it} chosen by researcher...
- ATT: $\frac{1}{|\Omega_1|} \sum_{it \in \Omega_1} \tau_{it}$, i.e. $w_{it} = \frac{1}{|\Omega_1|}$ for all $it \in \Omega_1$

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				

- *Note:* this estimand is very sensitive to how many periods are observed when the effects exhibit dynamics

More estimands of interest

- ATT

- ATT $h \geq 0$ periods since treatment (typically fewer units for longer horizons)

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$	$h = 0$			
$t = 3$	$h = 1$	$h = 0$		
$t = 4$	$h = 2$	$h = 1$		
$t = 5$	$h = 3$	$h = 2$	$h = 0$	
$t = 6$	$h = 4$	$h = 3$	$h = 1$	

More estimands of interest

- ATT
- ATT $h \geq 0$ periods since treatment
- ATT $h \geq 0$ periods since treatment on a balanced set of units for different h

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$	$h = 0$			
$t = 3$	$h = 1$	$h = 0$		
$t = 4$	$h = 2$	$h = 1$		
$t = 5$	$h = 3$	$h = 2$	$h = 0$	
$t = 6$	$h = 4$	$h = 3$	$h = 1$	

More estimands of interest

- ATT
- ATT $h \geq 0$ periods since treatment
- ATT $h \geq 0$ periods since treatment on a balanced set of units for different h
- Difference of ATT across subgroups
- Size-weighted ATT; etc.

Outline

Conventional practice

By analogy with non-staggered DiD, it used to be common to estimate:

- Static TWFE specification — to get a single summary statistic of treatment effects:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau_{\text{static}} D_{it} + \varepsilon_{it}$$

- Event study (dynamic) specification:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{\substack{h=-a \\ h\neq -1}}^{b-1} \tau_h \mathbf{1}\left[t = E_i + h\right] + \tau_{b+} \mathbf{1}\left[t \geq E_i + b\right] + \varepsilon_{it}$$

  - Some dummies are often binned or dropped on the left and/or on the right
  - “Fully-dynamic” if $h = -1$ is the only omitted term
  - “Semi-dynamic” if include all $h \geq 0$ terms and nothing else

Static TWFE specification

- Does the static TWFE specification estimate the ATT?

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau_{\text{static}} D_{it} + \varepsilon_{it}$$

- de Chaisemartin and D'Haultfoeuille (AER 2020), Borusyak, Jaravel, and Spiess (REStud 2024) (BJS):
  - Yes if the effects are homogeneous across units and periods. Not otherwise!
  - Under PTA, estimand $\tau_{\text{static}} = \sum_{it \in \Omega_1} w_{it}^{\text{static}} \tau_{it}$ for some weights w_{it}^{static} that add up to one
  - But $w_{it}^{\text{static}} \neq \frac{1}{|\Omega_1|}$ and some can be negative due to "forbidden comparisons"...

Forbidden comparisons

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau_{\text{static}} D_{it} + \varepsilon_{it}$$

$\mathbb{E}[Y_{it}]$	$i = A$	$i = B$
$t = 1$	α_A	α_B
$t = 2$	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
$t = 3$	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

Here $\hat{\tau}_{\text{static}} = \left(Y_{A2} - Y_{B2} \right) - \frac{1}{2}\left(Y_{A1} - Y_{B1} \right) - \frac{1}{2}\left(Y_{A3} - Y_{B3} \right)$

- Future periods are used as reference periods for treated observations of early adopters
- Long-term effects for early adopters can get a negative weight: $\tau_{\text{static}} = \tau_{A2} + \frac{1}{2}\tau_{B3} - \frac{1}{2}\tau_{A3}$

Mechanics of negative weights

- By Frisch–Waugh–Lovell, $\hat{\tau}_{\text{static}}$ can be obtained from

$$Y_{it} = \tau_{\text{static}} \tilde{D}_{it} + \text{error}$$

  where $\tilde{D}_{it}$ are residuals from regressing $D_{it} = a_i + b_t + \text{error}$.

- Thus, $\hat{\tau}_{\text{static}} = \frac{\sum_{it} \tilde{D}_{it} Y_{it}}{\sum_{js} \tilde{D}_{js}^2}$. Weights $\frac{\tilde{D}_{it}}{\sum_{js} \tilde{D}_{js}^2}$ are easy to compute

- But they can be negative for some treated observations: where $\hat{a}_i$ is high (early adopters) and $\hat{b}_t$ is high (late periods if few never-treated units)
    - Angrist (1998) result does not apply to TWFE!

Characterizing negative weights

- You can compute w_{it}^{static} (by observation or group totals) and total negative weights
- Goodman-Bacon (2021) provides a decomposition of τ_{static} as convex weighted average of several types of comparisons (package *bacondecomp*):
    - Treated vs. never treated *(good)*
    - Early adopters vs. late adopters *(good)*
    - Late adopters vs. early adopters *(forbidden)*
    - Treated during the sample vs. always-treated *(forbidden)*
- *Note*: only useful if you plan to use the static TWFE specification

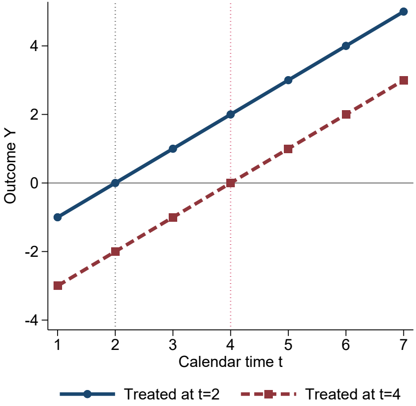
Under-identification of the fully-dynamic specification

- Fully-dynamic specification:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h \neq -1} \tau_h \mathbf{1}\left[t = E_i + h\right] + \varepsilon_{it}$$

- **Proposition** (BJS): Without never-treated units, the path $\{\tau_h\}_{h \neq -1}$ is not point identified in the fully-dynamic specification. Adding any linear trend to this path, $\{\tau_h + \kappa(h+1)\}$, fits the data equally well
  - This transformation corresponds to adding $\kappa\left(t - (E_i - 1)\right)$ to fitted values, which can be offset by changing unit and time FEs by $\kappa\left(E_i - 1\right)$ and $-\kappa t$ resp.

Under-identification of the fully-dynamic specification



(from BJS)

- Diff-in-diff doesn't work without some assumption of no anticipation effects!

Spurious identification of very long-run effects

- Semi-dynamic specification can be estimated:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h \geq 0} \tau_h \mathbf{1}\left[t = E_i + h\right] + \varepsilon_{it}$$

- **Proposition** (BJS): Without never-treated units and with heterogeneous effects, long-run effects ($h \geq \max_i E_i - \min_i E_i$) are not identified by PTA, while the semi-dynamic specification produces some (heroic/spurious) estimates

$\mathbb{E}\left[Y_{it}\right]$	$i = A$	$i = B$
$t = 1$		
$t = 2$	✓	
$t = 3$	✗	✗

Spurious identification of very long-run effects

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h=0}^{1} \tau_h \mathbf{1}\left[t = E_i + h\right] + \varepsilon_{it}$$

$\mathbb{E}\left[Y_{it}\right]$	$i = A$	$i = B$
$t = 1$	α_A	α_B
$t = 2$	$\alpha_A + \beta_2 + \tau_0$	$\alpha_B + \beta_2$
$t = 3$	$\alpha_A + \beta_3 + \tau_1$	$\alpha_B + \beta_3 + \tau_0$

- Here $\hat{\tau}_1 = \left(Y_{A3} - Y_{B3} \right) + \left(Y_{A2} - Y_{B2} \right) - 2\left(Y_{A1} - Y_{B1}\right)$

Spurious identification of very long-run effects

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{h=0}^{1} \tau_h \mathbf{1}\left[t = E_i + h\right] + \varepsilon_{it}$$

$\mathbb{E}\left[Y_{it}\right]$	$i = A$	$i = B$
$t = 1$	α_A	α_B
$t = 2$	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
$t = 3$	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$

- Here $\hat{\tau}_1 = \left(Y_{A3} - Y_{B3}\right) + \left(Y_{A2} - Y_{B2}\right) - 2\left(Y_{A1} - Y_{B1}\right)$
- Estimand $\tau_1 = \tau_{A3} + \tau_{A2} - \tau_{B3}$ inevitably involves extrapolation that is invalid with heterogeneous effects

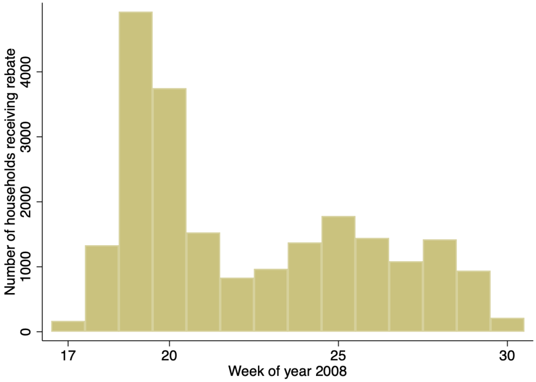
Cross-horizon contamination

Sun and Abraham (2021):

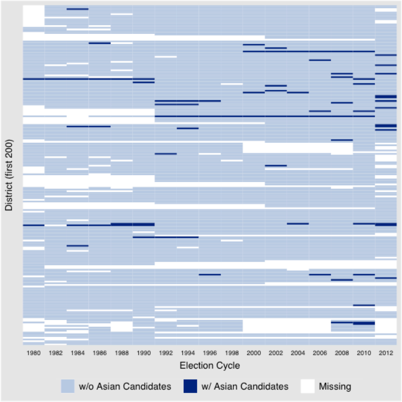
- Similar problems occur even for short-run effects in dynamic specification
    - Estimand τ_h is not an average of horizon-h effects: contaminated by heterogeneity of effects at other horizons
- And pre-trend coefficients are contaminated by treatment effect heterogeneity
    - Can be significant even if PTA holds!
- *Note*: these problems tend to be small in practice

Outline

Plot treatment timing

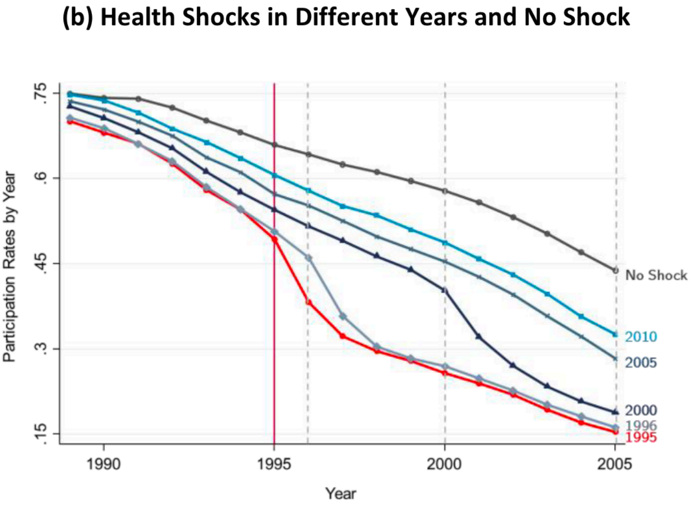


(From BJS, with no never-treated)



(from Chiu, Lan, Liu, and Xu (2023); package *panelView* in Stata & R)

Plot raw outcome data by cohort



(from Fadlon and Nielsen (2015))

Estimation robust to heterogeneous effects

- The problems arise from conventional specifications being too restrictive
- They are not fundamental to staggered adoption DiD
  - Under PTA, there are many valid 2x2 contrasts
- How to combine them?
  - Manual averaging approaches
    - Yield simpler estimators, more closely parallel conventional event studies
  - Imputation approaches
    - Transparently link assumptions to estimators, more versatile, often more efficient
  - Regression-based approaches

Manual averaging estimators

de Chaisemartin and D'Haultfoeuille (2020) for $h = 0$:

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$	$h = 0$			
$t = 3$		$h = 0$		
$t = 4$				
$t = 5$			$h = 0$	
$t = 6$				

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$	$h = 0$			
$t = 3$		$h = 0$		
$t = 4$				
$t = 5$			$h = 0$	
$t = 6$				

- For each cohort $E_i = e$: form the clean control group; compute **cohort-average treatment effect** ($CATT_{e,e+0}$) by comparing $Y_{ie} - Y_{i,e-1}$
    - *Note:* they compare to $e - 1$ only, ignoring earlier periods
- Average across cohorts weighting by cohort size

Manual averaging estimators (2)

Callaway and Sant'Anna (2021) for any $h \geq 0$:

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$	$h = 1$			
$t = 4$		$h = 1$		
$t = 5$				
$t = 6$			$h = 1$	

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$	$h = 1$			
$t = 4$		$h = 1$		
$t = 5$				
$t = 6$			$h = 1$	

- For each cohort $E_i = e$: form the control group as cohorts not treated by $e + h$; compute $CATT_{e,e+h}$ by comparing $Y_{i,e+h} - Y_{i,e-1}$

Manual averaging estimators (3)

Callaway and Sant'Anna (2021, 2nd estimator) and Sun and Abraham (2021): same but use never-treated controls only

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$	$h = 1$			
$t = 4$		$h = 1$		
$t = 5$				
$t = 6$			$h = 1$	

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$	$h = 1$			
$t = 4$		$h = 1$		
$t = 5$				
$t = 6$			$h = 1$	

- (If no never-treated, use latest-treated cohort instead)

Manual averaging: Pre-trend tests

de Chaisemartin and D'Haultfœuille (2024) pre-trend tests:

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$			$\ell = 3$	
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				

- For cohort e and lead $\ell > 1$, measure $Y_{i,e-\ell} - Y_{i,e-1}$
- Compare to the control group of units not treated by e
- Average across cohorts; test it's zero

Manual averaging: Pre-trend tests

Callaway and Sant'Anna (2021) default pre-trend tests:

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				

- Pre-trends always compare consecutive periods $\Rightarrow$ magnitudes not comparable to post-treatment coefficients (Roth (2024))

Imputation estimators

- PTA requires $\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t$
- No anticipation effects means $Y_{it} = Y_{it}(0)$ for untreated observations ($it \in \Omega_0$: both never-treated and pre-treatment)
- Imputation approach:
  1. Estimate α_i and β_t from untreated observations, e.g. by OLS
     - ⋆ Need pre-treatment obs for every unit to get $\hat{\alpha}_i$
     - ⋆ Need untreated obs in every period to get $\hat{\beta}_t$
  2. For each treated observation $it \in \Omega_1$, compute $\hat{\tau}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t$
     - ⋆ Each $\hat{\tau}_{it}$ is very noisy!
  3. Estimate $\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$ by $\hat{\tau}_w = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$
     - ⋆ Averaging across many units makes $\hat{\tau}_w$ consistent

Efficiency of imputation

Proposition (BJS) If $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$ for homoskedastic and serially uncorrelated ε_{it}, estimating $\hat{\alpha}_i, \hat{\beta}_t$ by OLS in the untreated sample yields most efficient $\hat{\tau}_w$ for any τ_w

Proof idea:

- This imputation estimator can be obtained by OLS from a very flexible regression

  $$Y_{it} = \alpha_i + \beta_t + \tau_{it} D_{it} + \varepsilon_{it}$$

  which dummies out each treated observation

- By Gauss-Markov, OLS is efficient for the vector of τ_{it} and for any linear combination τ_w

Comparison to manual averaging

- **Proposition** (BJS): Any unbiased estimator for τ_w under arbitrary heterogeneity of treatment effects can be represented as an imputation estimator for some unbiased $\hat{\alpha}_i, \hat{\beta}_t$
- E.g. manual averaging estimators are (implicitly) imputation estimators that estimate $\hat{\alpha}_i, \hat{\beta}_t$ differently than by OLS
- They use less information... although that's not always bad
  - Harmon (2022): If ε_{it} is a random walk, the de Chaisemartin and D'Haultfoeuille (2020) estimator is efficient for $h = 0$
  - Outcomes at $E_i - 1$ contain all useful information; earlier periods only add noise

BJS: Standard errors

- Represent $\hat{\tau}_w = \sum_{it} v_{it} Y_{it}$: $v_{it} = w_{it}$ for $it \in \Omega_1$; v_{it} can be computed for Ω_0
- True variance: $\mathrm{Var}\left[\hat{\tau}_w\right] = \mathrm{Var}\left[\sum_{it} v_{it}\varepsilon_{it}\right] = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$
- Plug-in clustered estimator: $\hat{\sigma}_w^2 = \sum_i \left(\sum_t v_{it}\hat{\varepsilon}_{it}\right)^2$
- For untreated obs, $\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t$
- Key challenge: for treated obs, $\hat{\varepsilon}_{it}^{\text{naive}} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{\tau}_{it} = 0$ by construction

BJS: Standard errors (2)

- If cohorts are large, instead use

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{\tau}_{E_i t} \approx \varepsilon_{it} + (\tau_{it} - \bar{\tau}_{E_i t})$$

  where $\hat{\tau}_{E_i t} =$ avg of $\hat{\tau}_{jt}$ in the same cohort $E_j = E_i$

  - SE are conservative when there is variation in τ_{it} within cohorts; otherwise asymptotically exact

- If cohorts are small, either replace $\hat{\tau}_{E_i t}$ with averages that pool multiple cohorts or use leave-out estimation; see BJS

  - *Note:* SE for all other estimators are only proved with large cohorts

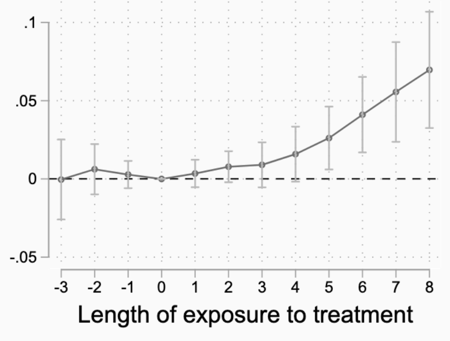
BJS: Pre-trend and other falsifications tests

- To test for pre-trends, always use only untreated observations
- Null hypothesis: $Y_{it} = \alpha_i + \beta_t + \varepsilon_{it}$
- Choose a richer alternative model: $Y_{it} = \alpha_i + \beta_t + \eta' W_{it} + \varepsilon_{it}$
  - E.g. anticipation effects: W_{it} are $\mathbf{1}\left[t = E_i - 1\right], \ldots, \mathbf{1}\left[t = E_i - L\right]$
  - Non-parallel linear trends: W_{it} are cohort dummies $\times t$
  - Structural break: W_{it} are cohort dummies $\times$ post financial crisis
- Use F-test for $\eta = 0$. (Don't include too many covariates to avoid low power)
- *Note 1:* not all violations affect causal estimates much
- *Note 2:* η should not be directly compared to causal effect estimates

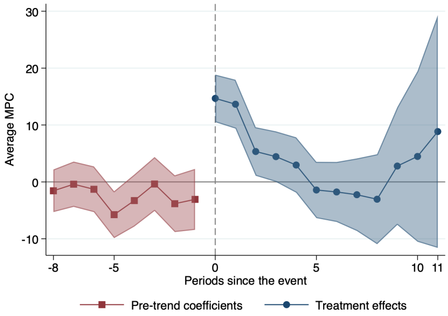
Estimators equivalent to imputation

- Gardner (2021): Two-Stage Diff-in-Diff estimator:
  - Estimate α_e, β_t using untreated observations (note cohort FEs instead of unit FEs)
  - Regress $Y_{it} - \hat{\alpha}_{E_i} - \hat{\beta}_t$ on the treatment dummy or event study indicators
- Liu, Wang, and Xu (AJPS, 2022): Counterfactual Estimators
  - Applied to both TWFE and richer Interactive FE models of $Y_{it}(0)$
- Be mindful of some differences in pre-trend tests, SE computation, behavior with unbalanced panels

Event study graphs from two groups of methods



(de Chaisemartin and D'Haultfœuille (2024) Figure 1)

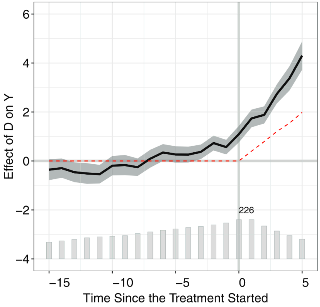


(BJS Figure 2a, using `event_plot` Stata command)

- Note different reference groups and different behavior of SE
- See Roth (2024) and Li and Strezhnev (2025) on correct interpretations of pre-trend coefficients with different methods

Liu et al. (2022) pre-trend test and graph

- Both for $h \geq 0$ and $h < 0$, plot averages of $Y_{it} - \hat{\alpha}_i - \hat{\beta}_t$



- Like with de Chaisemartin and D'Haultfœuille (2024) graphs, comparable on the left and right (although see Li and Strezhnev (2025))
- But no single reference period because imputation-based
- Unlike BJS, no explicit alternative hypothesis

Extended TWFE regression

- Can the convenience of OLS be preserved without negative weights and other problems?
- Wooldridge (2025): the problem with old-school estimators is restrictive specifications

$\Rightarrow$ Let's keep running regressions but more flexibly

$$Y_{it} = \alpha_i + \beta_t + \sum_e \sum_{s \geq e} \tau_{es} \mathbf{1}\left[E_i = e\right] \times \mathbf{1}\left[t = s\right] + \text{error},$$

where τ_{et} estimates CATT for cohort e in period t; then aggregate estimates as required

Extended TWFE regression

	$e=2$	$e=3$	$e=5$	$e=\infty$
$t=1$				
$t=2$	τ_{22}			
$t=3$	τ_{23}	τ_{33}		
$t=4$	τ_{24}	τ_{34}		
$t=5$	τ_{25}	τ_{35}	τ_{55}	
$t=6$	τ_{26}	τ_{36}	τ_{56}	

- Numerically equivalent to BJS for CATT-based estimands in complete panels

Local projection DiD

- Dube, Girardi, Jorda, and Taylor (2023) propose to estimate, for each h,

$$Y_{i,t+h} - Y_{i,t-1} = \beta_{ht} + \tau_h \mathbf{1}\left[t = E_i\right] + \text{error}$$

  on the subsample where $E_i = t$ or $E_i > t + h$

  - Combines the local projections estimator for time series of Jorda (2005) with "stacking" approach of Cengiz, Dube, Lindner, and Zippere (2019, Appendix D)

- τ_h estimates a convex weighted average of treatment effects

  - Can be reweighted to get Callaway and Sant'Anna (2021)

  - Instead of subtracting $Y_{i,t-1}$ can subtract average of several pre-period outcomes $\implies$ closer to imputation

Does heterogeneity-robust estimation matter?

BJS application with no never-treated units:

- Yes, relative to specifications that restrict dynamics
- Yes, for the spurious long-run coefficients
- For the short-run, the semi-dynamic specification seems fine

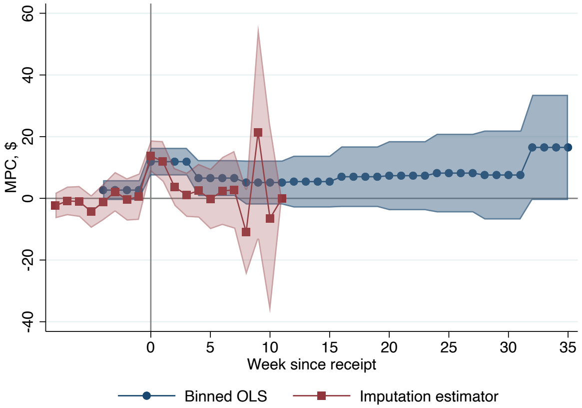
Setting: marginal propensity to spend (MPC) from the 2008 Economic Stimulus Payments (tax rebates)

- Staggered disbursement of rebates. Weekly spending data from Nielsen

Broda and Parker (2014) find a very large MPC from a monthly-binned specification:

$$Y_{it} = \alpha_i + \beta_t + \sum_{m=-1}^{\infty} \tau_m \mathbf{1}\left[t - E_i \in \{4m-3, \ldots, 4m\}\right] + \text{error}_{it}$$

BJS results

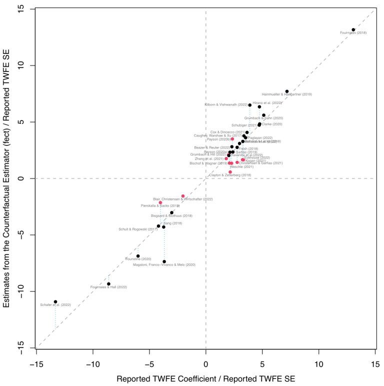


- Because the effects are fast decaying, binned specification overstates them in the first month
- Binned specification also extrapolates them to all future months

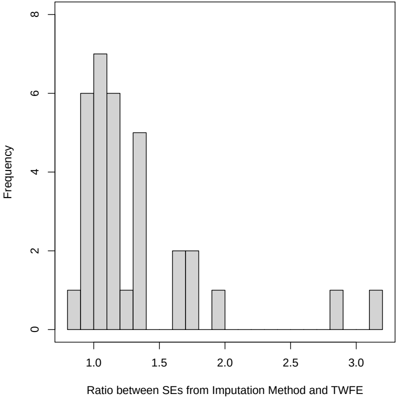
Chiu, Lan, Liu, and Xu (2023) are less convinced

Chiu et al. (2023) reanalyze main significant TWFE coefficients in 37 political science papers and don't find big differences — but also no cost of being robust

Figure 4. TWFE vs. Imputation Method



(From Chiu et al.)



(Computation by Yixing Xu)

Software packages

Method	Stata	R	Python
Borusyak et al. 2024	`did_imputation`	`didimputation` *(limited)*	*(coming soon)*
Callaway and Sant'Anna 2021	`csdid`, `csdid2`	`did`	`csdid`
de Chaisemartin, D'Haultfoeuille 2020, 2024	`did_multiplegt_dyn`	`did_multiplegt_dyn`	*
Sun and Abraham 2021	`eventstudyinteract`	*	*
Plotting any event studies	`event_plot`	—	*(coming soon)*
Visualizing event timing	panelview	panelView	—

* Callaway and Sant'Anna's packages can be used instead

References I

Angrist, J. (1998): "Estimating the labor market impact of voluntary military service using social security data on military applicants," *Econometrica*, 66, 249–288.

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 91, 3253–3285.

Broda, C. and J. A. Parker (2014): "The economic stimulus payments of 2008 and the aggregate demand for consumption," *Journal of Monetary Economics*, 68, S20–S36, publisher: Elsevier.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*.

Cengiz, D., A. Dube, A. Lindner, and B. Zippere (2019): "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, 1405–1454.

Chiu, A., X. Lan, Z. Liu, and Y. Xu (2023): "What To Do (and Not to Do) with Causal Panel Analysis under Parallel Trends: Lessons from A Large Reanalysis Study," .

de Chaisemartin, C. and X. D'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–2996, arXiv: 1803.08807v1.

References II

de Chaisemartin, C. and X. D'Haultfœuille (2024): "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *Review of Economics and Statistics*, arXiv: 2007.04267v1.

Dube, A., D. Girardi, O. Jorda, and A. M. Taylor (2023): "A Local Projections Approach to Difference-in-Differences Event Studies," *Working Paper*.

Fadlon, I. and T. H. Nielsen (2015): "Household Responses to Severe Health Shocks and the Design of Social Insurance," *Working Paper*.

Gardner, J. (2021): "Two-stage differences in differences," *Working Paper*.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277, publisher: Elsevier B.V.

Harmon, N. A. (2022): "Difference-in-Differences and Efficient Estimation of Treatment Effects," .

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

Li, Z. and A. Strezhnev (2025): "Benchmarking parallel trends violations in regression imputation difference-in-differences," *Working Paper*.

References III

Liu, L., Y. Wang, and Y. Xu (2022): "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *American Political Science Review*.

Roth, J. (2024): "A Cautionary Note on Interpreting Event-Studies from Recent Difference-in-Differences Methods," .

Sun, L. and S. Abraham (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*.

Wooldridge, J. M. (2025): "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Empirical Economics*.