A Local Projections Approach to Difference-in-Differences

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

June 2025

[†] University of Massachusetts, Amherst; NBER; and IZA
[*] King's College London;
[‡] Federal Reserve Bank of San Francisco; University of California, Davis; and CEPR
[§] Columbia University; Bank of England; NBER; and CEPR

Outline

- How to estimate DiD with staggered treatment?
  - Recent literature shows that conventional TWFE implementations can be severely biased.
- A new regression-based framework: LP-DiD.
  - Local projections (Jordà 2005) + clean controls (CDLZ 2019).
  - Can yield convex VWATT or equally-weighted ATT.
  - Allows for covariates and non-absorbing treatment
  - `lpdid` STATA command (Busch and Girardi 2023)

Outline

- How to estimate DiD with staggered treatment?
    - Recent literature shows that conventional TWFE implementations can be severely biased.
- A new regression-based framework: LP-DiD.
    - Local projections (Jordà 2005) + clean controls (CDLZ 2019).
    - Can yield convex VWATT or equally-weighted ATT.
    - Allows for covariates and non-absorbing treatment
    - `lpdid` STATA command (Busch and Girardi 2023)
- Montecarlo simulation to assess its performance.
- Empirical applications:
    - Effect of banking deregulation on the wage share.
    - Democracy & growth

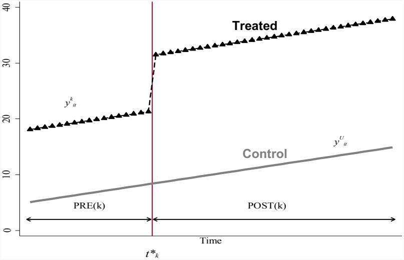
Why do we need yet another DiD estimator?

Advantages of LP-DiD:

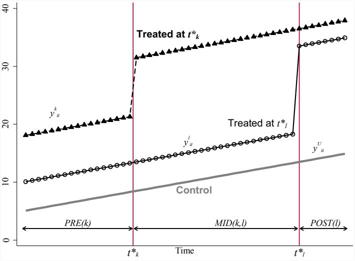
- Simple and fast to implement.
- Transparent in defining treated and control units.
- Flexible: easily accommodates different settings, weighting schemes, and target estimands.
- General: encompasses other recent DiD estimators as specific sub-cases.

Difference-in-Differences (DiD)

2x2 Setting



Staggered Setting



(Visual examples from Goodman-Bacon, 2021)

Background

The conventional (until recently) DiD estimator: TWFE

- Static TWFE

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

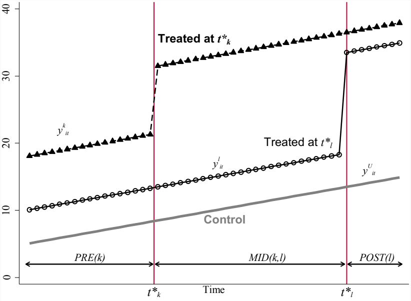
- Event-study (distributed lags) TWFE

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

- 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)

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

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)

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

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

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

  - Weights can be negative (bad!)

A LP-DiD Estimator

Baseline version

Setting & Assumptions:

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

A LP-DiD estimator

Baseline version

Estimating equation:

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

restricting the estimation sample to observations that are either

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

A LP-DiD estimator

Baseline version

Estimating equation:

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

restricting the estimation sample to observations that are either

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

Estimates are obtained from a set of ‘clean’ comparisons between newly treated units and not-yet treated ones → no negative weighting

What does LP-DiD identify?

- OLS estimation of the LP-DiD specification yields a variance-weighted average effect:

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

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

- No negative weights.

LP-DiD Estimator

What does LP-DiD identify?

- OLS estimation of the LP-DiD specification yields a variance-weighted average effect:

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

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

- No negative weights.

- Weights depend on subsample size & treatment variance:

$$\omega_{g,h}^{LP-DiD} = \frac{N_{CCS_{g,h}}[n_{gh}(1-n_{gh})]}{\sum_{g \neq 0} N_{CCS_{g,h}}[n_{gh}(1-n_{gh})]},$$

  - o $N_{CCS_{g,h}}$ = size of subsample including cohort g & its clean controls.
  - o $[n_{gh}(1-n_{gh})]$ = treatment variance in that subsample.

Alternative weighting schemes

- Variance-weighting gives more weight to more precisely estimated cohort-specific effects
- But you can apply any desired weights through weighted regression.
- For the equally-weighted ATT:
  - weighted regression with weights $= (\omega_{g,h}^{LP-DiD}/N_g)^{-1}$
  - can get the same using regression adjustment.

LP-DiD encompasses other recent DiD estimators

- Baseline OLS LP-DiD
  $\leftrightarrow$ stacked estimator (CDLZ, 2019)
  - But no need to stack the data!
- Reweighted LP-DiD for equal weights
  $\leftrightarrow$ CS estimator
- Reweighted PMD LP-DiD
  $\approx$ BJS estimator.
  - PMD means using $y_{i,t+h} - \frac{1}{k}\sum_{\tau=t-k}^{t-1} y_{i,\tau}$ as outcome

Extended settings

- Covariates
- Non-absorbing treatment
- In future work, can be extended to continuous treatment

LP-DiD with covariates

- Parallel trends conditional on **x**.
- A Regression Adjustment LP-DiD specification controlling for **x** yields the (equally-weighted) ATT.
- Example of STATA implementation:

```
teffects ra (Dhy i.time x1 x2) (dtreat)
if D.treat==1 | Fh.treat==0, atet vce(cluster unit)
```

- If effects are independent of covariates, adding covariates directly to an OLS LP-DiD specification yields convex VWATT with same weights as in baseline.

LP-DiD with non-absorbing treatment

- Tackled by adapting the clean control condition.

LP-DiD with non-absorbing treatment

- Tackled by adapting the clean control condition.
- Effect of entering treatment for the first time:

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } j \geq 1)\,, \\ \text{or clean control} & D_{i,t-j} = 0 \text{ for } j \geq -h\,. \end{cases}$$

.

LP-DiD with non-absorbing treatment

- Tackled by adapting the clean control condition.
- Effect of entering treatment for the first time:

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } j \geq 1)\,, \\ \text{or clean control} & D_{i,t-j} = 0 \text{ for } j \geq -h\,. \end{cases}$$

.

- Average effect of a treatment event:
  - o Assume treatment effects *stabilize* after L periods. Then use:

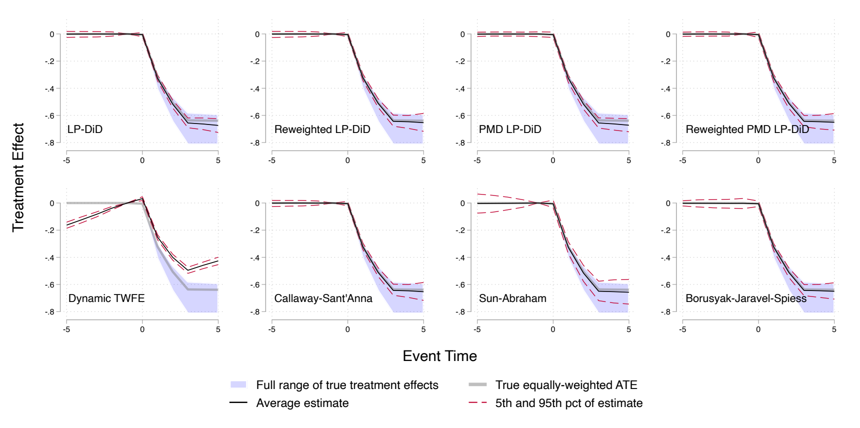
$$\begin{cases} \text{treatment} & (\Delta D_{it} = 1) \quad \& \quad (\Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L; j \neq 0) \\ \text{clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L \end{cases}$$

Simulation

- Calibrated on empirical application: banking deregulation (treatment) and wage share (outcome) in US states
- Simulate wage share data (y) for 46 states over 26 years.
  - DGP: $y_{it}(0) = \lambda_i \gamma_t e_{it}$, with $e_{it} = (1-\rho)\epsilon_{it} + \rho e_{i,t-1}$
  - $\lambda_i, \gamma_t, \epsilon_{it} \sim Beta$, parameters estimated from wage share data.
- Same treatment rollout as banking deregulation laws.
- TE grows in time for 4 years, is stronger for early adopters.
- Given multiplicative DGP, we estimate a log specification.

Simulation

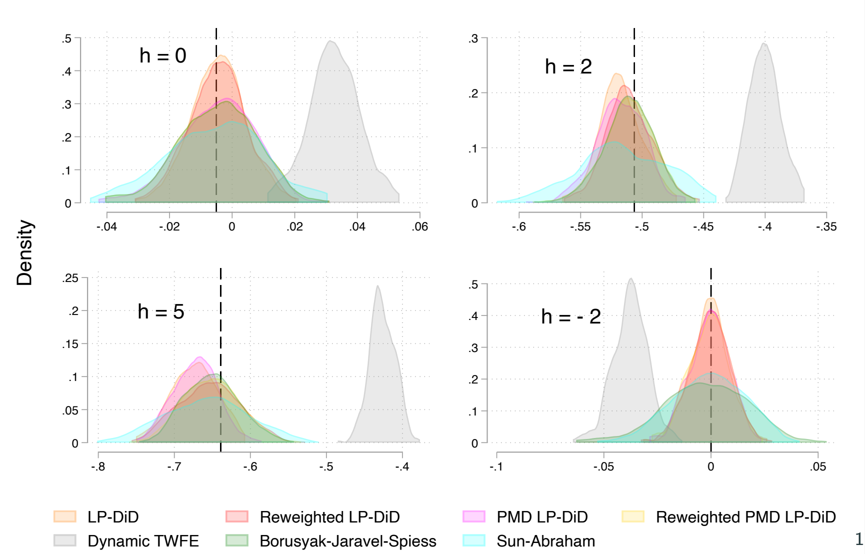
True effect path and estimates from 200 replications



- LP-DiD performs well and similarly to other recent estimators;
- Variance-weighted LP-DiD has the lowest RMSE of all estimators at time horizons where treatment effect heterogeneity is less large.

Simulation

Distribution of estimates from 200 replications.



Computational speed

Estimating the treatment effect path in a single repetition of the simulation (seconds):

Panel size	Dynamic TWFE	LP-DiD	PMD LP-DiD	Rw LP-DiD	Rw PMD LP-DiD	CS	SA	BJS
N=46; T=27; 13 events	.24	.12	.13	.20	.19	4.46	1.09	.24
N=184; T=54; 26 events	.22	.16	.19	.26	.29	137.5	105.5	.54

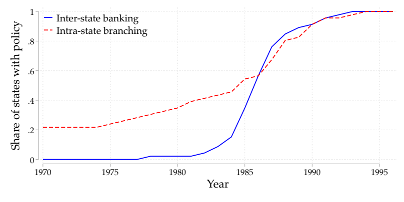
Notes: Computation times in a single repetition of the simulated datasets described in Section 5, measured in seconds. Recorded on a laptop with M2 Apple Chip processor and 8 GB of RAM, using the STATA software. Rw = reweighted (see Sec 3.3); PMD = pre-mean-differenced (see Sec 3.4); CS = Callaway and Sant'Anna, 2020; SA = Sun and Abraham, 2020; BJS = Borusyak, Jaravel, and Spiess (2024).

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

Banking Deregulation and the Labor Share

1970-1996: US states deregulate banking in a staggered fashion.

- Inter-state banking deregulation
- Intra-state branching deregulation

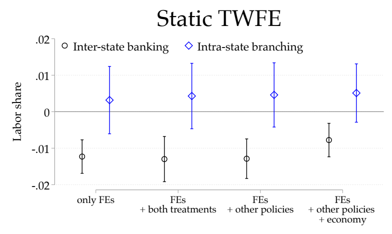


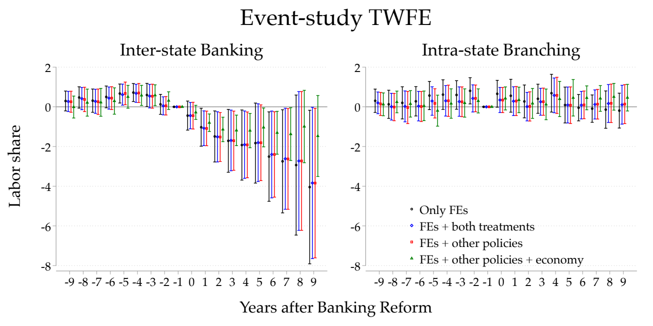
- Leblebicioglu & Weinberger (EJ, 2020) use static & event-study TWFE to estimate effects on the labor share.

Empirical Application

TWFE estimates

- Negative effect of *inter-state* bank deregulation (≈ -1pp).
- No effect of *intra-state* branching deregulation.





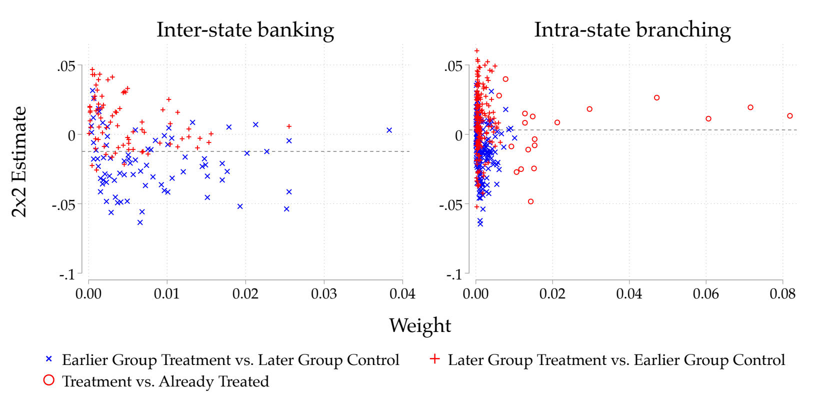
Empirical Applications (1)

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:
  - 36% for inter-state banking deregulation;
  - 70% for intra-state branching deregulation.

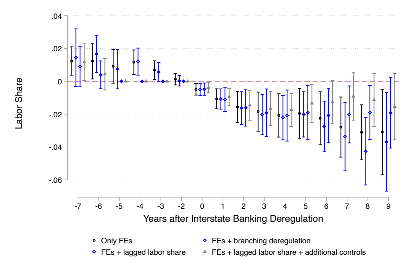
Empirical Application

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

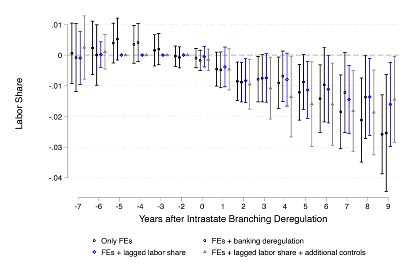


Effect of banking deregulation on the labor share: LP-DiD estimates

(a) Inter-state banking deregulation



(b) Intra-state branching deregulation



- Same conclusion re: *inter-state banking* deregulation
- But dramatically different re: *intra-state branching* deregulation, where unclean comparisons had large influence in TWFE.
- Also *intra-state branching* deregulation reduces the labor share!

lpdid STATA command (Busch and Girardi, 2023)

```
ssc install lpdid, replace
```

```
. use http://fmwww.bc.edu/repec/bocode/l/lpdidtestdata1.dta

. lpdid Y, time(time) unit(unit) treat(treat) pre(5) post(10)
lpdid Y, unit(unit) time(time) treat(treat) pre_window(5) post_window(10)
```

LP-DiD Event Study Estimates

E-time	Coeffic~t	SE	t	P>|t|	[95% co~.	interval]	obs
pre5	-.0425659	.9483544	-.04	.9642	-1.902432	1.817301	40662
pre4	.6403343	.9588844	.67	.5043	-1.240183	2.520852	42662
pre3	1.079831	.8967272	1.2	.2287	-.6787866	2.838449	44662
pre2	1.45865	.8264465	1.76	.0777	-.1621368	3.079437	46662
pre1	0	.	.	.	.	.	.
tau0	3.640153	.7948942	4.58	0	2.081246	5.199061	48662
tau1	7.11248	.9093428	7.82	0	5.329121	8.895838	46662
tau2	9.749811	.9573893	10.18	0	7.872225	11.6274	44662
tau3	14.68331	.9699534	15.14	0	12.78109	16.58554	42662
tau4	19.87852	1.013118	19.62	0	17.89164	21.8654	40662
tau5	28.50038	1.014339	28.1	0	26.51111	30.48965	38662
tau6	34.7144	1.021342	33.99	0	32.7114	36.71741	36662
tau7	42.87508	1.034415	41.45	0	40.84643	44.90372	34662
tau8	53.21209	1.094259	48.63	0	51.06608	55.35809	32662
tau9	62.63418	1.108112	56.52	0	60.46101	64.80736	30662
tau10	72.63583	1.193887	60.84	0	70.29444	74.97723	28709

LP-DiD Pooled Estimates

	Coeffic~t	SE	t	P>|t|	[95% co~.	interval]	obs
Pre	.7840624	.7374264	1.06	.2878	-.6621424	2.230267	40662
Post	31.79438	.7559724	42.06	0	30.3118	33.27696	28709

For “manual” implementation of LP-DiD (which is quite easy), see example codes at `https://github.com/danielegirardi/lpdid`

Conclusions

Khoa Vu
@KhoaVuUmn

Arin Dube @arindube · May 1

🚨 Difference-in-differences working paper alert 🚨

Our Local-Projections DiD offers a unified approach that encompasses many popular alternatives as specific instances; allows for extensions; and does it all using an OLS regression.

nber.org/papers/w31184

🧵

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.

Obtaining an equally-weighted ATT

- Baseline weights $\omega_{g,h}^{LP-DiD}$ depend on cohort size & treatment variance.
- But you can apply any desired weights using weighted regression.
- Equally-weighted ATE: Reweight by

$$1/(\omega_{g,h}^{LP-DiD}/N_g).$$

- $\omega_{g,h}^{LP-DiD}$ easy to compute from 'residualized' treatment indicator $\Delta\tilde{D}$.
- Can also use regression adjustment.

Empirical Applications (2)

Application: Democracy and economic growth

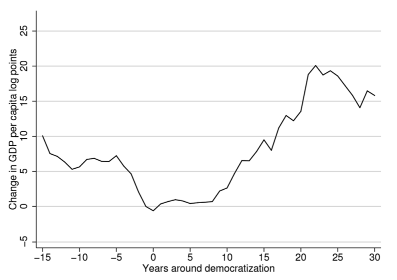
- Acemoglu, Naidu, Restrepo and Robinson (2019).
- 1960-2010 panel on 175 countries & binary measure of democracy.
- Potential for negative weights.
- Non-absorbing treatment.
- Selection based on pre-treatment GDP dynamics.

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.

GDP per capita around democratization



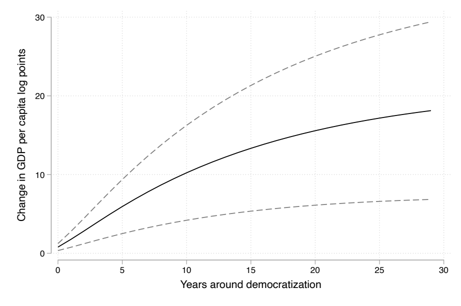
Effect of democracy on growth: dynamic panel estimates

- 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

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

restricting the estimation sample to:

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

Empirical Applications (2)

Effect of democracy on growth: LP-DiD specification

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

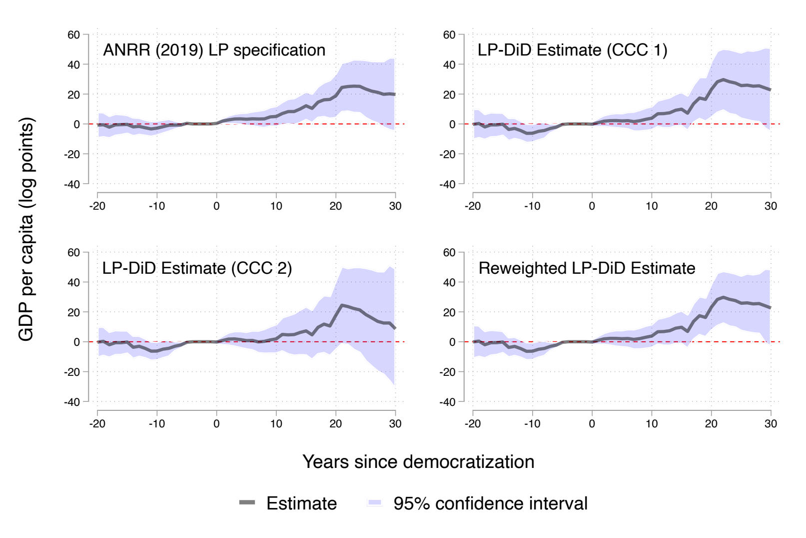
restricting the estimation sample to:

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

- We set L=20 years.
- Acemoglu et al. LP analysis: a version of this, but (implicitly) L=1.

Empirical Applications (2)

Effect of democracy on growth: LP-DiD estimates



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.