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

Arindrajit Dube* Daniele Girardi** Óscar Jordà*** Alan M. Taylor****

July 12 2022

* 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
Research question

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

- Recent literature shows that the TWFE implementation of DiD (static or distributed lags) can be severely biased.
- Estimate is an average with possibly negative weights. Bad!

A new regression-based framework: LP-DiD.

- Basically, local projections (Jordà 2005) + clean controls (Cengiz et al. 2019).
- We derive weights placed on each treatment event.
- No negative weights. Good!
- Simple reweighting to recover ATT.

Simulation evidence to assess its performance.

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

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

- Recent literature shows that the TWFE implementation of DiD (static or distributed lags) can be severely biased.
  - Estimate is an average with possibly negative weights. **Bad!**
- A new regression-based framework: **LP-DiD**.
- We derive weights placed on each treatment event
  - No negative weights. **Good!**
  - Simple reweighting to recover ATT
Research question

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

- Recent literature shows that the TWFE implementation of DiD (static or distributed lags) can be severely biased.
  - Estimate is an average with possibly negative weights. Bad!

- A new regression-based framework: LP-DiD.


- We derive weights placed on each treatment event
  - No negative weights. Good!
  - Simple reweighting to recover ATT

- 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.
- With exogenous treatment, the estimates are identical to the increasingly popular "stacked regression" approach of Cengiz et al. (2019), but easier to implement, and to generalize.
- 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
  \[ y_{it} = \alpha_i + \delta_t + \sum_{m=-Q}^{M} \beta^{TWFE}_m D_{it-m} + \epsilon_{it} \]

- \textbf{OK} in the 2x2 setting, or when treatment occurs at the same time.
- \textbf{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)
  \[
  \lim_{N \to \infty} \hat{\beta}^{TWFE} = VWATT + VWCT - \Delta ATT
  \]
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)

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

- Weights can be negative!
A Local Projections Diff-in-Diff Estimator (LP-DiD)
No Covariates, Outcome Lags

\[ y_{i,t+k} - y_{i,t-1} = \beta^k_{LP-DiD} \Delta D_{it} \quad \} \text{treatment indicator} \]
\[ + \delta^k_t \quad \} \text{time effects} \]
\[ + e^k_{it} ; \quad \text{for } k = 0, \ldots, K . \]

restricting the sample to observations that are either:

\[
\begin{cases}
\text{treatment} & \Delta D_{it} = 1 , \\
\text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k .
\end{cases}
\]
A Local Projections Diff-in-Diff Estimator (LP-DiD)
No Covariates, Outcome Lags

\[ y_{i,t+k} - y_{i,t-1} = \beta_{k LP-DiD} \Delta D_{it} \] treatment indicator
\[ + \delta_k t \] time effects
\[ + e_{it} ^k \]

restricting the sample to observations that are either:

\[ \begin{cases} 
\text{treatment} & \Delta D_{it} = 1, \\
\text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k. 
\end{cases} \]

Key advantage of LP over distributed lags TWFE formulation of DiD: the differencing is in outcomes, and not treatments. Allows for easy use of the clean control sample restriction.
A Local Projections Diff-in-Diff Estimator (LP-DiD)

\[ y_{i,t+k} - y_{i,t-1} = \beta_k \, LP-DiD \, \Delta D_{it} \]
\[ + \sum_{p=1}^{P} \gamma^k_{0,p} \Delta y_{i,t-p} \]
\[ + \sum_{m=1}^{M} \sum_{p=0}^{P} \gamma^k_{m,p} \Delta x_{m,i,t-p} \]
\[ + \delta^k_t \]
\[ + e^k_{it} ; \]

restricting the sample to observations that are either:

\[ \left\{ \begin{array}{l}
\text{treatment} \quad \Delta D_{it} = 1, \\
\text{clean control} \quad \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k .
\end{array} \right. \]
An equivalent specification to implement LP-DiD

• Instead of restricting the sample, we can use interaction terms to rule out unclean controls.

\[ y_{i,t+k} - y_{i,t-1} = \beta^k \text{LP-DiD} \Delta D_{it} + \theta^k \text{UC}_{i,t} \]
\[ + \sum_{p=1}^{P} \gamma_{0,p}^k (1 + \rho_{0,p}^k \text{UC}_{i,t}) \Delta y_{i,t-p} \]
\[ + \sum_{m=1}^{M} \sum_{p=0}^{P} \gamma_{m,p}^k (1 + \rho_{m,p}^k \text{UC}_{i,t}) \Delta x_{m,i,t-p} \]
\[ + \delta_t^k (1 + \phi_t^k \text{UC}_{i,t}) + e_{it}^k ; \]

} \text{treatment indicator} \]
{ \text{UC indicator} \]
{ \text{outcome lags } \times \text{UC} \]
{ \text{covariates } \times \text{UC} \]
{ \text{time effects } \times \text{UC} \]

\text{for } k = 0, \ldots, K .

• UC_{it} = 1 \text{ if previously treated.}

• With absorbing treatment, UC_{it} = \sum_{j=-H(j \neq 0)}^{k} \Delta D_{i,t+j} \]
What does LP-DiD identify?

- A variance-weighted average effect:

\[
E(\hat{\beta}^k_{LP-DiD}) = \sum_{g \neq 0} \omega^\text{LP-DiD}_{g,k} \tau_g(k)
\]

- \( \tau_g(k) = k\)-periods forward ATT for treatment-cohort \( g \).
What does LP-DiD identify?

- A variance-weighted average effect:

\[
E(\hat{\beta}_k^{LP-DiD}) = \sum_{g \neq 0} \omega_{g,k}^{LP-DiD} \tau_g(k)
\]

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

- Weights are always positive and depend on subsample size & treatment variance

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

where

- \( CCS_{g,k} \) is a subsample including group \( g \) and its 'clean controls'.
- \( n_{g,k} = N_g/N_{CCS_{g,k}} \) is the share of treated units in \( CCS_{g,k} \).
- \( n_{c,g,k} = N_{c,g,k}/N_{CCS_{g,k}} \) is the share of control units in \( CCS_{g,k} \).
LP-DiD as a ‘swiss knife’
LP-DiD Estimator

LP-DiD as a ‘swiss knife’

1. Flexibility in defining the treatment & control units

Some examples:

- absorbing treatment: can use $D_{i,t+k} = 0$ to define clean controls.
- repeated treatment: select an appropriate time-window $H$.
- continuous treatment: can define clean controls as ‘stayers’ or ‘quasi-stayers’ (as in deChaisemartin et al., 2022)
LP-DiD as a ‘swiss knife’

1. **Flexibility in defining the treatment & control units**
   *Some examples:*
   - absorbing treatment: can use $D_{i,t+k} = 0$ to define clean controls.
   - repeated treatment: select an appropriate time-window $H$.
   - continuous treatment: can define clean controls as ‘stayers’ or ‘quasi-stayers’ (as in deChaisemartin et al., 2022)

2. **Flexibility in choosing a weighting scheme**
   - Can apply any desired weights through weighted regression.
   - Equally-weighted ATT: reweight observations by $1/(\omega_g^{LP-DiD} / N_g)$.
   - $\omega_g^{LP-DiD}$ easy to compute empirically from ‘residualized’ treatment indicator.
   - Can use p-score reweighting instead of regression adjustment for covariates and lagged outcomes
Uses (a Very Simple) Regression to Solve the TWFE Regression’s Problems

Peter Hull @instrumenthull · Jun 6
For the record, I still love regression

apoorva.lal @Apoorva_Lal · Jun 6
Replying to @instrumenthull
Always nice when the fixes are 'just run a different regression'
Simulation Evidence

Simulation

- $N=500$ units; $T=50$ time periods.
- 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 Evidence

Simulation

- N=500 units; T=50 time periods.
- 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 \( \tau = 11, 13, 15 \ldots, 27 \).
Simulation Evidence

Simulation

- N=500 units; T=50 time periods.
- 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 \( \tau = 11, 13, 15 \ldots, 27 \).
  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 Evidence

Simulation results – exogenous treatment scenario

Average estimates and 95% and 5% percentiles from 200 replications.
Simulation results – endogenous treatment scenario

Average estimates and 95% and 5% percentiles from 200 replications.
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-Abramah (2020)**: 198.5 seconds

(using a laptop with 2.80 GHz Quad-core Intel i7 Processor and 16 GB of Ram)
Seeking leads on estimating Callaway/Sant'Anna doubly robust DID estimates and SEs when N is somewhat large (>1mil). The SE bootstrapping has been running on a pretty powerful computer for about 48 hours now and I am not feeling optimistic.
Application: Banking Deregulation and the Labor Share


- Leblebicioglu & Weinberger (2020) use static & event-study TWFE to estimate effects on the labor share.
- Negative effect of *inter-state* banking deregulation ($\approx -1$ p.p.).
- No effect of *intra-state* branching deregulation.
Effect of banking deregulation on the labor share: TWFE estimates
Forbidden comparisons in the TWFE specification

- TWFE uses ‘forbidden’ comparisons: earlier liberalizers are controls for later liberalizers.
- We employ Goodman-Bacon (2021) decomposition to assess their influence.
- Contribution of unclean comparisons to TWFE estimates:
  - 36% for inter-state banking deregulation;
  - 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

- LP-DiD avoids unclean comparisons & allows controlling for $y$ lags.
- Negative effect of inter-state branching deregulation is confirmed.
- But also intra-state branching deregulation has negative effect.
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: \( \frac{\hat{\beta}}{1-\sum_{j=1}^{p} \hat{\gamma}_j} = 21 \text{pp (s.e. 7pp)} \)

IRF from the dynamic panel estimates
Effect of democracy on growth: LP-DiD specification

\[ y_{c,t+k} - y_{c,t-1} = \beta^k \Delta D_{ct} + \delta^k + \sum_{j=1}^{p} \gamma_j^k y_{c,t-j} + \epsilon_{ct}. \]

restricting the sample to:

\[
\begin{cases}
\text{democratizations} & D_{it} = 1, D_{i,t-1} = 0 \\
\text{clean controls} & D_{i,t+h} = 0 \text{ for } -H \leq h \leq k.
\end{cases}
\]
Effect of democracy on growth: LP-DiD specification

\[ y_{c,t+k} - y_{c,t-1} = \beta_k^k \Delta D_{ct} + \delta_k^t + \sum_{j=1}^{p} \gamma_j^k y_{c,t-j} + \epsilon_{ct}. \]

restricting the sample to:

\[
\begin{cases}
\text{democratizations} & D_{it} = 1, D_{i,t-1} = 0 \\
\text{clean controls} & D_{i,t+h} = 0 \text{ for } -H \leq h \leq k.
\end{cases}
\]

- Acemoglu et al. LP analysis: a version of this, but controls defined by \( D_{it} = D_{i,t-1} = 0 \).

- They still include countries that slide into autocracy at or before \( t - 1 \), and countries that transition between \( t \) and \( t + k \).
Effect of democracy on growth: LP-DiD estimates

- ANRR (2019) LP specification
- LP-DiD with $H=1$
- LP-DiD with $H=20$
- LP-DiD with $H=40$
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 control units based on the setting.
- Allows matching on pre-treatment outcomes and other time-varying covariates.
Additional Slides
A1 - Other new DiD estimators
de Chaisemartin & D’Haultfoeiuille estimator

- For a given time-horizon $\ell$, it estimates the average effect of having switched in or out of treatment $\ell$ 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).
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.
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 $D_{it}^k$ (i.e., 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.
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.