Part D: Panel Data Methods

D4: Difference-in-Differences: Extensions

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

Extensions

Feature	Baseline	Extensions
Model of $Y(0)$:	TWFE (PTA)	Including covariates Partial identification approach Multiplicative models Factor models
Model of τ_{it}:	Arbitrary heterogeneity	Ex ante restrictions
Maintained assumptions:	No anticipation effects No spillovers	Some anticipation effects Some spillovers
Treatment:	Binary Absorbing	Discrete or continuous Treatment reversals
Data structure:	Panel	Repeated cross-sections Two-dimensional cross-sections Triple-difference designs

D4 outline

1. Richer models of $Y(0)$: covariates and partial identification
2. Restrictions on treatment effects
3. Anticipation and spillover effects
4. Treatments beyond binary absorbing
5. Beyond panel data

DiD with covariates (1)

Two ways of thinking about covariates:

1. Borusyak, Jaravel, and Spiess (2024): make the model of $Y_{it}(0)$ richer,

$$\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t + \gamma' X_{it}$$

   - Continuous covariates (avoid bad controls!)
   - Unit-specific linear trends $\gamma_i \cdot t$
   - Time-interacted baseline characteristics $\gamma_t' X_i$,
     e.g. state-year FE in a county-level panel
   - *Exercise:* What do these mean in first-differences?

$\Rightarrow$ Estimate γ from untreated observations only (along with α_i, β_i), then imputation

DiD with covariates (2)

2. Impose PTA conditional on baseline characteristics

- Abadie (2005) and Sant'Anna and Zhao (2020) for non-staggered DiD:

$$\mathbb{E}\left[\Delta Y_{it}(0) \mid G_i = 1, X_i\right] = \mathbb{E}\left[\Delta Y_{it}(0) \mid G_i = 0, X_i\right]$$

  - A standard problem of selection on observables for ΔY_{it} (for ATT)

  $\Rightarrow$ Propose to use inverse probability weighting (Abadie (2005)), augmented IPW, or other covariate adjustment (Sant'Anna and Zhao (2020))

- Callaway and Sant'Anna (2021) extend this approach to manual averaging under staggered adoption
  - Impose $\mathbb{E}\left[\Delta Y_{it}(0) \mid X_i, E_i = e\right] = \mathbb{E}\left[\Delta Y_{it}(0) \mid X_i, E_i > s\right]$ for all $s \geq t \geq e$
  - Involves estimating a separate pscore for each e, t

DiD with covariates: What NOT to do

- With heterogeneous effects, do not just add covariates to TWFE and event study regressions even in non-staggered settings, e.g. avoid

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \tau D_{it} + \gamma' X_{it} + \text{error}$$

  - $\hat{\gamma}$ will use treated observations too, misattributing some treatment effects

- *Example:* Wolfers (2006) studies the effect of unilateral divorce laws on divorce rates

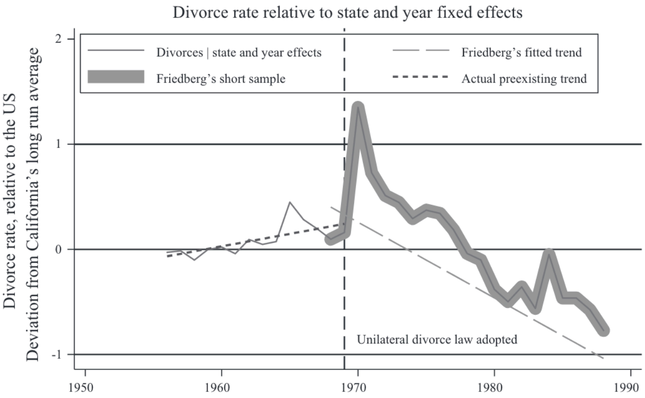
  - Reanalyzes Friedberg (1998) who found a positive effect when including state-specific linear trends

Results without and with state-specific trends

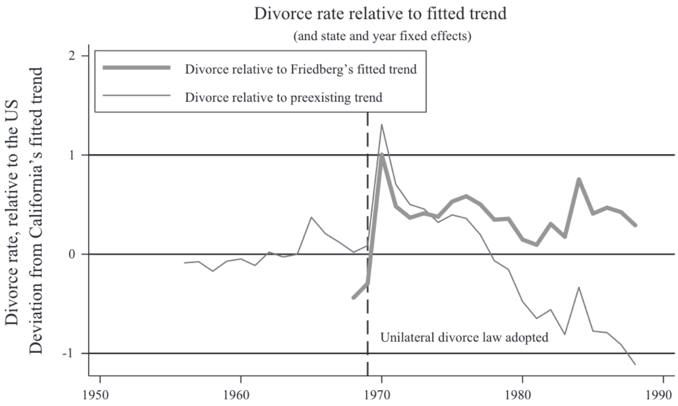
	(1) Basic specification	(2) State-specific trends linear
	Panel A. Friedberg (1998)	
Unilateral	0.004	0.447
	(0.056)	(0.050)
Year effects	F = 89.0	F = 95.3
State effects	F = 217.3	F = 196.2
State trend, linear	No	F = 24.7
State trend, quadratic	No	No
Adjusted R^2	0.946	0.976

Which results should we trust: with extra controls or without?

Example of California



Example of California (2)



- With appropriate estimators, state-specific trends don't change estimates (de Chaisemartin and D'Haultfœuille (2023))

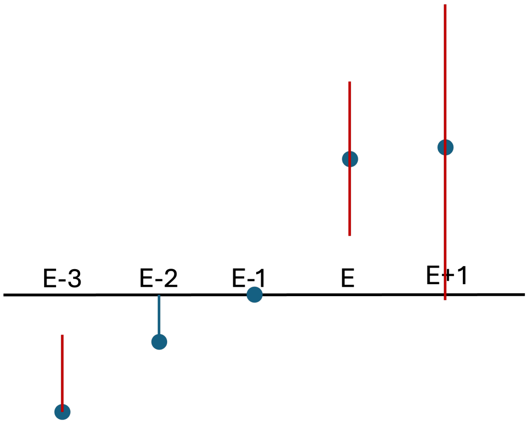
Partial identification approach

Rambachan and Roth (2023) for non-staggered DiD: checking robustness to PTA violations

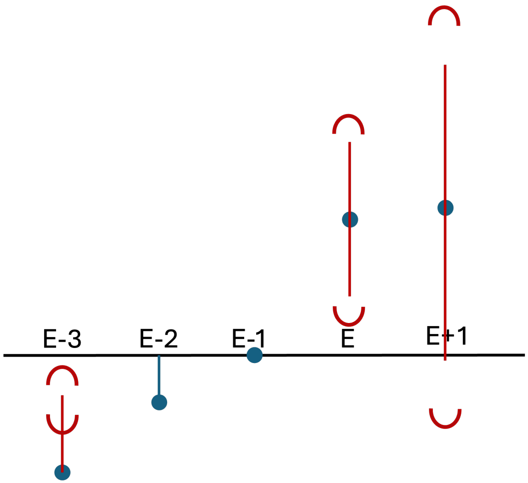
- Let $\delta_t = \mathbb{E}\left[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = 1\right] - \mathbb{E}\left[Y_{it}(0) - Y_{i,t-1}(0) \mid G_i = 0\right]$
- PTA requires $\delta_t = 0$; Estimable for $t < E$ but not $t \geq E$
- Weaker assumptions for $t \geq E$ as robustness checks:
  0. *(Not in the paper) Exercise:* what would $\delta_t = \text{const}$ mean?
  1. Differential trends are not too large: $|\delta_t| \leq M \cdot \max_{s<E} |\delta_s|$

Simplified case: $N = \infty$

Imagine a stylized event study plot [*not* a plot of δ_t]:



With sampling noise



Partial identification approach (cont'd)

Rambachan and Roth (2023) weaker assumptions for $t \geq E$:

1. Differential trends are not too large: $|\delta_t| \leq M \cdot \max_{s<E} |\delta_s|$
2. Differential trends are smooth: $|\delta_t - \delta_{t-1}| \leq M$
    - *Note:* for $M = 0$ this is not PTA but a linear differential trend

- When pre-trends are noisily estimated, conf.interval will be wider
- Need to pick M or compute the largest M that doesn't kill your findings
- See Liu (2025) for extensions to staggered adoption via manual averaging and imputation

Outline

Restrictions on treatment effects

So far we have considered estimators robust to arbitrary effect heterogeneity

- This is a strong demand on the estimator $\Longrightarrow$ can't get all estimands; others can be noisy
- Model is very asymmetric: strong assumptions on $Y_{it}(0)$ and none on $Y_{it}(1)$
- With so much heterogeneity possible, are ATTs informative about future policy?
  "Anyone who makes a living out of data analysis probably believes that heterogeneity is limited enough that the well-understood past can be informative about the future" (Angrist and Pischke, 2010)

Restrictions on treatment effects (2)

Can we identify more estimands (e.g. long-run effects) or get more power for the same estimands using extra restrictions on τ_{it} and $Y_{it}(1)$?

- Impose some simple model of treatment effects, $\tau_{it} = \Gamma_{it}'\theta$: e.g.
  - $\tau_{it} = \bar{\tau}$ is homogeneous across i and t $\implies$ *static TWFE specification*
  - $\tau_{it} \equiv \tau_{t-E_i}$ is homogeneous across i for any given horizon $\implies$ *semi-dynamic*
  - $\tau_{it} = 0$ when $t > E_i + K$ for some K
  - $\tau_{it+1} = \tau_{it}$ when $t > E_i + K$
- **Proposition** (BJS): Under PTA, spherical errors, and model of τ_{it}, the efficient unbiased estimator of τ_w is:
  1. Run OLS on the full sample for the "true" model: $Y_{it} = \alpha_i + \beta_t + D_{it}\Gamma_{it}'\theta + \varepsilon_{it}$
  2. Compute $\hat{\tau}_{it} = \Gamma_{it}'\hat{\theta}$ and $\hat{\tau}_w = \sum_{it\in\Omega_1} w_{it}\hat{\tau}_{it}$

Outline

Anticipation effects

- If anticipation effects of at most L periods are possible, redefine the treatment to L periods earlier
  - Interpret the estimates accordingly
  - Can separately estimate anticipation effects vs. post-treatment effects
- Same if there is a gap between announcement and implementation (that can vary by i)

Spillovers (1)

- Consider a 2-period DiD with spillovers: $Y_{i2} = Y_{i2}(\boldsymbol{D}_2) \equiv Y_{i2}(\mathbf{0}) + \tau_{i2}(\boldsymbol{D}_2)$
  - $\tau_{i2}(\boldsymbol{d}_2)$ is the effect of a vector of treatments $\boldsymbol{d}_2$ on person i's outcome relative to the counterfactual where nobody is treated
- Impose PTA on untreated outcomes:
  $\mathbb{E}\left[Y_{i2}(\mathbf{0}) - Y_{i1}(\mathbf{0}) \mid D_{i2} = 1\right] = \mathbb{E}\left[Y_{i2}(\mathbf{0}) - Y_{i1}(\mathbf{0}) \mid D_{i2} = 0\right]$
- Then DiD yields *relative* effect $\mathbb{E}\left[\tau_{i2}(\boldsymbol{D}_2) \mid D_{i2} = 1\right] - \mathbb{E}\left[\tau_{i2}(\boldsymbol{D}_2) \mid D_{i2} = 0\right]$
- Identifies the *aggregate* ATT if and only if $\mathbb{E}\left[\tau_{i2}(\boldsymbol{D}_2) \mid D_{i2} = 0\right] = 0$
  - Still cannot separate direct and indirect effects

Spillovers (2)

What to do if control group can be affected?

- Specify which untreated observations can experience spillover effects (e.g., because within a short geographic distance from a treated)
- View them as a separate treated group
- If there are enough observations that are not treated directly or indirectly, use them as controls
- Identify both causal effects on the treated and spillover effects on the untreated as two estimands

Outline

Continuous treatment intensity (1)

When the treatment is not binary, there are multiple scenarios

- Are there untreated units in every period and untreated periods for every unit?
- **Yes:** e.g. in the BJS application, the tax rebate amount is a continuous treatment but that's no problem
  - PTA on $Y_{it}(0)$ still identifies counterfactual trends
  - Can binarize the treatment (without loss) and impute $\tau_{it} = Y_{it}(D_i) - Y_{it}(0)$
  - Then identify e.g. (1) average MPC as % of household-specific rebate or (2) the total increase in spending as % of total rebates:

$$\tau_w^{(1)} = \frac{1}{|\Omega_{1,h}|} \sum_{t-E_i=h} \frac{\tau_{it}}{\text{Rebate}_i} \quad \text{or} \quad \tau_w^{(2)} = \frac{\sum_{t-E_i=h} \tau_{it}}{\sum_{t-E_i=h} \text{Rebate}_i}$$

Continuous treatment intensity (2)

- Are there untreated units in every period, untreated periods for every unit?
- **No:** Then PTA on $Y_{it}(0)$ is unhelpful
- Can make progress by randomization, restrictions on heterogeneous effects, or other/stronger PTAs
- de Chaisemartin and D'Haultfœuille (2025) develop a "random coefficients distributed-lag linear [causal] model":

$$Y_{it} = \alpha_i + \beta_t + \sum_{h=0}^{H} \tau_{ih} D_{i,t-h} + \varepsilon_{it}$$

  - where effects τ_{ih} vary across i and can be correlated with treatments;
  - but don't depend on t (only on h); only appear up to a known lag H; are linear in dosage
- Expect more work on imputation with other restrictions on heterogeneity

Continuous treatment intensity (3)

Making progress with no untreated units via other/stronger PTAs

- Appendix of de Chaisemartin and D'Haultfoeuille (2020) considers settings where D_{it} is discrete and there are many "stayers" ($D_{it} = D_{i,t-1}$)
  - Decide to compare stayers and switchers with the same initial treatment

$\Rightarrow$ Impose $\mathbb{E}\left[\Delta Y_{i2}(d) \mid D_{i1} = d, D_{i2}\right] = \mathbb{E}\left[\Delta Y_{i2}(d) \mid D_{i1} = d\right]$

  - Can identify e.g. $\mathbb{E}\left[Y_{i2}(D_{i2}) - Y_{i2}(D_{i1}) \mid D_{i2} > D_{i1}\right]$

Continuous treatment intensity (4)

Making progress with no untreated units via other/stronger PTAs

- de Chaisemartin, d'Haultfoeuille, Pasquier, Sow, and Vazquez-Bare (2025) extend this to continuous treatments:
    - E.g. D_{it} = state-level minimum wage; changes only when there is a local reform
    - For switchers, they can't find stayers with exact same $D_{i1} \implies$ estimate $\mathbb{E}\left[\Delta Y_{i2}(d) \mid D_{i1} = D_{i2} = d\right]$ nonparametrically for stayers
- See Callaway, Goodman-Bacon, and Sant'Anna (2021) for another version of PTA

Treatment reversals

What if treatment is not an absorbing state?

- E.g. Adda (2016): Economic effects of epidemics (e.g., flu); D_{it} = dummy for school holidays
- Key assumption: **no carryover effects**
  - If lagged treatment can have effects far into the future, there is no control group
- Imputation approach extends immediately
  - Estimate α_i, β_t from untreated observations both before and after treatment
- de Chaisemartin and D'Haultfoeuille (2020) use groups with $D_{i,t-1} = D_{it} = 1$ as controls for "leavers" who switch from $D_{i,t-1} = 1$, $D_{it} = 0$
  - Requires assuming PTA on $Y_{it}(1)$ for them. Parameter is no longer ATT
- The no-carryover effects assumption should be tested

Outline

Repeated cross-sections

- Instead of a state-year panel, observe a new random sample of individuals i in each state s and year t
  - Can't include individual FE but don't need to
  - Model $\mathbb{E}[Y_i(0)] = \alpha_{s(i)} + \beta_{t(i)}$

Two-dimensional cross-sections

- Suppose we have data by region i and age group g in a single period
- $D_{ig} = \mathbf{1}[g < E_i]$, e.g. a policy applies to exactly 18 year-olds but was rolled out in a staggered way across regions
  - Older age cohorts in the same region experienced it, younger didn't
  - Some age cohorts are treated in states that adopted early but not others
  - Impose $\mathbb{E}[Y_{ig}(0)] = \alpha_i + \beta_g$; use older groups and late-adopting regions as controls
- Both manual averaging and imputation extend by redefining variables
  - E.g. $D_{ib} = \mathbf{1}[b \geq B_i]$ where b is birth year and B_i is the earliest birth year eligible

Triple-differences (1)

- Suppose for each state $i = NJ, PA$ and period $t = 1, 2$ we observe groups $g = L, H$: low- and high-wage occupations
- Min.wage should affect group L only: $D_{igt} = \mathbf{1}[i = NJ] \times \mathbf{1}[g = L] \times \mathbf{1}[t = 2]$
- Strategy 1 (**not** triple-differences):
  - Use DiD on data from group L to learn the ATT for them
  - Also run **placebo** DiD for group H: test $\tau_H = 0$ in

$$Y_{iHt} = \alpha_{iH} + \beta_{Ht} + \tau_H \mathbf{1}[i = NJ] \times \mathbf{1}[t = 2] + \varepsilon_{iHt}$$

Triple-differences (2)

- Strategy 2 (**triple-differences**):
  - Allow non-parallel trends between NJ and PA: placebo would fail
  - But assume violations of PTA are the same for $g = L, H$:

$$\mathbb{E}\left[\Delta Y_{NJ,L}(0) - \Delta Y_{PA,L}(0)\right] = \mathbb{E}\left[\Delta Y_{NJ,H}(0) - \Delta Y_{PA,H}(0)\right] \equiv \alpha$$

  - Equivalent to $\mathbb{E}\left[\Delta Y_{ig}(0)\right] = \alpha_i + \beta_g$ and $\mathbb{E}\left[Y_{igt}(0)\right] = \alpha_{it} + \beta_{gt} + \gamma_{ig}$
  - Can estimate $ATT = \tau_{NJ,L,2}$ from 2

$$Y_{igt} = \tilde{\alpha}_{it} + \tilde{\beta}_{gt} + \tilde{\gamma}_{ig} + \tau D_{igt} + \text{error}_{igt}$$

Triple-differences (3)

With multiple groups and periods:

- Manual averaging can be manually extended
- Imputation extends immediately: estimate

$$Y_{igt}(0) = \alpha_{it} + \beta_{gt} + \gamma_{ig} + \text{error}$$

on all untreated observations


Jonathan Roth
@jondr44

"the credibility of the method decreases in the square of the number of differences you need"

8:45 AM · Aug 1, 2024 · **3,681** Views

Triple-diff $\neq$ effect heterogeneity

Do not confuse triple-differences with estimating effect heterogeneity: in triple-diff,

- Group H is untreated in NJ at $t = 2$
- Differential trends for group H between NJ and PA are not a consequence of the treatment, but unexplained
- The estimand is the causal effect on group L in NJ at $t = 2$, not some difference in effects
- The regression may look the same, but the underlying assumptions and the estimand are different

References I

Abadie, A. (2005): "Semiparametric Difference-in-Difference Estimators," *Review of Economic Studies*, 72, 1–19, iSBN: 0034-6527.

Adda, J. (2016): "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data," *Quarterly Journal of Economics*, 131, 891–941, arXiv: 1011.1669v3 ISBN: 9788578110796.

Angrist, J. D. and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.

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

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," *Working paper*, arXiv: 2107.02637.

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

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

References II

de Chaisemartin, C., X. d'Haultfoeuille, F. Pasquier, D. Sow, and G. Vazquez-Bare (2025): "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period," *Working Paper*.

de Chaisemartin, C. and X. D'Haultfœuille (2025): "Treatment-Effect Estimation in Complex Designs under a Parallel-trends Assumption," *Technical Report*.

de Chaisemartin, C. and X. D'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *The Econometrics Journal*, 26, C1–C30.

Friedberg, L. (1998): "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data," *NBER Working Paper*.

Liu, Z. (2025): "Cohort-Anchored Robust Inference for Event-Study with Staggered Adoption," *arXiv preprint*.

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

References III

Sant'Anna, P. H. and J. Zhao (2020): "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 219, 101–122, arXiv: 1812.01723 Publisher: Elsevier B.V.

Wolfers, J. (2006): "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results," *American Economic Review*, 1802–1820.