A Local Projections Approach to Difference-in-Differences Event Studies

Arindrajit Dube[†]   Daniele Girardi[*]   Òscar Jordà[‡]   Alan M. Taylor[§]

February 2023

[†] University of Massachusetts, Amherst; NBER; and IZA
[*] University of Massachusetts, Amherst;
[‡] Federal Reserve Bank of San Francisco; University of California, Davis; and CEPR
[§] University of California, Davis; NBER; and CEPR

How to estimate Difference-in-Differences (DiD) with multiple treated groups & treatment periods?

- Recent literature shows that conventional TWFE implementations can be severely biased.

- A new regression-based framework: LP-DiD.
  - Local projections (Jordà 2005) + clean controls (Cengiz et al 2019).

- Simulation evidence to assess its performance.

- Empirical applications:
  1. The effect of banking deregulation on the wage share.
 2. Democracy & growth.

Why do we need yet another DiD estimator?
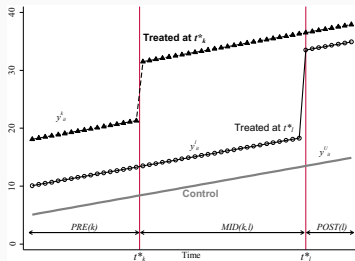
Advantages of LP-DiD:

- Simpler, more transparent, easier to code, and faster to compute than other recent DiD estimators.

- Flexible: offers a general framework that can easily accommodate different settings.

- Allows matching on pre-treatment outcomes and other time-varying covariates.

Difference-in-Differences (DiD)
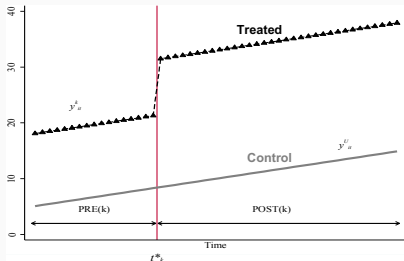
2x2 Setting   **Staggered Setting**

(Visual examples from Goodman-Bacon, 2021)

The conventional (until recently) DiD estimator: TWFE

- Static TWFE
$$y_{it} = \alpha_i + \delta_t + \beta^{TWFE} D_{it} + \epsilon_{it}$$

- Event-study (distributed lags) TWFE

$$y_{it} = \alpha_i + \delta_t + \sum_{h=-Q}^{H} \beta_h^{TWFE} D_{it-h} + \epsilon_{it}$$

- OK in the 2x2 setting.
- Biased even under parallel trends with staggered treatment, if treatment effects are dynamic and heterogeneous.

The problems with TWFE in the staggered setting

- TWFE as weighted-average of 2x2 comparisons (Goodman-Bacon 2021)
    1. Newly treated vs Never treated;
    2. Newly treated vs Not-yet treated;
    3. Newly treated vs Earlier treated.
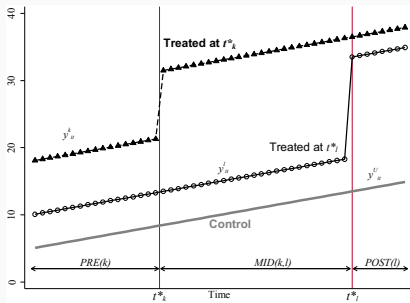
The problems with TWFE in the staggered setting

- TWFE as weighted-average of 2x2 comparisons (Goodman-Bacon 2021)
  1. Newly treated vs Never treated;
  2. Newly treated vs Not-yet treated;
  3. Newly treated vs Earlier treated.

- Bias formula for TWFE (Goodman-Bacon 2021)

$$p\lim_{N\to\infty} \hat{\beta}^{TWFE} = VWATT + VWCT - \Delta ATT$$

The problems with TWFE in the staggered setting

- TWFE as weighted-average of 2x2 comparisons (Goodman-Bacon 2021)

  1. Newly treated vs Never treated;
  2. Newly treated vs Not-yet treated;
  3. Newly treated vs Earlier treated.

- Bias formula for TWFE (Goodman-Bacon 2021)

$$p \lim_{N \to \infty} \hat{\beta}^{TWFE} = VWATT + VWCT - \Delta ATT$$

- TWFE as a weighted-average of cell-specific ATTs (de Chaisemartin & D'Haultfoeuille 2020)

$$E\left[\hat{\beta}^{TWFE}\right] = E\left[\sum_{(g,t):D_{gt}=1} \frac{N_{g,t}}{N_1} w_{g,t} \Delta_{g,t}\right]$$

  o Weights can be negative!

A Local Projections Diff-in-Diff Estimator (LP-DiD)
Baseline version

Setting & Assumptions:

- Binary absorbing treatment.
- Staggered adoption.
- Treatment effects can be dynamic & heterogeneous.
- No anticipation.
- Parallel trends.

A Local Projections Diff-in-Diff Estimator (LP-DiD)
Baseline version

Estimating equation:

$$
\begin{aligned}
y_{i,t+h} - y_{i,t-1} = \quad & \beta_h^{LP-DiD} \Delta D_{it} && \} \text{ treatment indicator} \\
& + \delta_t^h && \} \text{ time effects} \\
& + e_{it}^h ; && \text{for } h = 0, \ldots, H .
\end{aligned}
$$

restricting the sample to observations that are either:

$$
\begin{cases}
\text{newly treated} & \Delta D_{it} = 1 , \\
\text{or clean control} & D_{i,t+h} = 0
\end{cases}
$$

What does LP-DiD identify?

- A variance-weighted average effect:

$$E(\hat{\beta}_h^{LP-DiD}) = \sum_{g \neq 0} \omega_{g,h}^{LP-DiD} \tau_g(h)$$

  ○ $\tau_g(h) = h$-periods forward ATT for treatment-cohort g.

- No negative weights.

What does LP-DiD identify?

- A variance-weighted average effect:

$$E(\hat{\beta}_h^{LP-DiD}) = \sum_{g \neq 0} \omega_{g,h}^{LP-DiD} \tau_g(h)$$

  - $\tau_g(h) = h$-periods forward ATT for treatment-cohort g.

- No negative weights.

- Weights depend on subsample size & treatment variance:

$$\omega_{g,h}^{LP-DiD} = \frac{N_{CCS_{g,h}}[n_{gh}(n_{c,g,h})]}{\sum_{g \neq 0} N_{CCS_{g,h}}[n_{g,h}(n_{c,g,h})]} \, ,$$

  - $N_{CCS_{g,h}} = $ size of subsample including group g & its clean controls.
  - $[n_{gh}(n_{c,g,h})] = $ treatment variance in that subsample.

LP-DiD as a 'swiss knife'

Flexibility in choosing a weighting scheme

- Can apply any desired weights through weighted regression.

- Equally-weighted ATT: reweight by $1/(\omega_{g,h}^{LP-DiD}/N_g)$.

- $\omega_{g,h}^{LP-DiD}$ easy to compute from 'residualized' treatment indicator.

- Can also use regression adjustment.

LP-DiD as a 'swiss knife'

Easy to adapt to different settings

- Covariates & outcome lags
- Non-absorbing treatment
- Continuous treatment variable

LP-DiD with covariates

LP-DiD with covariates and outcome lags

Estimating equation:

$$
\begin{aligned}
y_{i,t+h} - y_{i,t-1} = \quad & \beta_h^{LP-DiD} \Delta D_{it} && \} \text{ treatment indicator} \\
& + \sum_{p=1}^{P} \gamma_p^h \Delta y_{i,t-p} && \} \text{ outcome lags} \\
& + \sum_{m=1}^{M} \sum_{p=0}^{P} \gamma_{m,p}^h \Delta x_{m,i,t-p} && \} \text{ covariates} \\
& + \delta_t^h && \} \text{ time effects} \\
& + e_{it}^h \, ; && \text{for } h = 0, \ldots, H \, ,
\end{aligned}
$$

restricting the sample to observations that are either

$$
\begin{cases}
\text{newly treated} & \Delta D_{it} = 1 \, , \\
\text{or clean control} & D_{i,t+h} = 0
\end{cases}
$$

LP-DiD with covariates

LP-DiD with covariates and outcome lags

Estimating equation:

$$
\begin{aligned}
y_{i,t+h} - y_{i,t-1} = \quad & \beta_h^{LP-DiD} \Delta D_{it} && \} \text{ treatment indicator} \\
& + \sum_{p=1}^{P} \gamma_p^h \Delta y_{i,t-p} && \} \text{ outcome lags} \\
& + \sum_{m=1}^{M} \sum_{p=0}^{P} \gamma_{m,p}^h \Delta x_{m,i,t-p} && \} \text{ covariates} \\
& + \delta_t^h && \} \text{ time effects} \\
& + e_{it}^h \ ; && \text{for } h = 0, \ldots, H \,,
\end{aligned}
$$

restricting the sample to observations that are either

$$
\begin{cases}
\text{newly treated} & \Delta D_{it} = 1 \,, \\
\text{or clean control} & D_{i,t+h} = 0
\end{cases}
$$

- Covariates will generally alter the weights.

- Can use p-score methods to make sure weights remain non-negative, or regression adjustment to get equally-weighted ATT.

12

LP-DiD with non-absorbing or continuous treatment

- In general: Adapt the clean control condition to the specific setting.

LP-DiD with non-absorbing or continuous treatment

- In general: Adapt the clean control condition to the specific setting.
- Example for non-absorbing treatment:

$$\begin{cases} \text{treatment} & (\Delta D_{it} = 1) \quad \& \quad (\Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L; j \neq 0) \\ \text{clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L \end{cases}$$

LP-DiD with non-absorbing or continuous treatment

- In general: Adapt the clean control condition to the specific setting.
- Example for non-absorbing treatment:

$$\left\{ \begin{array}{ll} \text{treatment} & (\Delta D_{it} = 1) \quad \& \quad (\Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L; j \neq 0) \\ \text{clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L \end{array} \right.$$

- Example for continuous treatment X_{it}:

$$\left\{ \begin{array}{ll} \text{movers} & (|\Delta X_{it}| > c) \quad \& \quad (|\Delta X_{i,t-j}| \leq c \text{ for } -h \leq j \leq L; j \neq 0) \\ \text{quasi-stayers} & |\Delta X_{i,t-j}| \leq c \text{ for } -h \leq j \leq L \end{array} \right.$$

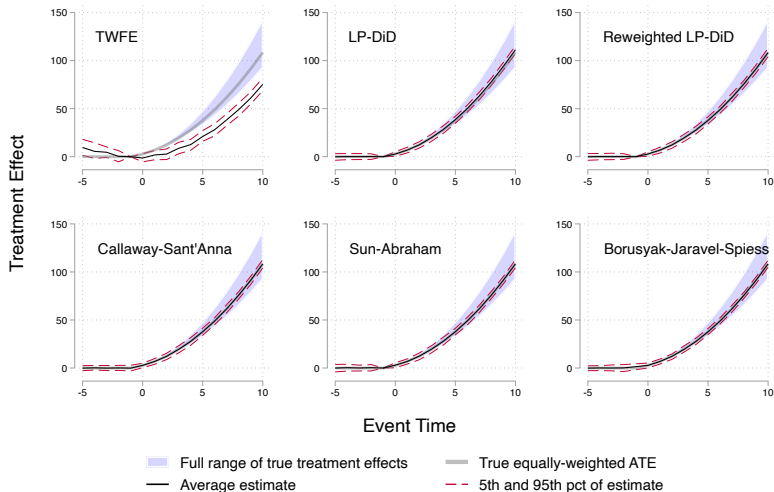- Underlying assumption: treatment effects *stabilize* after L periods.

13

Simulation

- N=500; T=50.
- DGP:

  $Y_{0it} = \rho Y_{0,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}; \quad -1 < \rho < 1; \quad \lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25)$

- Binary staggered treatment.
- TE grows in time for 20 periods, and is stronger for early adopters.

Simulation

- N=500; T=50.

- DGP:
  $$Y_{0it} = \rho Y_{0,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}; \quad -1 < \rho < 1; \quad \lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25)$$

- Binary staggered treatment.

- TE grows in time for 20 periods, and is stronger for early adopters.

1 Exogenous treatment

  - Units randomly assigned to 10 groups of size $N/10$
  - One group never treated; others treated at $t = 11, 13, 15 \ldots, 27$.

Simulation

- N=500; T=50.

- DGP:
  $$Y_{0it} = \rho Y_{0,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}; \quad -1 < \rho < 1; \quad \lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25)$$

- Binary staggered treatment.

- TE grows in time for 20 periods, and is stronger for early adopters.

1 Exogenous treatment
    - Units randomly assigned to 10 groups of size $N/10$
    - One group never treated; others treated at $t = 11, 13, 15 \ldots, 27$.

2 Endogenous treatment
    - Probability of treatment depends on past outcome dynamics.
    - Negative shocks increase probability of treatment.
    - Parallel trends holds only conditional on outcome lag.

Simulation 1 – exogenous treatment scenario

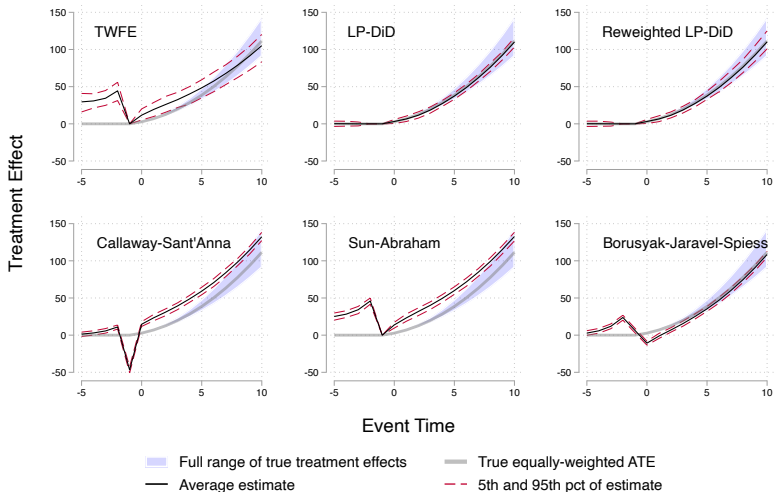## True effect path and estimates from 200 replications

Simulation 1 – exogenous treatment scenario
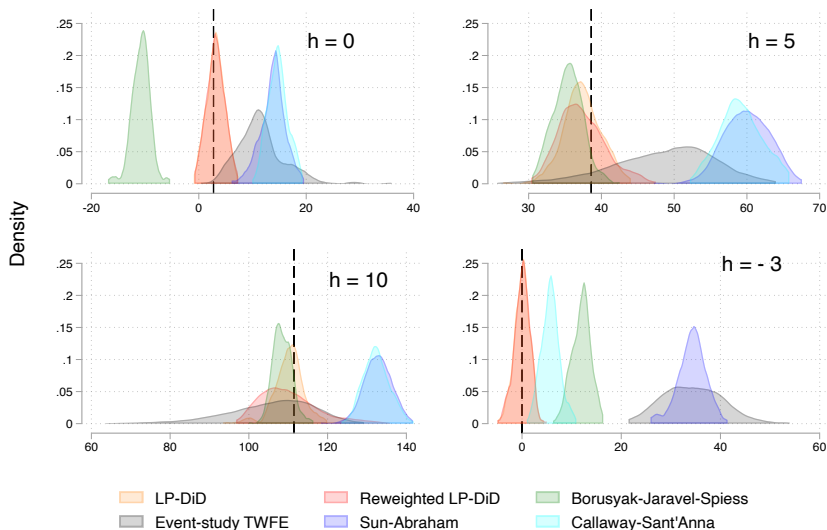
Distribution of estimates from 200 replications.

Legend: LP-DiD, Reweighted LP-DiD, Borusyak-Jaravel-Spiess, Event-study TWFE, Sun-Abraham

True effect path and estimates from 200 replications

Distribution of estimates from 200 replications.

18

Simulation Evidence

Computational speed

Estimating the treatment effect path in a single simulation of the synthetic dataset with exogenous treatment timing:

- **TWFE**: 1.04 seconds
- **LP-DiD**: 1.2 seconds
- **Callaway-Sant'Anna (2020)**: 144.6 seconds
- **Sun-Abraham (2020)**: 198.5 seconds

(using a laptop with 2.80 GHz Quad-core Intel i7 Processor and 16 GB of Ram)

Application: Banking Deregulation and the Labor Share

1970-1996: staggered introduction
of (inter-state & intra-state)
banking deregulation in US states.

— % of States that deregulated inter-state banking
-- % of States that deregulated intra-state branching
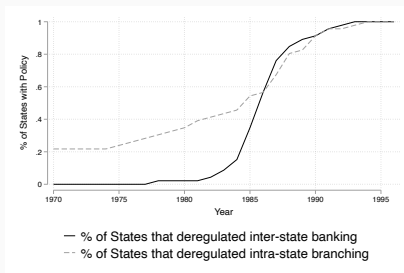
- Leblebicioglu & Weinberger (EJ, 2020) use static & event-study
  TWFE to estimate effects on the labor share.

Empirical Applications (1)

TWFE estimates

- Negative effect of *inter-state* bank deregulation (≈ -1pp).
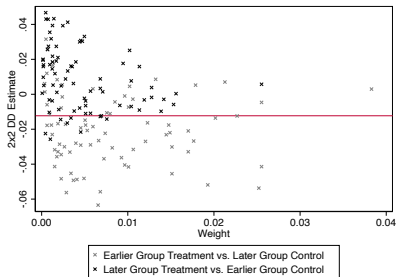- No effect of *intra-state* branching deregulation.
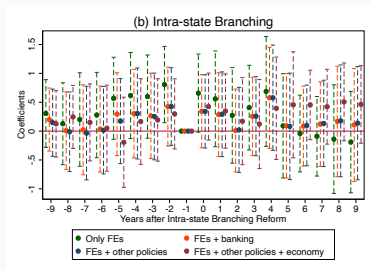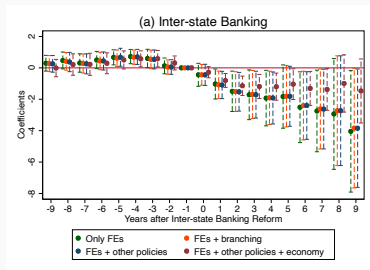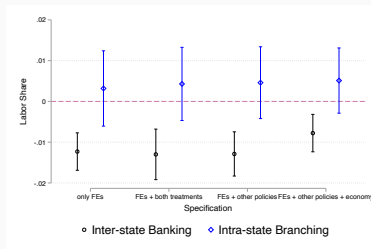
Forbidden comparisons in the TWFE specification

- TWFE uses 'forbidden' comparisons: earlier liberalizers are controls for later liberalizers.

- Goodman-Bacon (2021) decomposition to quantify their influence.

- Contribution of unclean comparisons to TWFE estimates:
    o 36% for inter-state banking deregulation;
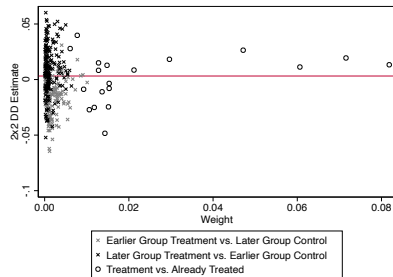   o 70% for intra-state branching deregulation.

Goodman-Bacon (2021) decomposition diagnostic for the static TWFE estimate

(a) Inter-state banking deregulation
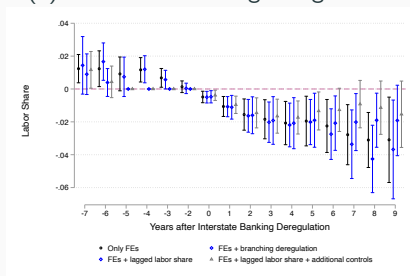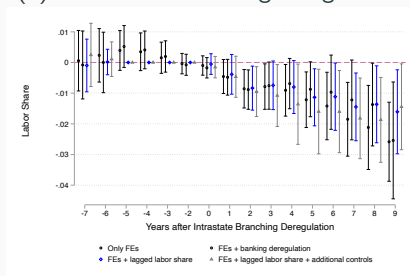
(b) Intra-state branching deregulation

Effect of banking deregulation on the labor share:

LP-DiD estimates

(a) Inter-state banking deregulation

(b) Intra-state branching deregulation

Empirical Applications (2)

Application: Democracy and economic growth

- Acemoglu, Naidu, Restrepo and Robinson (2019).

- 1960-2010 panel on 175 countries & binary measure of democracy.

- Potential for negative weights.

- Non-absorbing treatment.

- Selection based on pre-treatment GDP dynamics.

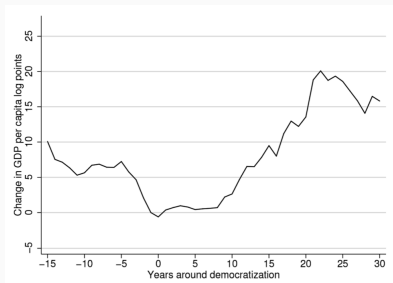### Application: Democracy and economic growth

- Acemoglu, Naidu, Restrepo and Robinson (2019).

- 1960-2010 panel on 175 countries & binary measure of democracy.

- Potential for negative weights.

- Non-absorbing treatment.

- Selection based on pre-treatment GDP dynamics.

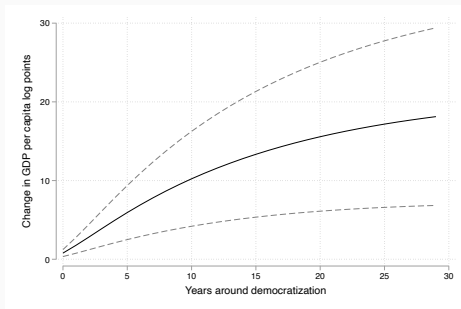GDP per capita around democratization

25

Effect of democracy on growth: dynamic panel estimates

- Dynamic fixed effects specification:

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct},$$

- Long-run effect: $\frac{\hat{\beta}}{1-\sum_{j=1}^{p} \hat{\gamma}_j} = 21pp$ (s.e. 7pp)

IRF from the dynamic panel estimates

Empirical Applications (2)

Effect of democracy on growth: LP-DiD specification

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP\ DiD} \Delta D_{ct} + \delta_t^h + \sum_{j=1}^{p} \gamma_j^h y_{c,t-j} + \epsilon_{ct}^h. \qquad (1)$$

restricting the estimation sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1; D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L \\ \text{clean controls} & D_{i,t-j} = 0 \text{ for } 0 \leq j \leq L\,. \end{cases} \qquad (2)$$

Empirical Applications (2)

Effect of democracy on growth: LP-DiD specification

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP\ DiD} \Delta D_{ct} + \delta_t^h + \sum_{j=1}^{p} \gamma_j^h y_{c,t-j} + \epsilon_{ct}^h. \qquad (1)$$

restricting the estimation sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1; D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L \\ \text{clean controls} & D_{i,t-j} = 0 \text{ for } 0 \leq j \leq L. \end{cases} \qquad (2)$$

- We set L=20 years.

- Acemoglu et al. LP analysis: a version of this, but controls defined by $D_{it} = D_{i,t-1} = 0$, and treated defined as $\Delta D_{it} = 1$

- They still include countries that transition at or before $t - 1$.

Empirical Applications (2)

Effect of democracy on growth: LP-DiD estimates
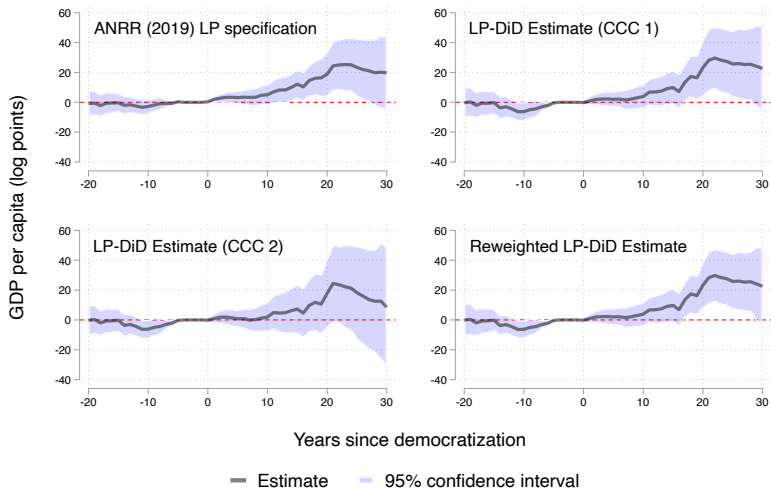
Conclusions

- LP-DiD offers a flexible overarching framework for DiD settings.

- Simpler and less computationally intensive than estimators that aggregate many group-specific averages.

- Flexibility in defining the treatment and ('clean') control units based on the setting.

- Allows matching on pre-treatment outcomes and other time-varying covariates.

Additional Slides

Identification Assumptions (baseline specification)

No anticipation

$$E\left[y_{it}(p) - y_{it}(0)\right] = 0, \text{ for all } p \text{ and } t \text{ such that } t < p.$$

Units do not respond in anticipation of a future treatment.

Parallel trends

$$E\left[y_{it}(0) - y_{i1}(0)|p_i = p\right] = E\left[y_{it}(0) - y_{i1}(0)\right],$$
$$\text{for all } t \in \{2, ..., T\} \text{ and for all } p \in \{1, ..., T, \infty\}.$$

Average trends in untreated potential outcomes do not depend on treatment status.

A1 - Other new DiD estimators

Alternative estimators: de Chaisemartin & D'Haultfoeiulle

de Chaisemartin & D'Haultfoeiulle estimator

- For a given time-horizon ℓ, it estimates the average effect of having switched in or out of treatment ℓ periods ago.
- A weighted average, across time periods t and possible values of treatment d, of 2x2 DiD estimators.
- The constituent 2x2 DiDs compare the $t - \ell - 1$ to t outcome change, in groups with a treatment equal to d at the start of the panel and whose treatment changed for the first time in $t - \ell$ (the first-time switchers) and in control groups with a treatment equal to d from period 1 to t (not-yet switchers).

Alternative estimators: Callaway-Sant'Anna

Callaway-Sant'Anna estimator

- Estimates each group specific effect at the selected time horizon.
- Take long-differences in the outcome variable, and compare each treatment group g with its control group.
- To control for covariates, re-weight observations based on outcome regression (OR), inverse-probability weighting (IPW) or doubly-robust (DR) estimation.
- Aggregate group-time effects into a single overall ATT using some weights.

Alternative estimators: Sun-Abraham

Sun-Abraham interaction-weighted estimator

- Event-study DiD specification, with leads and lags of the treatment variable.

- Includes a full set of interaction terms between relative time indicators D_{it}^k (ie, leads and lags of the treatment variable) and treatment cohort indicators $1\{G_g = g\}$ (dummies for when a unit switches into treatment).

- Then calculates a weighted average over cohorts g for each time horizon, in order to obtain a standard event-study plot.

Alternative estimators: Borusyak-Jaravel-Spiess

Borusyak-Jaravel-Spiess imputation estimator

- Estimate unit and time FEs only using untreated sample.
- Take them out from Y to form counterfactual Y'.
- Then for any treatment group, just compare Y and Y' for treated units around event time.
- Average these across events to get an average effect.