

NBER WORKING PAPER SERIES

A LOCAL PROJECTIONS APPROACH TO DIFFERENCE-IN-DIFFERENCES

Arindrajit Dube  
Daniele Girardi  
Òscar Jordà  
Alan M. Taylor

Working Paper 31184  
<http://www.nber.org/papers/w31184>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 2023, revised May 2024

This paper previously circulated as "A Local Projections Approach to Difference-in-Differences Event Studies." We thank Alexander Busch, Colin Cameron, Sophie Hatte, Thomas Høgholm Jørgensen, and seminar participants at UC Davis, University of Milan La Statale, 2022 NBER Summer Institute, 2023 Annual AEA Meetings, 2023 EEA-ESEM Annual Congress, PolMeth Europe 2023, Asia School of Business (ASB), Sant'Anna School of Advanced Studies, UMass Amherst, and University of Pisa for helpful comments and exchanges. The `lpdid` STATA command, which implements the LP-DiD estimators discussed in this paper, can be found at <https://econpapers.repec.org/software/bocbocode/S459273.htm>. Additional STATA example files are available at <https://github.com/danielegirardi/lpdid/>. The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco, the Federal Reserve System, or the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w31184>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Arindrajit Dube, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Local Projections Approach to Difference-in-Differences  
Arindrajit Dube, Daniele Girardi, Òscar Jordà, and Alan M. Taylor  
NBER Working Paper No. 31184  
April 2023, revised May 2024  
JEL No. C1,C23,C5

### **ABSTRACT**

We propose a local projection (LP) based difference-in-differences approach that subsumes many of the recent solutions proposed in the literature to address possible biases arising from negative weighting. We combine LPs with a flexible ‘clean control’ condition to define appropriate sets of treated and control units. Our proposed LP-DiD estimator can be implemented with various weighting and normalization schemes for different target estimands, can be extended to include covariates or accommodate non-absorbing treatment, and is simple and fast to implement. A simulation and two empirical applications demonstrate that the LP-DiD estimator performs well in common applied settings.

Arindrajit Dube  
Department of Economics  
University of Massachusetts  
Crotty Hall  
412 N. Pleasant Street  
Amherst, MA 01002  
and NBER  
adube@econs.umass.edu

Daniele Girardi  
Department of Economics  
University of Massachusetts  
Crotty Hall  
412 N. Pleasant Street  
Amherst, MA 01002  
dgirardi@umass.edu

Òscar Jordà  
Economic Research, MS 1130  
Federal Reserve Bank of San Francisco  
San Francisco, CA 94105  
and University of California, Davis and also CEPR  
oscar.jorda@gmail.com

Alan M. Taylor  
Columbia University  
1313 International Affairs  
420 West 118th Street  
New York, NY 10027  
and CEPR  
and also NBER  
amt2314@columbia.edu

A STATA example files illustrating the implementation of the LP-DiD estimators discussed in this paper can be

found here: is available at  
[https://github.com/  
danielegirardi/lpdid/](https://github.com/danielegirardi/lpdid/)

# 1 Introduction

Difference-in-differences (DiD) designs are commonly used to estimate treatment effects with observational data. In canonical form, with two periods and two groups (of which one is treated), and under suitable assumptions (no anticipation and parallel trends), the DiD comparison is an unbiased estimator of the average treatment effect on the treated.

Yet, as the scale and scope of DiD applications have widened over time and expanded into multi-period settings, its underpinnings have been stretched and doubts about the generality of its underlying assumptions have proliferated, as highlighted in many notable recent studies. The central matter of concern has been the appropriate implementation of DiD in an expanded set of situations where the dates of treatment can differ across groups (i.e., are *staggered*), and the treatment effects can differ across groups (i.e., are *heterogeneous*) and can occur gradually over time (i.e., are *dynamic*).<sup>1</sup>

In this more general setting, the two-way fixed effects (TWFE) estimate has been shown to be a weighted average of heterogeneous group-specific treatment effects where the weights may be negative, leading to potentially severe bias.<sup>2</sup> As a result of this ‘negative weights’ bias, the TWFE estimate could even lie outside of the range of group-specific treatment effects. This problem has led to the search for, and proliferation of, several alternative estimators.<sup>3</sup>

In this paper we propose a framework for DiD estimation that exploits an important link to *local projections* (LPs), a statistical technique introduced in a time-series context in [Jordà \(2005\)](#). The LP approach has been widely used to estimate dynamic impulse responses in time series or panel data.<sup>4</sup> By its very design, and as used in applied macroeconomics, the LP approach was set up to estimate *dynamic* average treatment responses. As we show in this paper, an under-appreciated feature of the LP framework is that it is straightforward to limit the set of permissible comparisons based on a desired

---

<sup>1</sup>See for example [de Chaisemartin and D’Haultfœuille \(2020\)](#); [Sun and Abraham \(2020\)](#); [Goodman-Bacon \(2021\)](#); [Borusyak, Jaravel, and Spiess \(2021\)](#); [Gardner \(2021\)](#); [Baker, Larcker, and Wang \(2022\)](#). Surveys of this ‘new DiD’ literature are provided in [Roth, Sant’Anna, Bilinski, and Poe \(2023\)](#) and [de Chaisemartin and D’Haultfœuille \(2022\)](#).

<sup>2</sup>See the literature cited in footnote 1 above. In Section 2 of this paper we show how these problems would affect a ‘naive’ LP implementation (similar to the negative results in [de Chaisemartin and D’Haultfœuille \(2022, pp.33-34\)](#) but considering a different specification of the LP regression).

<sup>3</sup>Alternative DiD estimators, that do not suffer from the shortcomings of TWFE in the staggered setting, are proposed, for example, in [Sun and Abraham \(2020\)](#); [Callaway and Sant’Anna \(2020\)](#); [Borusyak, Jaravel, and Spiess \(2021\)](#); [de Chaisemartin and D’Haultfœuille \(2020\)](#); [de Chaisemartin, D’Haultfœuille, Pasquier, and Vazquez-Bare \(2022\)](#); [Gardner \(2021\)](#); [Wooldridge \(2021\)](#).

<sup>4</sup>A recent innovative contribution is [Viviano and Bradic \(2023\)](#). Like us, they employ a local projections framework in a panel data setting. Unlike us, however, their focus is not on the DiD design with its parallel trends assumptions, but on a different dynamic treatment design that relies on sequential ignorability assumptions.

criterion, such as past treatment history. Leveraging this feature, we further develop these LP techniques from macroeconometrics in conjunction with the potential outcomes approach of microeconometrics, to derive a simple, general and flexible regression-based framework for DiD.

Our proposed LP-DiD approach employs local projections to estimate dynamic effects alongside a flexible ‘clean control’ condition in the spirit of [Cengiz, Dube, Lindner, and Zipperer \(2019\)](#) to avoid the bias that can plague fixed-effects estimators when treatment adoption is staggered. Intuitively, the bias of fixed-effects estimators arises because previously treated units, which might still be experiencing lagged time-varying and heterogeneous treatment effects, are implicitly used as controls for newly treated units. The LP-DiD clean control condition avoids this bias by restricting the estimation sample so that ‘unclean’ observations, whose outcome dynamics are still potentially influenced by a previous change in treatment status, are not part of the control group.

Under the usual DiD assumptions, LP-DiD unbiasedly estimates a convex weighted average of potentially heterogeneous cohort-specific treatment effects. We explicitly characterize the weights assigned to each cohort-specific effect, and show that they are always positive and depend on treatment variance and subsample size. As we will explain, however, it is easy to implement a different weighting scheme with our LP-DiD estimator – including an equally-weighted average effect or any other desired scheme.

We also show how the LP-DiD approach can accommodate extensions to the basic case, such as the inclusion of covariates or non-absorbing treatment, and we suggest how future research could extend it to the case of continuous treatment.

Importantly, we show that many of the recent DiD estimators can be reproduced as specific sub-cases of our general approach based on either weights assigned to particular treatment events, or the choice of a base period for constructing the local projection. In a baseline version without covariates and with binary absorbing treatment, and using not-yet-treated observations as controls, the LP-DiD estimate is identical to the estimate from a stacked regression approach as implemented in [Cengiz, Dube, Lindner, and Zipperer \(2019\)](#), while the re-weighted LP-DiD regression that recovers an equally-weighted ATT is numerically equivalent to the estimator proposed by [Callaway and Sant’Anna \(2020\)](#). Moreover, we will show that yet another version of LP-DiD, reweighted and with an alternative pre-treatment base period, is very close to the [Borusyak, Jaravel, and Spiess \(2021\)](#) imputation estimator. However, the LP-DiD implementation is simpler and computationally faster, and can be more easily generalized to include control variables and deal with non-absorbing treatment or continuous treatment.

Evidence from a Monte Carlo simulation suggests that LP-DiD performs well in

staggered DiD settings, also in comparison with other recently proposed estimators. We consider a binary staggered treatment with dynamic and heterogeneous effects, where the parallel trends assumption holds but conventional TWFE performs poorly because of negative weights bias. We show that LP-DiD performs similarly to the [Sun and Abraham \(2020\)](#), [Callaway and Sant’Anna \(2020\)](#) and [Borusyak, Jaravel, and Spiess \(2021\)](#) estimators, while being computationally simpler and faster.

Our two empirical applications employ LP-DiD to estimate the impact of banking deregulation on the labor share (replicating and extending [Leblebicioğlu and Weinberger 2020](#)) and the effect of democratization on economic growth (replicating and extending [Acemoglu, Naidu, Restrepo, and Robinson 2019](#)). These are two important empirical settings where conventional estimates are potentially subject to bias because of previously treated units being effectively used as controls. These applications demonstrate the applicability and simplicity of implementation of the LP-DiD approach both under absorbing and under non-absorbing treatment, and both when the parallel trends assumption holds unconditionally, and when it holds only conditional on pre-determined covariates.

Our proposed LP-DiD estimator is, in our view, a useful addition to the growing toolkit of DiD techniques which are being developed in reaction to the recently uncovered shortcomings of conventional TWFE implementations. First, the LP-DiD framework is especially simple and computationally fast to implement. Its regression-based formulation allows estimation and inference using well-understood standard methods.<sup>5</sup> Second, the clean control condition employed by LP-DiD defines the appropriate set of treated and control observations in a way that is transparent and therefore easy to understand, communicate, and evaluate. Third, the LP-DiD framework is especially flexible: it is easy to flexibly adapt the clean control condition to specific empirical settings; to implement different weighting schemes; to employ alternative pre-treatment base periods; and to pool estimates over a desired pre- or post-treatment horizon.

The rest of this paper is organized as follows. In sections 2 to 4 we draw a connection between DiD designs and the LP estimator and present our proposed LP-DiD specification. In Section 5 we use a simulation to assess the performance of our LP-DiD approach, with comparisons to other methods in the recent literature. In Section 6 we apply the LP-DiD estimator in two empirical applications. Section 7 concludes.

---

<sup>5</sup>Moreover, in estimating pre-treatment trends, the LP-DiD approach does not suffer from the potential difficulties of interpretation highlighted in [Roth \(2024\)](#), since pre- and post-treatment coefficients are estimated symmetrically.

## 2 Local Projections and Difference-in-Differences

In this section we clarify the connection between difference-in-differences (DiD) designs and the local projection (LP) method. While we start from simpler settings for the sake of clarity (Sections 2.2 to 2.4), the core of this section is the discussion of the case of binary staggered treatment with dynamic and heterogeneous treatment effects (Section 2.5). In this setting, ‘negative weights’ bias analogous to that incurred by conventional two-way fixed effects regression would apply to a naive LP implementation. Section 3 will then show how a properly specified LP approach (which we call LP-DiD) successfully addresses these problems and provides a simple, general and flexible regression-based framework for DiD.

### 2.1 General setup and notation

We consider the following general setup. An outcome  $y_{it}$  is observed for  $i = 1, \dots, N$  units over  $t = 1, \dots, T$  time periods. Units can receive a binary treatment, denoted by  $D_{it} \in \{0, 1\}$ . For now, treatment is permanent (or *absorbing*), so we have  $D_{is} \leq D_{it}$  for  $s < t$ . We let  $p_i$  denote the period in which unit  $i$  enters treatment for the first time, with the convention  $p_i = \infty$  if unit  $i$  is never treated during the observed sample.

Define groups (or treatment cohorts)  $g \in \{0, 1, \dots, G\}$  as exhaustive, mutually exclusive sets of units. Groups are defined so that all units within a group enter treatment at the same time, and two units belonging to different groups enter treatment at different times. Group  $g = 0$  is the never-treated group (i.e., the set of units with  $p_i = \infty$ ). We denote the time period in which group  $g$  enters treatment as  $p_g$ .

Using the potential outcomes framework (Rubin, 1974), we let  $y_{it}(0)$  denote the potential outcome that unit  $i$  would experience at time  $t$  if they were to remain untreated throughout the whole sample period (that is, if  $p_i = \infty$ ). We let  $y_{it}(p)$  denote the outcome for unit  $i$  at time  $t$ , if unit  $i$  were to enter treatment at time  $p \neq \infty$ . Observed outcomes can then be written as  $y_{it} = y_{it}(0) + \sum_{p=1}^T (y_{it}(p) - y_{it}(0)) \times \mathbf{1}\{p_i = p\}$ .<sup>6</sup>

Define the (unit- and time-specific) treatment effect at time  $t$  for unit  $i$  which enters treatment at time  $p_i \neq \infty$  as

$$\tau_{it} = y_{it}(p_i) - y_{it}(0).$$

We are generally interested in some average of these treatment effects across units. To this end, we define the (group-specific and dynamic) average treatment effect on the treated

---

<sup>6</sup>Similar notation is used, for example, in Callaway and Sant’Anna (2020) and Sun and Abraham (2020).

(ATT) at time horizon  $h$  for group  $g$  which enters treatment at time  $p_g$  as

$$\tau_h^g = E \left[ y_{i,p_g+h}(p_g) - y_{i,p_g+h}(0) | p_i = p_g \right]. \quad (1)$$

In other words,  $\tau_h^g$  represents the average dynamic effect,  $h$  periods after entering into treatment, for all units belonging to a group  $g$  that enters treatment at time  $p_g$ .<sup>7</sup>

Throughout the following discussion, we will make use of the assumptions of parallel trends and no anticipation, the two essential assumptions that underpin the DiD approach.

**Assumption 1. No anticipation**

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

This ensures that units do not respond now in anticipation of a future treatment.

**Assumption 2. Parallel trends**

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

This assumption ensures that, had treated units been left counterfactually untreated, they would have evolved over time in the same manner as the control units have.

It is convenient for the exposition of our methods to be more specific and assume a simple data-generating process (DGP) for untreated potential outcomes, which respects the parallel trends assumption. Following the recent DiD literature, we will also assume

$$E[y_{it}(0)] = \alpha_i + \delta_t, \quad (2)$$

where  $\alpha_i$  is a unit-specific fixed effect, and  $\delta_t$  is a time-specific effect common to all units.

Finally, let us define three regression specifications of interest, which can be estimated in our panel of  $N$  units and  $T$  time periods, or in some subset of it: static two-way fixed-effects (static TWFE); event-study two-way fixed-effects (event study TWFE); and local projections (LP).

**Specification 1. Static two-way fixed-effects regression (static TWFE)**

$$y_{it} = \alpha_i^{STWFE} + \delta_t^{STWFE} + \beta^{STWFE} D_{it} + e_{it}^{STWFE}, \quad (3)$$

---

<sup>7</sup>This object is analogous to the cohort-specific treatment effect on the treated (CATT) defined in [Sun and Abraham \(2020\)](#).

where the  $\alpha$  are unit-specific intercepts and the  $\delta$  are common time-specific intercepts, and we denote with  $e$  the error term.

**Specification 2. Event study two-way fixed-effects regression (event study TWFE)**

$$y_{it} = \alpha_i^{ETWFE} + \delta_t^{ETWFE} + \sum_{h=-Q}^H \gamma_h^{ETWFE} D_{i,t-h} + e_{it}^{ETWFE}; \quad Q \geq 0, \quad (4)$$

where  $\beta_h^{ETWFE} = \sum_{j=0}^h \gamma_j^{ETWFE}$  provides the event study TWFE estimate for the effect at horizon  $h$  after treatment ( $0 \leq h \leq H$ ), and  $\beta_{-h}^{ETWFE} = -\sum_{j=-h}^{-1} \gamma_j^{ETWFE}$  is likewise an estimate of possible pre-trends at horizon  $h$  before treatment ( $-Q \leq h \leq -1$ ). An equivalent specification of the event study TWFE regression uses the first difference of the treatment indicator  $\Delta D$  instead of its level  $D$ , except for the  $H$ -th lag, which is taken in level.<sup>8</sup>

**Specification 3. Local Projections regression (LP)**

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP} \Delta D_{it} + e_{it}^h; \quad \text{for } h = -Q, \dots, 0, \dots, H; \quad Q \geq 0; \quad h \neq -1. \quad (5)$$

As a result of the differencing, note that the LP specification no longer includes unit fixed effects for comparability with the two previous specifications. Moreover, a different regression is needed for each time horizon  $h$ , in contrast to the previous two specifications. Given our assumption of staggered absorbing treatment, if treatment is administered at time  $s$ , then  $D_{it} = 1$  for  $t \geq s$  and hence  $\Delta D_{it} = 1$  for  $s = t$ , but  $\Delta D_{it} = 0$  for  $t \neq s$ .

Turning to the coefficients of interest, in all specifications the  $\beta$  terms are population regression coefficients, while the OLS estimates of these coefficients will be denoted by  $\hat{\beta}$ .

**2.2 Basic DiD setting with two groups and two time periods**

The link between LP and DiD is easiest to see in a basic 2-groups/2-periods (2x2) setting. In this setting, an LP regression at horizon  $h = 0$  is equivalent to a first-difference regression or a static TWFE regression, both widely-used DiD implementations.

Hence, assume two groups of units and two time periods. In the first period (pre-treatment) no unit is treated. In the second period (post-treatment) one group of units is treated while the other remains untreated. In terms of the general setup and notation

---

<sup>8</sup>Yet another specification of event study TWFE, often used in applications, uses the first difference of the treatment indicator  $\Delta D$  and normalizes estimates by subtracting the coefficient on the first lead.



introduced above, we are setting  $T = 2$ , and therefore  $t \in \{1, 2\}$ . Moreover, we have  $g \in \{0, 1\}$ , where group 0 is the control group and group 1 the treatment group. For units in the treatment group  $p_i = p_1 = 2$ . For units in the control group  $p_i = p_0 = \infty$ .

Our interest is in estimating the ATT in period  $t = 2$ , defined as  $E[y_{i2}(2) - y_{i2}(0)|p_i = 2]$ . Given the no-anticipation and parallel trends assumptions (Assumptions 1 and 2) stated earlier, the ATT in this setting can be rewritten as follows,

$$\begin{aligned}
ATT &\equiv E[y_{i2}(2) - y_{i2}(0)|p_i = 2] \\
&= E[(y_{i2}(2) - y_{i1}(0)) - (y_{i2}(0) - y_{i1}(0))|p_i = 2] \\
&= E[y_{i2}(2) - y_{i1}(0)|p_i = 2] - E[y_{i2}(0) - y_{i1}(0)|p_i = \infty] \\
&= E[\Delta y_{i2}|p_i = 2] - E[\Delta y_{i2}|p_i = \infty] \equiv \beta^{2x2},
\end{aligned}$$

where  $\beta^{2x2}$  is the well-known 2x2 DiD estimand (Angrist and Pischke, 2009, pp. 227–233).

Now consider an LP regression (Equation 5) with time horizon  $h = 0$ . In this 2x2 setting, this boils down to a simple first-difference regression

$$\Delta y_{it} \equiv y_{i2} - y_{i1} = \delta + \beta_0^{LP} \Delta D_{i2} + e_{i2}.$$

Since  $\Delta D_{i2} = D_{i2}$  in this simple case, we therefore have that

$$\beta_0^{LP} = E[\Delta y_{i2}|D_{i2} = 1] - E[\Delta y_{i2}|D_{i2} = 0] = \beta^{2x2} = ATT.$$

Thus, in the 2x2 setting, the LP regression at horizon  $h = 0$  is equivalent to a first-difference regression, and its population coefficient corresponds to the 2x2 DiD estimand  $\beta^{2x2}$ , which (given no-anticipation and parallel trends) equals the ATT. As is well known, in this setting also the estimand  $\beta^{TWFE}$  from the static TWFE regression of Equation 3 is equivalent to the coefficient from a first-difference regression and corresponds to  $\beta^{2x2}$  (Angrist and Pischke, 2009, pp. 233–236). We thus have  $\beta_0^{LP} = \beta^{STWFE} = \beta^{2x2} = ATT$ .

### 2.3 Two groups and multiple time periods

We now consider a slightly extended setting, with two groups (treated and control), multiple time periods  $T > 2$ , and where all treated units enter treatment in the same time period. Also in this setting, we show that an LP regression is a way to implement the DiD method and recover the (dynamic) ATT.

Specifically, assume that all units in the treatment group enter treatment at time  $s$ , with  $1 < s < T$ , and remain treated thereafter, while control units are never treated over the sample period. In terms of our general setup and notation, we are setting  $g \in \{0, 1\}$ ,

where group 0 is the control group and group 1 the treatment group. For all units in the treatment group,  $p_i = p_1 = s$ . For all units in the control group,  $p_i = p_0 = \infty$ . With only one treated cohort, the dynamic ATT (Equation 1) does not need the treatment group indicator, and becomes simply  $\tau_h = E[y_{i,s+h}(s) - y_{i,s+h}(o) | p_i = s]$ .

Again, via no-anticipation and parallel trends assumptions (Assumptions 1 and 2),

$$\begin{aligned}\tau_h &\equiv E[y_{i,s+h}(s) - y_{i,s+h}(o) | p_i = s] \\ &= E[(y_{i,s+h}(s) - y_{i,s-1}(o)) - (y_{i,s+h}(o) - y_{i,s-1}(o)) | p_i = s] \\ &= E[y_{i,s+h}(s) - y_{i,s-1}(o) | p_i = s] - E[y_{i,s+h}(o) - y_{i,s-1}(o) | p_i = \infty] \\ &= E[y_{i,s+h} - y_{i,s-1} | p_i = s] - E[y_{i,s+h} - y_{i,s-1} | p_i = \infty] \quad \equiv \beta_h^{DiD},\end{aligned}$$

where  $\beta_h^{DiD}$  is the DiD estimand for the dynamic ATT  $h$  periods after treatment.

The population coefficient  $\beta_h^{LP}$  from an LP regression (Equation 5) corresponds exactly to this estimand. To see this, note that in this setting the LP regression of Equation 5 is equivalent to the following cross-sectional regression, estimated on a subsample including all units, but only for the time period  $t = s$ ,

$$y_{i,s+h} - y_{i,s-1} = \delta^h + \beta_h^{LP} \Delta D_{i,s} + e_{i,s}^h.$$

Therefore we have

$$\beta_h^{LP} = E[y_{i,s+h} - y_{i,s-1} | \Delta D_{i,s} = 1] - E[y_{i,s+h} - y_{i,s-1} | \Delta D_{i,s} = 0] = \beta_h^{DiD} = \tau_h.$$

This equivalence holds because when  $t \neq s$  there is no variation in the regressor  $\Delta D_{it}$ . Hence, observations with  $t \neq s$  do not contribute to the estimated coefficient  $\beta_h^{LP}$ , and the coefficient  $\beta_h^{LP}$  is only identified using observations for time  $t = s$ .

From results in the recent literature on DiD (for example de Chaisemartin and D'Haultfœuille 2020; Gardner 2021; Sun and Abraham 2020; Goodman-Bacon 2021), we know that in this setting, with only one treated cohort, and under Assumptions 1 and 2, it also follows that the coefficients in the event-study TWFE regression (Equation 4) correspond to the  $\tau_h$  estimands.<sup>9</sup> Moreover, the  $\beta^{STWFE}$  estimand from the static TWFE

<sup>9</sup>This can be seen using the decomposition of the event-study TWFE coefficients ( $\beta_h^{ETWFE}$  in our notation) provided by Sun and Abraham (2020). This shows that  $\beta_h^{ETWFE}$  is equal to  $\tau_h$  plus a bias term that can arise if the ATE is heterogeneous across cohorts. With only one treatment cohort, obviously, heterogeneity across cohorts cannot arise, and  $\beta_h^{ETWFE} = \tau_h$ .

regression (Equation 3) equals the ATT, defined as  $E[\tau_{it}|D_{it} = 1]$ .<sup>10</sup>

## 2.4 Staggered treatment adoption with dynamic but homogeneous treatment effects

We now allow for multiple treated groups which enter treatment at different points in time (treatment is *staggered*). For now, we assume that the average treatment effect trajectory (or path) does not differ across treatment cohorts (i.e., we assume that treatment effects are *homogeneous*). In terms of our general setup and notation, we now have  $G > 1$ , meaning that we have more than one treatment group, and  $\tau_h^g = \tau_h$  for all  $g > 0$ .

In this setting with staggered treatment and dynamic but homogeneous treatment effects, we still have that an LP regression (Equation 5) augmented with an adequate number of lags and leads of the treatment indicator is able to recover the average treatment effect path under the parallel trends and no-anticipation assumptions introduced earlier.

Here is how we arrive at this result. Under Assumptions 1 and 2 and assuming that treatment effects are homogeneous, mean observed outcomes at time  $t + h$  are given by

$$\begin{aligned} E[y_{i,t+h}] &= E[y_{i,t+h}(o)] + \sum_{p=1}^T [(E(y_{i,t+h}(p)) - y_{i,t+h}(o)) \times \mathbf{1}\{p_i = p\}] \\ &= E[y_{i,t+h}(o)] + \sum_{j=-h}^{\infty} \tau_{h+j} \times \mathbf{1}\{p_i = t-j\} \\ &= \alpha_i + \delta_{t+h} + \tau_h \Delta D_{i,t} + \sum_{\substack{j=-h \\ h \neq 0}}^{\infty} \tau_{h+j} \Delta D_{i,t-j}. \end{aligned} \quad (6)$$

Hence, by now subtracting  $E[y_{i,t-1}]$  from both sides of the previous expression and defining  $\delta_t^h = \delta_{t+h} - \delta_{t-1}$ , we obtain,<sup>11</sup>

$$E[y_{i,t+h} - y_{i,t-1}] = \delta_t^h + \tau_h \Delta D_{i,t} + \sum_{j=1}^h \tau_{h-j} \Delta D_{i,t+j} + \sum_{j=1}^{\infty} [\tau_{h+j} - \tau_{j-1}] \Delta D_{i,t-j}.$$

Therefore the dynamic ATT  $\tau_h$  corresponds to the  $\beta_h^{LP}$  population coefficient in the

<sup>10</sup>One way to see this is to use the decomposition of the static TWFE into a weighted average of treatment-cohort specific ATTs (de Chaisemartin and D'Haultfœuille 2020, p. 2970; Gardner 2021, p. 7). This decomposition implies that, when there is only one treatment cohort and the panel is balanced,  $\beta^{STWFE}$  corresponds to an equally-weighted average of all the cell-specific ATTs.

<sup>11</sup>Note that  $E[y_{i,t-1}] = \alpha_i + \delta_{t-1} + \sum_{j=1}^{\infty} \tau_{j-1} \Delta D_{i,t-j}$ .

following LP regression,

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP} \Delta D_{it} + \sum_{\substack{j=-h \\ j \neq 0}}^{\infty} \theta_j^h \Delta D_{i,t-j} + e_{it}^h. \quad (7)$$

This LP regression includes lags of the differenced treatment indicator, but also its leads up to period  $t+h$ . Leads are necessary to account for the possibility that a unit might enter treatment between period  $t+1$  and period  $t+h$ .

What do the static and event-study TWFE specifications of [Equation 3](#) and [Equation 4](#) identify in this setting with staggered treatment and dynamic but homogeneous effects?

Results from the recent DiD literature show that a static TWFE regression ([Equation 3](#)) can suffer from bias if treatment effects are dynamic (in the sense that  $\tau_h \neq \tau_{h+1}$  for some  $h$ ), even under parallel trends, no-anticipation, and homogeneity across treatment cohorts.

Intuitively, the bias comes from the fact that previously treated units are effectively used as controls for newly treated units. Since previously treated units might still be experiencing a delayed dynamic response to treatment, these treatment effect dynamics are effectively subtracted from the static TWFE treatment effect estimate ([Goodman-Bacon, 2021](#)). That is, delayed dynamic responses to treatment can enter the static TWFE estimate ([Equation 3](#)) with a negative weight ([de Chaisemartin and D'Haultfœuille, 2020](#)).

Under the assumption of homogeneous treatment effects, however, event-study TWFE regression ([Equation 4](#)) does not suffer from this bias and, like the LP regression with lags and leads of treatment discussed above, unbiasedly estimates the average treatment effect path under parallel trends and no anticipation, as long as a sufficient number of lags of the treatment indicator is included (see [Proposition 4](#) and [Eq. 19](#) in [Sun and Abraham, 2020](#)). Intuitively, the lagged treatment indicators control for lagged dynamic effects of previous treatments, which in this setting are the same in expectation for all units.

The difference between [Equation 4](#) and [Equation 7](#) arises from how they handle unit fixed effects. The former (TWFE) removes unit fixed effects with mean differencing using the full sample, while the latter (LP) does so by differencing around the treatment times. The latter could be advantageous in reducing bias if the unit effects are not fully fixed (i.e., there are violations in the parallel trends assumption). For an example of this, see [Cengiz et al. \(2019\)](#).

## 2.5 Staggered treatment adoption with dynamic and heterogeneous treatment effects

We now abandon the assumption of homogeneity of the treatment effect path, and allow for heterogeneous treatment effects across different cohorts. Formally, we have  $\tau_h^g \neq \tau_h^{g'}$  for at least some time-horizon  $h$  and some pair of groups  $g' \neq g$ .

It is now well understood that in this setting conventional TWFE implementations can be biased (see the surveys in Roth et al., 2023; de Chaisemartin and D’Haultfœuille, 2022). The static TWFE estimator (Equation 3) is biased because of dynamic lagged effects (de Chaisemartin and D’Haultfœuille, 2020; Goodman-Bacon, 2021).<sup>12</sup> The event-study TWFE specification (Equation 4) accommodates dynamics but can fail under heterogeneity (Sun and Abraham, 2020).<sup>13</sup>

To understand the relation between LP and DiD in this setting, we can start by noting that here  $E[y_{i,t+h}]$  is determined as follows,

$$\begin{aligned}
 E[y_{i,t+h}] &= E[y_{i,t+h}(0)] + \sum_{p=1}^T \left[ (E(y_{i,t+h}(p) - y_{i,t+h}(0)) \times \mathbf{1}\{p_i = p\}) \right] \\
 &= E[y_{i,t+h}(0)] + \sum_{g=1}^G \left[ \left( \sum_{j=-h}^{\infty} \tau_{h+j}^g \times \mathbf{1}\{p_g = t-j\} \right) \times \mathbf{1}\{p_i = p_g\} \right] \\
 &= E[y_{i,t+h}(0)] + \sum_{g=1}^G \left[ \sum_{j=-h}^{\infty} \left( \tau_{h+j}^g \times \mathbf{1}\{p_g = t-j\} \times \Delta D_{i,t-j} \right) \right] \\
 &= \alpha_i + \delta_{t+h} + \sum_{g=1}^G \left[ \tau_h^g \times \Delta D_{i,t} \times \mathbf{1}\{t = p_g\} \right] \\
 &\quad + \sum_{g=1}^G \left[ \sum_{j=1}^{\infty} \left( \tau_{h+j}^g \times \Delta D_{t-j} \times \mathbf{1}\{t = p_g + j\} \right) \right] \\
 &\quad + \sum_{g=1}^G \left[ \sum_{j=1}^h \left( \tau_{h-j}^g \times \Delta D_{t+j} \times \mathbf{1}\{t = p_g - j\} \right) \right].
 \end{aligned}$$

Subtracting  $E[y_{i,t-1}]$  from both sides, we obtain,<sup>14</sup>

$$\begin{aligned}
 E[y_{i,t+h} - y_{i,t-1}] &= \delta_t^h + \sum_{g=1}^G \left[ \tau_h^g \times \Delta D_{i,t} \times \mathbf{1}\{t = p_g\} \right] \\
 &\quad + \sum_{g=1}^G \left[ \sum_{j=1}^{\infty} \left( (\tau_{h+j}^g - \tau_{j-1}^g) \times \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\} \right) \right] \\
 &\quad + \sum_{g=1}^G \left[ \sum_{j=1}^h \left( \tau_{h-j}^g \times \Delta D_{i,t+j} \times \mathbf{1}\{t = p_g - j\} \right) \right].
 \end{aligned} \tag{8}$$

<sup>12</sup>de Chaisemartin and D’Haultfœuille (2020) show that under no-anticipation and parallel trends  $\beta^{STWFE}$  in Equation 3 equals in expectation a weighted average of all cohort-and-time specific ATTs, but with weights that can be negative. Negative weights introduce bias: for example, positive effects can enter the formula for the TWFE coefficient with a negative sign. The Goodman-Bacon (2021) decomposition theorem provides another perspective on the same problem: the static TWFE estimator of Equation 3 is a variance-weighted average of all potential 2x2 comparisons in the data, but it includes ‘unclean’ comparisons in which previously treated units are used as controls for newly-treated units, leading to potential bias.

<sup>13</sup>Specifically, Sun and Abraham (2020) show that the relative-period coefficients (i.e., the coefficients on leads and lags of treatment in Equation 4) can be contaminated by effects from other periods.

<sup>14</sup>Note that in this setting  $E[y_{i,t-1}] = \alpha_i + \delta_{t-1} + \sum_{g=1}^G \left[ \sum_{j=1}^{\infty} \tau_{j-1}^g \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\} \right]$ .

Without appropriate adjustment to take into account the last two sums on the right-hand side of Equation 8, the simplest LP regression of Equation 5 would be mis-specified in this setting. The easiest way to see this and understand the sources of bias is to consider the special case where  $\delta_t^h = \delta^h$  for all  $t$  and  $h$ .<sup>15</sup> In this special case, we have

$$\begin{aligned}
E[\hat{\beta}_h^{LP}] &= E[y_{i,t+h} - y_{i,t-1} | \Delta D_{it} = 1] - E[y_{i,t+h} - y_{i,t-1} | \Delta D_{it} = 0] \\
&= E \left[ \sum_g^G (\tau_h^g \times \mathbf{1}\{t = p_g\}) | \Delta D_{it} = 1 \right] \\
&\quad - E \left[ \sum_{g=1}^G \left[ \sum_{j=1}^{\infty} \left( (\tau_{h+j}^g - \tau_{j-1}^g) \times \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\} \right) \right] | \Delta D_{it} = 0 \right] \\
&\quad - E \left[ \sum_{g=1}^G \left[ \sum_{j=1}^h \left( \tau_{h-j}^g \times \Delta D_{i,t+j} \times \mathbf{1}\{t = p_g - j\} \right) \right] | \Delta D_{it} = 0 \right].
\end{aligned} \tag{9}$$

Equation 9 shows that, without appropriate adjustment, the simplest LP regression of Equation 5 suffers from two sources of bias.<sup>16</sup>

The first source of bias is the presence of previously treated units in the control group, i.e., observations such that  $\Delta D_{it} = 0$  but  $\Delta D_{i,t-j} \neq 0$  for some  $j \geq 1$ . These previously treated units contribute to the estimated counterfactual for units entering treatment at time  $t$ , as if they were untreated, although they might in fact be experiencing dynamic treatment effects. This bias exists as long as, for some treatment cohort  $g$  at some time-horizon  $h + j$ , we have  $\tau_{h+j}^g \neq \tau_{j-1}^g$ , meaning that treatment effects evolve gradually over time. As a result, the dynamic changes in treatment effects that these previously treated units might be experiencing enter Equation 9 with a *negative* sign. This is a manifestation of the ‘negative weights’ bias discussed in the recent literature on DiD.

Moreover, in the LP setting, a second potential source of bias is the presence in the control group of units that are treated between  $t + 1$  and  $t + h$ , i.e., observations such that

<sup>15</sup>We focus on this special case here for ease of exposition. When studying our proposed LP-DiD estimator below, we will analyse the general case with unrestricted common time effects (Section 3.2).

<sup>16</sup>de Chaisemartin and D’Haultfoeuille (2022, pp. 33-34) present similar negative results about a naive application of LP with panel data, although they consider a LP specification in levels, with  $y_{i,t+h}$  (rather than  $y_{i,t+h} - y_{i,t-1}$ ) as the dependent variable and with  $D_{it}$  (rather than  $\Delta D_{it}$ ) on the right-hand side. The LP specification in levels studied by de Chaisemartin and D’Haultfoeuille (2022) has sometimes been used in applied work and suffers from additional problems, which de Chaisemartin and D’Haultfoeuille (2022, pp.33-34) analyse in detail.

$\Delta D_{it} = 0$  but  $\Delta D_{i,t+j} \neq 0$  for some  $j$  in  $1 \leq j \leq h$ .<sup>17</sup>

To summarize, with staggered treatment, a *naive* LP regression might suffer from the ‘negative weights’ bias highlighted by recent studies, arising from unclear comparisons using previously treated units as controls. The next Section presents our main contribution: we will show how a properly specified LP regression solves these problems and provides a simple, general and flexible regression-based framework for DiD.

### 3 LP-DiD estimator

This section presents our main contribution: a properly specified LP regression, which we call *LP-DiD*, consistently estimates a convex weighted average treatment effect without incurring in the ‘negative weights’ problem discussed above. The key is to define appropriate treated and control units following the ‘clean control approach’ of [Cengiz et al. \(2019\)](#). We explicitly characterize the weights assigned to each cohort-specific treatment effect, and show that they are non-negative and proportional to group size and treatment variance (Section 3.2). We then show how a simple re-weighted LP regression can recover an equally-weighted average treatment effect on the treated (3.3). Moreover, we show how to employ alternative pre-treatment base periods (3.5), pool estimates over a post-treatment horizon (3.5), and avoid composition effects (3.6). Finally, we lay out the relation with other recent DiD estimators (3.7).

This Section considers the setup, notation and identification assumptions introduced in the previous Section, and in particular in Section 2.5: treatment is binary, staggered and absorbing; the no anticipation and parallel trends assumptions hold unconditionally;

---

<sup>17</sup>As [Equation 8](#) shows, one solution would be a LP regression that estimates separately the effect for each group by interacting group indicators with the contemporaneous differenced treatment indicator, while at the same time controlling for interaction terms between group indicators and the leads and lags of the differenced treatment indicator. These interaction terms ‘clean’ the estimated counterfactual from the bias coming from the influence of previously treated units. Moreover, this is equivalent to interacting the leads and lags of the differenced treatment indicator with time indicators. One could then obtain an overall ATE by computing some convex combination of all the individual group-specific effects. This solution could be fruitful in some settings and has similarities with the interactive fixed effects estimator proposed by [Sun and Abraham \(2020\)](#), but generally has some drawbacks. In practical applications, it involves estimating a potentially very large number of interaction terms, where the coefficients are of no economic interest. Moreover, our aim in this paper is to show that it is possible to directly estimate a convex combination of all the cohort-specific effects, without having to first estimate them separately and then aggregate.



treatment effects can be dynamic and heterogeneous.<sup>18</sup>

### 3.1 LP-DiD specification

LP-DiD consists, in essential form, of estimating the LP specification of Equation 5 in a restricted sample that only includes newly treated observations ( $\Delta D_{it} = 1$ ) and not-yet treated ones ( $\Delta D_{i,t-j} = 0$  for  $-h \leq j \leq \infty$ ). Under the assumption of absorbing binary treatment, the restriction imposed on the control group ( $\Delta D_{i,t-j} = 0$  for  $-h \leq j \leq \infty$ ) simplifies to  $D_{i,t+h} = 0$ . Intuitively, as recent literature has made clear and as Equation 8 and Equation 9 illustrate, ‘negative weights’ bias comes from unclean comparisons in which previously treated units are used as controls for newly-treated units. Excluding these ‘unclean’ observations from the control group eliminates the bias.

Formally, consider the following specification of an LP-DiD regression,

**LP-DiD regression** Estimate the regression

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

by restricting the estimation sample to observations that are either

$$\begin{cases} \text{newly treated:} & \Delta D_{it} = 1, \\ \text{or clean control:} & D_{i,t+h} = 0. \end{cases} \quad (10)$$

The clean control condition in Equation 10 ensures that the estimate is obtained from a set of clean comparisons between units entering treatment at  $t$  and units that are not yet treated at  $t+h$ . This sample restriction eliminates the need to explicitly model (or ‘control for’) the lagged effects of previous treatments or of treatments happening between  $t+1$  and  $t+h$ , because it removes previously treated observations and observations treated

---

<sup>18</sup>The LP-DiD approach presented in this Section can be valuable also in settings in which treatment effects are assumed to be homogeneous. As shown above (Section 2.4), under homogeneous effects, simple LP or dynamic TWFE specifications are sufficient to obtain an unbiased estimate, provided that a sufficient number of lags of the treatment indicator is included. However, there are two reasons for still using LP-DiD (with a clean control condition) also in that setting. First, and most obviously, LP-DiD is robust to possible failure of the homogeneous effects assumption. Second, even if homogeneity holds, LP-DiD relieves the researcher from the problem of selecting the appropriate number of lags.



between  $t + 1$  and  $t + h$  from the control group.<sup>19</sup>

The  $\beta_h^{LP-DiD}$  coefficient from Equation 10 consistently estimates a convex combination of all group-specific effects  $\tau_h^g$ . Before analysing the general case (in Section 3.2 below), it is instructive to consider again the special case with  $\delta_t^h = \delta^h$ . In this special case, we have

$$\begin{aligned} E[\hat{\beta}_h^{LP-DiD}] &= E(y_{i,t+h} - y_{i,t-1} | \Delta D_{it} = 1) - E(y_{i,t+h} - y_{i,t-1} | \Delta D_{it} = 0, D_{i,t+h} = 0) \\ &= E \left[ \sum_g^G \left( \tau_h^g \times \mathbf{1}\{t = p_g\} \right) | \Delta D_{it} = 1 \right]. \end{aligned}$$

### 3.2 Weights of the LP-DiD estimator

We can explicitly characterize the weights assigned to each cohort-specific effect  $\tau_h^g$  when the LP-DiD specification (Equation 10) is estimated with OLS in the general case with unrestricted common time effects. The key result is that, under parallel trends and no-anticipation (Assumptions 1 and 2), the LP-DiD specification of Equation 10 consistently estimates a weighted average of all cohort-specific treatment effects, with weights that are *always positive* and depend on treatment variance and subsample size. Here we present this result. A simple formal derivation based on the Frisch-Waugh-Lovell theorem is provided in Appendix A.

To illustrate the result, we need to introduce further definitions. Recall that the time period in which group  $g$  enters treatment is  $p_g$ . For each treatment group  $g > 0$ , define the clean control sample (CCS) for group  $g$  at time horizon  $h$  (denoted as  $CCS_{g,h}$ ) as the set of observations for time  $t = p_g$  that satisfy the sample restriction in Equation 10. Therefore  $CCS_{g,h}$  includes the observations at time  $p_g$  for all units that either enter treatment at  $p_g$  or are still untreated at  $p_{g+h}$ . In other words,  $CCS_{g,h}$  includes observations at  $p_g$  for group  $g$  and its *clean controls*.

Under parallel trends and no anticipation (Assumptions 1 and 2), the LP-DiD estimator  $\beta_h^{LP-DiD}$  recovers the following weighted average effect,

$$E(\hat{\beta}_h^{LP-DiD}) = \sum_{g \neq 0} \omega_{g,h}^{LP-DiD} \tau_h^g. \quad (11)$$

---

<sup>19</sup>The clean control condition of Equation 10 can equivalently be implemented by estimating Equation 5 in the *full sample* with a full set of interaction terms between a binary indicator for ‘unclean’ (i.e., previously treated) control observations and time effects. With absorbing treatment, this binary indicator is equal to  $UC_{it} = \mathbf{1}\{\Delta D_{it} = 0\} \times \mathbf{1}\{D_{i,t+h} = 1\}$ . With control variables (which we will discuss in Section 4 below), the  $UC_{it}$  indicator should also be interacted with any included covariates. Of course, this is just another and completely equivalent way of excluding unclean controls from the estimation sample.

The weight attributed to each group-specific effect is given by

$$\omega_{g,h}^{LP-DiD} = \frac{N_{CCS_{g,h}} [n_{gh}(n_{c,g,h})]}{\sum_{g \neq 0} N_{CCS_{g,h}} [n_{g,h}(n_{c,g,h})]}, \quad (12)$$

where  $N_{CCS_{g,h}}$  is the number of observations in the clean control sample for group  $g$  at time-horizon  $h$ ;  $n_{g,h} = N_g / N_{CCS_{g,h}}$  is the share of treated units in the  $CCS_{g,h}$  subsample; and  $n_{c,g,h} = N_{c,g,h} / N_{CCS_{g,h}}$  is the share of control units in the  $CCS_{g,h}$  subsample.<sup>20</sup>

Thus, in short, the LP-DiD estimator  $\beta_h^{LP-DiD}$  recovers a variance-weighted ATT (VWATT in the terminology of [Goodman-Bacon, 2021](#)).

### 3.3 Obtaining an equally-weighted average effect

If a researcher is instead interested in an equally-weighted ATT, there are two equivalent ways to obtain it within an LP-DiD framework. The first is to employ a re-weighted regression. The second is to use regression adjustment.

Indeed, Equations 11-12 imply that estimation of an LP-DiD regression ([Equation 10](#)) through weighted least squares, assigning to an observation belonging to  $CCS_{g,h}$  a weight equal to  $1 / (\omega_{g,h}^{LP-DiD} / N_g)$ , allows unbiased estimation of the equally-weighted ATT.

In practical applications, the weight  $(\omega_{g,h}^{LP-DiD} / N_g)$  can be obtained by computing subsamples sizes and shares of treated and control units in the sample and using [Equation 12](#), or through an auxiliary regression. Specifically, consider an auxiliary regression of  $\Delta D$  on time indicators in the sample defined by [Equation 10](#). Define  $\Delta \tilde{D}_{g,p_g}$  as the residual at time  $p_g$  for a unit belonging to group  $g$ .<sup>21</sup> The Frisch-Waugh-Lovell theorem implies that

$$(\omega_{g,h}^{LP-DiD} / N_g) = \frac{\Delta \tilde{D}_{g,p_g}}{\sum_{g \neq 0} N_g \Delta \tilde{D}_{g,p_g}},$$

where further discussion can be found in [Appendix A](#).

Another equivalent way to obtain an equally-weighted ATT is to estimate the LP-DiD specification with the clean control condition of [Equation 10](#) through regression adjustment. Here, regression adjustment uses clean control units to estimate a counterfactual outcome change for each treated unit, and then compute an average estimated effect assigning equal weight to each treated unit, thus estimating an equally-weighted ATT.

In particular, regression adjustment can be implemented as follows. Regress  $y_{i,t+h} -$

<sup>20</sup>The derivation of these weights is in [Appendix A](#).

<sup>21</sup>Note that  $\Delta \tilde{D}_{g,p_g}$  will be identical for all units belonging to the same group.

$y_{i,t-1}$  on time indicators using only clean control observations (i.e., observations with  $D_{i,t+h} = 0$ ). Use the estimated coefficients to get a predicted value in absence of treatment  $y_{i,t+h} - \widehat{y}_{i,t-1}$  for each treated unit. The ATT is then estimated as

$$N_{TR}^{-1} \sum_{i \in TR} [(y_{i,t+h} - y_{i,t-1}) - (y_{i,t+h} - \widehat{y}_{i,t-1})],$$

where  $TR$  is the set of treated observations (i.e., the set of observations with  $\Delta D_{i,t} = 1$ ).<sup>22</sup> This regression-adjustment implementation of LP-DiD constitutes an imputation estimator, in the same sense as in [Borusyak, Jaravel, and Spiess \(2021\)](#).

### 3.4 Alternative pre-treatment base periods

In some settings, there can be efficiency gains from adopting an alternative LP-DiD specification, one in which the long difference of the outcome variable is taken relative to its average value over some interval before  $t$ , instead of relative to just its first lag. Formally, this alternative specification uses  $y_{i,t+h} - \frac{1}{k} \sum_{\tau=t-k}^{t-1} y_{i,\tau}$  instead of  $y_{i,t+h} - y_{i,t-1}$  as the dependent variable.

The motivation for considering such an alternative base period in the long difference of the outcome is a possible efficiency-related concern with the LP-DiD specification of [Equation 10](#) (as well as the simple LP regression of [Equation 5](#)). Typically, an LP uses the long difference  $y_{i,t+h} - y_{i,t-1}$  as the dependent variable. Period  $t-1$  is thus used as the pre-treatment base period: for a treatment event occurring at time  $s$ , the expected value of the outcome in the pre-treatment period in the treated group and its clean controls are estimated from  $y_{i,s-1}$ . However, the number of time periods available for estimating the expected value of the outcome in the pre-treatment period is larger than just  $s-1$ : observations for all time periods  $t < s$  can potentially be used. For this reason, using a single pre-treatment period as the baseline may be inefficient.

This concern can be accommodated by using an average of pre-treatment observations ( $\frac{1}{k} \sum_{\tau=t-k}^{t-1} y_{i,\tau}$ ) as the baseline, instead of just  $y_{i,t-1}$ . The dependent variable thus becomes  $y_{i,t+h} - \frac{1}{k} \sum_{\tau=t-k}^{t-1} y_{i,\tau}$ . This gives rise to the following ‘pre-mean-differenced’ (PMD hereafter) specification of LP-DiD,

<sup>22</sup>It is easy to implement this method for recovering the equally-weighted ATT using standard statistical software. The following is an example in STATA syntax:

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

where  $y$  is the variable measuring the outcome of interest;  $h$  is the time-horizon of the estimate;  $Dhy = y_{i,t+h} - y_{i,t-1}$ ;  $dtreat$  is the first difference of the binary treatment indicator;  $time$  is a variable indexing time periods;  $unit$  is a variable indexing units, and we are clustering standard errors at the level of units.

## PMD LP-DiD regression

$$\begin{aligned}
 y_{i,t+h} - \frac{1}{k} \sum_{\tau=t-k}^{t-1} y_{i,\tau} &= \beta_h^{PMD \text{ LP-DiD}} \Delta D_{it} && \text{treatment indicator} \\
 &+ \delta_t^h && \text{time effects} \\
 &+ e_{it}^h && \text{for } h = 0, \dots, H,
 \end{aligned}$$

restricting the estimation sample to observations that are either:

$$\left\{ \begin{array}{l} \text{newly treated: } \Delta D_{it} = 1, \\ \text{or clean control: } D_{i,t+h} = 0. \end{array} \right. \quad (13)$$

(Note here that by setting  $k = t - 1$ , one can use all available observations for estimating the expected value of the outcome in the pre-treatment period.)

The results presented earlier in this Section and in Appendix A imply that  $\beta_h^{PMD \text{ LP-DiD}}$  recovers a convex weighted average of cohort-specific effects, with the same weights  $\omega^{LP-DiD}$  discussed in Section 3.2. Also in this case, weighted regression or regression adjustment can be employed to obtain an equally-weighted ATE (Section 3.3).

The potential advantages and risks of differencing with respect to the pre-treatment average ('pre-mean differencing') relative to differencing over a single lag ('first-lag differencing') have been discussed in the recent literature (a review of these discussions is provided in [de Chaisemartin and D'Haultfœuille 2022](#), pp. 18–19). The potential advantage of pre-mean differencing is the efficiency gain discussed above. This advantage is greater the lower the autocorrelation in untreated potential outcomes ([Borusyak, Jaravel, and Spiess, 2021](#); [Harmon, 2022](#))<sup>23</sup>. A potential risk is that, under some deviations from the parallel trends assumption, pre-mean differencing can amplify the bias relative to first-lag differencing. If parallel trends holds between periods  $s$  and  $s + h$  ( $s$  being the time of treatment), but not in earlier pre-treatment periods, first-lag differencing will still be unbiased, while pre-mean differencing will be biased. In this sense, first-lag differencing relies on a weaker parallel trends assumption than pre-mean differencing ([Marcus and Sant'Anna, 2021](#)). Moreover, if parallel trends does not fully hold at any time period, and the gap in average untreated potential outcomes between treated and controls increases over time, then pre-mean differencing will be more biased than first-lag differencing.

The ability of LP-DiD to allow researchers to flexibly choose the appropriate pre-

---

<sup>23</sup>[Borusyak et al. \(2021\)](#) prove that in the extreme case of no autocorrelation, differencing with respect to the pre-treatment average is more efficient than differencing over the last pre-treatment lag. However, [Harmon \(2022\)](#) finds that in the opposite polar case where errors follow a random walk (very high autocorrelation), this efficiency ranking reverses, with differencing over the last pre-treatment period being more efficient.

treatment base period, and to easily test robustness to alternative choices, is an additional advantage of LP-DiD. Other available DiD estimators (for example [Callaway and Sant’Anna \(2020\)](#), [Sun and Abraham \(2020\)](#) and [Borusyak et al. \(2021\)](#)) are less flexible in this respect. Of course, as with all specification choices, researchers should commit to a preferred choice of pre-treatment base period ex-ante, based on the features of their application and independent of estimation results, in order to avoid introducing bias.

### 3.5 Pooling over post-treatment periods

So far, the LP-DiD estimates have been based on long differenced outcomes,  $y_{i,t+h} - y_{i,t-1}$ , so the coefficients  $\beta_h^{LP-DiD}$  trace out the impulse response  $h$  periods following a treatment. However, we may wish to estimate an *overall* DiD estimate averaged over the full post-treatment window,  $h \in 0, \dots, H$ . This can be easily obtained by estimating a single LP-DiD regression where the dependent variable is a post-treatment period mean of long differenced outcomes:  $\frac{1}{H+1} \sum_{h=0}^H y_{i,t+h} - y_{i,t-1}$ . The resulting coefficient provides the treatment effect over the post-treatment window relative to the event date  $-1$ .

More generally, any linear combination of  $\beta_h^{LP-DiD}$  coefficients can be estimated and tested by appropriately redefining the dependent variable in the LP-DiD regression. This includes testing for differences in treatment effects across event times, or estimating a cumulative response of treatment summing over event time. One could also combine pooling over post-treatment periods with pre-mean differencing for additional power by using the outcome:  $\frac{1}{H+1} \sum_{h=0}^H y_{i,t+h} - \frac{1}{k} \sum_{\tau=t-k}^{t-1} y_{i,\tau}$ .

Of course, another simple way to test joint hypotheses about a combination of  $\beta_h^{LP-DiD}$  coefficients is to jointly estimate the individual LP regressions for the different  $H$  horizons as a system. In applications, this can be done for example by stacking the data. Ready-made packages to perform joint hypotheses from separate regressions are available in commonly used statistical software (e.g., the ‘`suest`’ command in STATA, which is equivalent to stacking the data).

### 3.6 Composition effects

In finite samples, the LP-DiD specification of [Equation 10](#) might suffer from composition effects because the set of treated and clean control units can change across different time horizons  $h$ . Note that the composition effect from a changing set of treated units is also present in other available DiD techniques, including conventional TWFE estimators, while composition effects from a changing control set are a result of the way the clean control condition is specified in [Equation 10](#).

There is a way to rule out composition effects, but at a cost, since it requires a reduction in the number of observations which can reduce statistical power. To keep the control set constant across time horizons, one can modify the clean control condition, defining clean controls at all horizons as units such that  $D_{i,t+H} = 0$ , where  $H$  is the maximum horizon considered in estimation. Moreover, to keep the set of treated units constant across time horizons, one can exclude from the estimation sample treatment events which occur after time period  $T - H$  (i.e., exclude treatment cohorts with  $p_g > T - H$ ).

### 3.7 Relation to other DiD estimators

Some recently proposed DiD estimators can be obtained as special cases of the LP-DiD approach, using specific weighting schemes or choosing particular base periods for constructing the local projection.

First, in the baseline version with absorbing binary treatment and no covariates, the LP-DiD estimator  $\beta_h^{LP-DiD}$  is numerically equivalent to the estimate from a stacked regression approach as implemented in [Cengiz et al. \(2019\)](#).<sup>24</sup> Moreover, the re-weighted version of LP-DiD (discussed in Section 3.3) is numerically equivalent to the estimator proposed by [Callaway and Sant’Anna \(2020\)](#).<sup>25</sup>

Futhermore, the PMD version of the LP-DiD estimator ([Equation 13](#)) is analogous to the estimator proposed by [Borusyak, Jaravel, and Spiess \(2021\)](#) (BJS thereafter), which also implicitly uses pre-mean differencing. In fact, in the special case of one single treated group, it is easy to see that PMD LP-DiD with  $k = t - 1$  is numerically equivalent to the BJS estimator.<sup>26</sup> With more than one treated group, the BJS estimator does not have a closed form expression, and it is therefore not straightforward to assess with precision its relation to our PMD LP-DiD estimator. However, the pre-period mean differencing means that the two estimators use similar information, and indeed our Monte Carlo simulations (presented in Section 5 below) show that with more than one treated group, when using

<sup>24</sup>See [Appendix A](#) for more details about this equivalence.

<sup>25</sup>In terms of our notation, the [Callaway and Sant’Anna \(2020\)](#) estimator of  $\tau_g^h$  is equal to  $E[y_{i,p_g+h} - y_{i,p_g-1} | \Delta D_{i,p_g} = 1] - E[y_{i,p_g+h} - y_{i,p_g-1} | D_{i,p_g+h} = 0]$ . An ATT is then estimated by taking an equally-weighted average across all treated units. In this baseline setting with absorbing treatment, no control variables and using only not-yet treated units as controls, re-weighted LP-DiD equals the same difference in means.

<sup>26</sup>With one single treated group  $g$  that enters treatment in period  $p_g$ , the BJS estimator has a closed form ([de Chaisemartin and D’Haultfœuille, 2022](#), pp.18-19). In terms of our notation, it is equal to

$$N_g^{-1} \sum_{i \in g} \left[ y_{i,p_g+h} - \frac{1}{p_g-1} \sum_{k=1}^{p_g-1} y_{i,k} \right] - N_{c,g,h} \sum_{i \neq g, i \in CCS_{g,h}} \left[ y_{i,p_g+h} - \frac{1}{p_g-1} \sum_{k=1}^{p_g-1} y_{i,k} \right].$$

In this one-group setting, this is exactly equal to the  $\beta_h^{PMD LP-DiD}$  estimator.

reweighting to obtain an equally-weighted ATT, the two estimators produce very similar (although not identical) point estimates.

Some additional considerations about the relation between PMD LP-DiD and BJS are in order. Although very similar, PMD LP-DiD might offer practitioners some advantages over the BJS estimator. First, it is easy to provide an analytical expression for PMD LP-DiD even in the case of more than one treated group, unlike for the BJS estimator.<sup>27</sup> Moreover, unlike the BJS estimator, PMD LP-DiD can be implemented using simple OLS (or weighted least squares) regression, employing commonly used and well understood methods for statistical inference. Third, as we will discuss in Section 4.1 and our empirical applications will illustrate, the structure of the LP-DiD specification allows one to control for *pre-treatment* lags of time-varying covariates.

That said, BJS prove the efficiency of their estimator of the equally weighted ATT under the Gauss-Markov assumptions; this implies that the PMD LP-DiD cannot be more efficient than BJS under these specialized conditions.

However, we note that in our simulations below, the point estimates from the reweighted PMD LP-DiD and BJS estimators are typically very close in most cases. Therefore, any efficiency advantage of the BJS estimator over PMD LP-DiD is likely to be small. Indeed, in our simulation (Section 5) the root mean squared error of the two estimators is nearly identical.

It also bears noting that efficiency of the BJS estimator is only guaranteed under the Gauss-Markov assumptions. These require, among other things, no auto-correlation in untreated potential outcomes, a polar assumption that might be seen as implausible in most panel data applications (Harmon, 2022; de Chaisemartin and D’Haultfoeuille, 2022, p. 18). Moreover, if there is little heterogeneity in treatment effects between treatment cohorts, variance-weighting (as done by LP-DiD or PMD LP-DiD without reweighting) can be more efficient than equal weights.

In general, the LP-DiD implementation is simpler and computationally faster than the alternatives. Relative to Cengiz et al. (2019) it is also less prone to errors in practical applications, given that it does not require the reshaping of the dataset in stacked format. Moreover, as we discuss below, the LP-DiD specification is especially easy to generalize to more complicated settings.

---

<sup>27</sup>With one single treated group, the PMD LP-DiD estimator (like the BJS estimator) is equal to the expression in footnote 26. With more than one treated group, the PMD LP-DiD estimator is equal to a weighted average of the expression in footnote 26 across all treated groups, with weights given by  $\omega_{g,h}^{LP-DiD}$  in Equation 12. The reweighted PMD LP-DiD estimator is equal to a simple average of the expression in footnote 26 across all treated units.



## 4 Extensions

In this section we extend the LP-DiD approach to alternative settings often encountered in empirical research. We begin by discussing the inclusion of covariates (Section 4.1). In 4.2 we consider non-absorbing treatment (units can enter and exit treatment multiple times). In addition, we briefly comment on non-binary treatments (4.3), although we leave a formal development of this extension for future work. This discussion helps us tee up a brief comment on the link between DiD and the impulse responses used in macroeconomics to establish the parts of commonality and departure (4.4).

### 4.1 Inclusion of covariates

The LP-DiD framework can be extended to include covariates. To do so, we consider a setting with staggered treatment adoption, dynamic and heterogeneous treatment effects, and in which parallel trends holds only conditional on a vector of covariates  $\mathbf{x}_i$ .<sup>28</sup>

Formally, using the notation introduced in Section 2.1, we make the following assumptions:<sup>29</sup>

#### Assumption 3. Conditional no anticipation

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

#### Assumption 4. Conditional parallel trends

$$E [y_{i,t+h}(0) - y_{i,t-1}(0) | \mathbf{x}_i; p_i = p] = E [y_{i,t+h}(0) - y_{i,t-1}(0) | \mathbf{x}_i],$$

for all  $t \in \{2, \dots, T\}$ , all  $h \in \{0, \dots, T-1\}$ , and all  $p \in \{1, \dots, T, \infty\}$ .

#### Assumption 5. Linear conditional expectation function

$$E [y_{i,t+h}(0) - y_{i,t-1}(0) | \mathbf{x}_i] = \delta_t^h + \gamma^h \mathbf{x}_i$$

---

<sup>28</sup>Although the covariates are only indexed by unit  $i$  in this discussion, one could always define  $\mathbf{x}_i$  as being a time-varying covariate (or its difference) observed at some time period. In this respect, an important advantage of the LP-DiD framework is that (unlike event study TWFE), if  $\mathbf{x}_i$  includes the lagged value of a time-varying covariate, this is measured pre-treatment.

<sup>29</sup>Note that these assumptions do not restrict how the treatment effect might vary depending on the value of the covariates  $\mathbf{x}$ . We assume linearity (Assumption 5) only for ease of exposition: as we will presently explain, it is possible to accommodate non-linearities in the CEF.



We can define the dynamic ATT at time-horizon  $h$  conditional on  $\mathbf{x}$  as

$$ATT^h(\mathbf{x}) = E \left[ y_{i,p_i+h}(p_i) - y_{i,p_i+h}(0) \mid p_i \neq \infty, \mathbf{x} = \mathbf{x}_i \right],$$

and the unconditional (equally-weighted) dynamic ATT as

$$\begin{aligned} ATT^h &= N_{TR}^{-1} \sum_{g \neq 0} N_g \tau_h^g = E \left[ y_{i,p_i+h}(p_i) - y_{i,p_i+h}(0) \mid p_i \neq \infty \right] \\ &= E \left\{ E \left[ y_{i,p_i+h}(p_i) - y_{i,p_i+h}(0) \mid p_i \neq \infty, \mathbf{x} = \mathbf{x}_i \right] \mid p_i \neq \infty \right\}, \end{aligned}$$

where again  $TR$  denotes the set of observations entering treatment (i.e., the set of observations with  $\Delta D_{i,t} = 1$ ), and the second line follows from the law of iterated expectations.

#### 4.1.1 Regression adjustment LP-DiD specification with covariates

In this setting, a regression-adjustment LP-DiD specification with covariates recovers the equally-weighted ATT.<sup>30</sup> The estimator can be written as follows:

$$\begin{aligned} \hat{\beta}_h^{LP-DiD,RA} &= N_{TR}^{-1} \sum_{i \in TR} \left[ (y_{i,t+h} - y_{i,t-1}) - \hat{E}(y_{i,t+h} - y_{i,t-1} \mid D_{i,t+h} = 0, \mathbf{x}_i) \right] = \\ &= N_{TR}^{-1} \sum_{i \in TR} \left[ (y_{i,t+h} - y_{i,t-1}) - \hat{\gamma}^h \mathbf{x}_i - \hat{\delta}_t^h \right], \end{aligned}$$

where  $\hat{E}[y_{i,t+h} - y_{i,t-1} \mid D_{i,t+h} = 0, \mathbf{x}_i]$  denotes the estimated conditional expectation function estimated in the subsample of clean control observations (ie, the subsample with  $D_{i,t+h}=0$ ) but evaluated at  $\mathbf{x}_i$  for treated unit  $i$ ;  $\hat{\gamma}^h$  and  $\hat{\delta}_t^h$  are estimated coefficients from the following regression model, estimated in the sample of clean control observations:

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \gamma^h \mathbf{x}_i + u_i.$$

Under the assumptions above,  $\hat{\beta}_h^{LP-DiD,RA}$  is an unbiased estimator of  $ATT^h$ . This can

---

<sup>30</sup>It is well known that a similar result holds in the canonical 2x2 setting (Roth et al., 2023, pp.2230-2231).

be shown as follows:

$$\begin{aligned}
E(\hat{\beta}_h^{LP-DiD,RA}) &= E(N_{TR}^{-1} \sum_{i \in TR} [(y_{i,t+h} - y_{i,t-1}) - \hat{E}(y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i)]) \\
&= E[E(y_{i,t+h} - y_{i,t-1} | \Delta D_{i,t} = 1, \mathbf{x}_i) - E(y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i) | \Delta D_{i,t} = 1] \\
&= E[E[y_{i,p_i+h}(1) - y_{i,p_i-1}(0) | p_i \neq \infty, \mathbf{x}_i] - E[y_{i,t+h}(0) - y_{i,t-1}(0) | p_i = \infty, \mathbf{x}_i] | p_i \neq \infty] \\
&= E\left\{E[y_{i,p_i+h}(p_i) - y_{i,p_i+h}(0) | p_i \neq \infty, \mathbf{x} = \mathbf{x}_i] | p_i \neq \infty\right\} \\
&= ATT^h,
\end{aligned} \tag{14}$$

where the second equality follows from the assumption that the conditional expectation function is correctly specified in the first-step regression and the fourth equality follows from conditional parallel trends and no anticipation.

Although we assumed linearity here for ease of exposition, this approach can also accommodate non-linear models for the conditional expectation function. Researchers can allow for non-linearities by obtaining the estimated conditional expectation  $\hat{E}[y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i]$  from a flexible non-parametric or semi-parametric model. Indeed, equation 14 continues to hold under non-linearity, as long as the conditional expectation function is correctly specified in the first-step regression.

#### 4.1.2 Controlling for covariates when treatment effects do not vary with covariates

Under more restrictive assumptions, namely that treatment effects do not vary based on the value of the covariates, estimation of a simple LP-DiD specification with control variables yields a convex variance-weighted effect.

In addition to assumptions 3 and 4 and 5, we now add:

##### **Assumption 6. Treatment effects are independent of covariates**

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \gamma^h \mathbf{x}_i + \tau_{i,t+h} D_{i,t+h} - \tau_{i,t-1} D_{i,t-1} + e_{i,t}^h,$$

Appendix A.2 shows that under these assumptions, simply augmenting the LP-DiD specification of equation 10 with the vector of covariates  $\mathbf{x}_i$  (yielding the specification of equation A.8) yields a convex variance-weighted average effect, with the weights of equation 12. If Assumption 6 fails, however, the weights are hard to characterize and might even be negative (see Appendix A.2).

Given that the RA specification presented in Section 4.1.1 yields a equally weighted ATT and requires weaker assumptions, it appears generally preferable in applications.

## 4.2 Non-absorbing treatment

In many applications, treatment is not absorbing: units can enter and exit treatment multiple times. The LP-DiD framework offers flexibility to accommodate the different target estimands and identification assumptions that might be appropriate under non-absorbing treatment. An appropriate modification of the ‘clean control’ sample restriction of Equation 10 will generally be necessary to implement LP-DiD in this setting. In what follows we extend our setup to allow for non-absorbing treatment and illustrate how LP-DiD can flexibly accommodate this setting.<sup>31</sup>

### 4.2.1 Setup and notation with non-absorbing treatment

To discuss the case of non-absorbing treatment, we extend the framework of Section 2.1 and allow potential outcomes to be a function of the treatment *path*, as in Robins (1986) and de Chaisemartin and D’Haultfoeuille (2022).<sup>32</sup>

Specifically, we make the following modifications. Treatment is still binary ( $D_{it} \in \{0, 1\}$ ), but we remove the assumption  $D_{is} \leq D_{it}$  for  $s < t$ . In other words, we allow units to enter and exit treatment. Let  $\mathbf{D}_i = (D_{i1}, \dots, D_{iT})$  denote the vector containing the observed time series of the treatment variable  $D$  over the sample period for unit  $i$ . The observed outcome can then be written as  $y_{it} = y_{it}(\mathbf{D}_i)$ . Let  $y_{it}(d_1, \dots, d_T)$  denote the potential outcome for unit  $i$  at time  $t$ , if its treatments over the sample period were equal to  $(D_{i1}, \dots, D_{iT}) = (d_1, \dots, d_T)$ .

For units that start untreated ( $D_{i1} = 0$ ), we let  $p_i^n$  denote the time period in which unit  $i$  enters treatment for the  $n$ -th time during the sample period, with the convention  $p_i^n = \infty$  if the unit enters treatment less than  $n$  times in the sample period. Therefore  $p_i^1$  is the time period in which unit  $i$  enters treatment for the first time and we have  $p_i^1 = \infty$  if the unit is never treated during the sample period (i.e., if  $\mathbf{D}_i = \mathbf{0}$ ). For units that are already treated at the start of the sample period ( $D_{i1} = 1$ ), we define  $p_i^1 = 0$ , while for  $n \geq 1$  we define  $p_i^{n+1}$  as the  $n$ -th period in which the unit enters treatment during the sample period, with the convention  $p_i^{n+1} = \infty$  if the unit enters treatment less than  $n$  times during the sample period. For example,  $p_i^2$  is the first time period in which a unit with  $D_{i1} = 1$  re-enters treatment during the sample period. For all units, we also define  $q_i^n$

---

<sup>31</sup>Later, in Section 6.2, we illustrate the use of LP-DiD under non-absorbing treatment in an empirical application which estimates the effect of democracy on economic growth. Moreover, the project repository <https://github.com/danielegirardi/lpdid/> provides a STATA example file illustrating the application of LP-DiD in a simulated dataset with non absorbing treatment.

<sup>32</sup>Recent articles discussing panel data estimation with non-absorbing treatment are for example de Chaisemartin and D’Haultfoeuille (2022); Bojinov et al. (2021); Viviano and Bradic (2023).

as the period in which unit  $i$  exits treatment for the  $n$ -th time during the sample period, with  $q_i^n = \infty$  if the unit exits treatment less than  $n$  times in the sample period.

Groups  $g \in \{1, \dots, G, G + 1\}$  are now defined so that all units within the same group experience the same treatment path over the sample period, while units in different groups do not. In other words, treatment is still assigned at the  $(g, t)$  level. We let  $\mathbf{D}_g$  denote the treatment path experienced by group  $g$ . Group  $g = 0$  again denotes a possible never treated group (i.e., a group with  $\mathbf{D}_g = \mathbf{0}$ ), while group  $g = G + 1$  denotes a possible ‘always treated’ group (i.e., a group with  $\mathbf{D}_g = \mathbf{1}$ ). Groups 1 to  $G$  experience some change in treatment status during the sample period.

The no-anticipation and parallel trends assumptions are rewritten as follows:

**Assumption 7. No anticipation (non-absorbing treatment version)**

$$y_{it}(d_1, \dots, d_T) = y_{it}(d_1, \dots, d_t) \text{ for all } i \text{ and } t.$$

**Assumption 8. Parallel trends (non-absorbing treatment version)**

$$E [y_{it}(\mathbf{0}) - y_{i1}(\mathbf{0}) | \mathbf{D}_i = \mathbf{D}] = E [y_{it}(\mathbf{0}) - y_{i1}(\mathbf{0})], \text{ for all } t \in \{2, \dots, T\} \text{ and all } \mathbf{D}.$$

The intuition for these two assumptions remains the same as in the absorbing treatment setting. As in the previous sections, we further assume  $E [y_{it}(\mathbf{0})] = \alpha_i + \delta_t$

**4.2.2 Estimating the effect of entering treatment for the first time**

We first consider the problem of estimating the effect of entering treatment for the first time and staying treated, relative to a counterfactual of remaining untreated.

We denote the set of groups which start untreated and which first treatment lasts at least  $h$  periods as  $g^h$ . Therefore  $g^h$  includes all groups such that  $D_{g,1} = 0$  and  $D_{g,p_g^1+j} = 1$  for  $0 \leq j \leq h$ .

We let  $\tau_h^g$  denote the  $h$ -periods horizon effect of entering treatment for the first time for group  $g$ , which enters treatment for the first time at  $p_g^1$  and then remains treated until at least period  $t + h$ . This can be written as

$$\tau_h^g = E [y_{i,p_g^1+h}(\mathbf{D}_g) - y_{i,p_g^1+h}(\mathbf{0}) | \mathbf{D}_i = \mathbf{D}_g, g \in g^h]$$

The estimand under consideration renders this setting analogous to the case of absorbing treatment discussed in the previous section, and similar results follow. A variance-weighted average effect  $\sum_{g \neq 0, g \in g^h} \omega_{g,h}^{LP-DiD} \tau_h^g$ , with strictly positive weights

determined as in Equation 12, can be recovered by estimating the LP-DiD specification of Equation 10 and applying the following ‘clean control’ sample restriction

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } j \geq 1), \\ \text{or clean control} & D_{i,t-j} = 0 \text{ for } j \geq -h. \end{cases} \quad (15)$$

Estimating the same specification using either a weighted regression or regression adjustment (as discussed in Section 3.3) allows unbiased estimation of an equally-weighted effect.

#### 4.2.3 Estimating average treatment effects under an ‘effect stabilization’ assumption

In numerous settings of practical importance, the estimand  $\tau_h^g$  introduced above and the resulting ‘clean control’ condition of Equation 15 might not be feasible or appropriate. Consider, for example, the problem of estimating the effect of minimum wage increases in a panel of regions. For later time periods  $t$ , there will be very few regions that have never experienced a minimum wage increase until period  $t + h$ . In fact, most regions are likely to have experienced some increase in the minimum wage before the start of the sample period. This case can be dealt with in a simple way in the LP-DiD framework, by focusing on the average effect of treatment events and under the additional assumption that dynamic effects stabilize after a finite number of periods.

Let us first define treatment and exit events. If group  $g$  enters treatment at time  $j$  (i.e., if  $\Delta D_{gj} = 1$ ), we say that all units in group  $g$  have a *treatment event* at time  $j$ . Similarly, if  $\Delta D_{gj} = -1$ , then  $g$  experiences an *exit event* at time  $j$ . We let  $N_g^T$  denote the number of treatment events experienced by group  $g$  in the sample period, and  $N_g^E$  the number of exit events.

We then introduce the vectors of counterfactual treatment series  $\mathbf{D}_{g,j,0}$  and  $\mathbf{D}_{g,j,1}$ . We let  $\mathbf{D}_{g,j,0}$  denote a  $(1 \times T)$  vector which is equal to  $\mathbf{D}_g$  up to column  $j-1$ , while its columns  $j$  to  $T$  are all equal to 0. Similarly,  $\mathbf{D}_{g,j,1}$  is equal to  $\mathbf{D}_g$  up to column  $j-1$ , but its columns  $j$  to  $T$  are all equal to 1. We thus have  $\mathbf{D}_{g,j,0} = (D_{g1}, \dots, D_{g,j-1}, \mathbf{0})$  and  $\mathbf{D}_{g,j,1} = (D_{g1}, \dots, D_{g,j-1}, \mathbf{1})$ .

The average dynamic effect of the  $n$ -th treatment event experienced by group  $g$  is denoted by  $\tau_h^{g,n}$  and defined as follows<sup>33</sup>

<sup>33</sup>If follows from our notation that, if a group is treated at the start of the sample period ( $D_{g1} = 1$ ), then the  $n$ -th treatment event is the  $n-1$ -th treatment event experienced within the sample period. For example, if  $D_{g1} = 1$ , the 2nd treatment event is the first treatment event experienced during the sample period.

$$\tau_h^{g,n} = E[y_{i,p_g^n+h}(\mathbf{D}_{g,p_g^n,1}) - y_{i,p_g^n+h}(\mathbf{D}_{g,p_g^n,0}) | \mathbf{D}_i = \mathbf{D}_g]$$

The average dynamic effect of the  $n$ -th exit event experienced by group  $g$ , denoted by  $\eta_h^{g,n}$ , is defined similarly:

$$\eta_h^{g,n} = E[y_{i,q_g^n+h}(\mathbf{D}_{g,q_g^n,1}) - y_{i,q_g^n+h}(\mathbf{D}_{g,q_g^n,0}) | \mathbf{D}_i = \mathbf{D}_g]$$

We assume that treatment effects (including exit effects) stabilize after a finite number of periods. Formally, we introduce the following assumption.

**Assumption 9. Dynamic effects stabilize after  $L$  periods:**

$$\lambda_L^{g,n} = \lambda_{L+l}^{g,n} \quad \text{for } l \geq 0, \lambda \in \{\tau, \eta\} \text{ and all groups } g = 1, \dots, G.$$

Assumption 9 implies that, for any group  $g$ , any time horizon  $h$ , any event  $n$ , and any  $j \geq L + 1$ , we have  $\tau_{h+j}^{g,n} = \tau_{j-1}^{g,n}$  (or, in the case of exit events  $\eta_{h+j}^{g,n} = \eta_{j-1}^{g,n}$ ).

In this setting and under assumptions 7, 8 and 9, we have

$$\begin{aligned} E[y_{t+h} - y_{t-1}] &= E[y_{i,t+h}(\mathbf{0})] - E[y_{i,t-1}(\mathbf{0})] + E[y_{i,t+h}(\mathbf{D}_i) - y_{i,t+h}(\mathbf{0})] - E[y_{i,t-1}(\mathbf{D}_i) - y_{i,t-1}(\mathbf{0})] \\ &= \delta_t^h + \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^T} \left( \tau_h^{g,n} \times \mathbf{1}\{t = p_g^n\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t} = +1\} \\ &\quad + \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^E} \left( \eta_h^{g,n} \times \mathbf{1}\{t = q_g^n\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t} = -1\} \\ &\quad + \sum_{j=1}^L \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^T} \left( \left( \tau_{h+j}^{g,n} - \tau_{j-1}^{g,n} \right) \times \mathbf{1}\{t = p_g^n + j\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t-j} = +1\} \\ &\quad + \sum_{j=1}^L \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^E} \left( \left( \eta_{h+j}^{g,n} - \eta_{j-1}^{g,n} \right) \times \mathbf{1}\{t = q_g^n + j\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t-j} = -1\} \\ &\quad + \sum_{j=1}^h \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^T} \left( \tau_{h-j}^{g,n} \times \mathbf{1}\{t = p_g^n - j\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t+j} = +1\} \\ &\quad + \sum_{j=1}^h \left[ \sum_{g=1}^G \sum_{n=1}^{N_g^E} \left( \eta_{h-j}^{g,n} \times \mathbf{1}\{t = q_g^n - j\} \times \mathbf{1}\{\mathbf{D}_i = \mathbf{D}_g\} \right) \right] \times \mathbf{1}\{\Delta D_{i,t+j} = -1\} \end{aligned} \tag{16}$$

where  $\delta_t^h = \delta_{t+h} - \delta_{t-1}$  as in the previous sections.

Equation 16 differs from its absorbing treatment counterpart (Equation 8) in two main ways: first, the same group can be under the influence of multiple previous treatment and exit events; second (because of Assumption 9), only events that occur between periods  $t-L$  and  $t+h$  affect the expected value of the long difference  $y_{t+h} - y_{t-1}$ .

A convex weighted ATT for the effect of entering treatment and staying treated is then obtained by estimating an LP-DiD specification with the following sample restriction:

$$\left\{ \begin{array}{ll} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L), \\ \text{or clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L. \end{array} \right. \quad (17)$$

For example, if the effect of treatment is assumed to stabilize after 15 periods ( $L=15$ ), then one only needs to exclude from the estimation sample observations that experience a change in treatment status between  $t-15$  and  $t-1$  or between  $t+1$  and  $t+h$ .<sup>34</sup>

Also in this case, regression-adjustment can be used to estimate an equally-weighted ATT instead of a variance-weighted one, as described in Section 3.3.

In this subsection we have provided two brief simple illustrations, likely relevant in empirical applications, of what one can do with the LP-DiD framework in settings with non-absorbing treatment. Of course, one could consider several alternative target estimands and assumptions. A detailed and comprehensive discussion of non-absorbing treatment would indeed require a whole article in itself. Our point here is simply that the LP-DiD framework and its clean control condition can be flexibly adapted to accommodate different estimands of interest and identification assumptions in settings with non-absorbing treatment.

### 4.3 Continuous treatment

A formal discussion of the issues that can arise under continuous treatment is outside the scope of this paper, and could represent a fruitful avenue for further research. Being based on simple regression, the LP-DiD approach could be extended in future work to normalize the treatment effect depending on treatment intensity (or *dose*). The flexibility of the framework could be exploited to accommodate the different definitions of the causal effect of interest and the different identification assumptions that can be appropriate with continuous treatment. For example, in future work on this topic, the clean control condition might be adapted to define clean controls as ‘stayers’ (or alternatively ‘quasi-stayers’), in the terminology of [de Chaisemartin et al. \(2022\)](#).

---

<sup>34</sup>In their empirical application, [Callaway and Sant’Anna \(2020\)](#) study the impact of minimum wage increases during 2001–2007, and use as controls all states that did not raise their minimum wage during this period. However, all states (including the control states) were affected by the federal minimum wage increases in 1996–1997, and there were no truly untreated states during the 2001–2007 period. Therefore, there is an assumption in [Callaway and Sant’Anna \(2020\)](#) that  $L$  is no greater than 4 years, although this assumption is not made explicit.

## 4.4 Identification and relation to impulse responses

Circling back to our initial motivation, we end this discussion by briefly outlining the differences and commonalities between the LP-DiD methods that we propose in this paper and the now-typical estimation of impulse responses by LPs in macroeconomics.

Perhaps the key difference is in the definition of the counterfactual experiment. In a traditional macro impulse response, treatment (the ‘shock’ in macro parlance), typically generates a series of later changes in the policy variable itself (i.e., subsequent treatments) as well as changes in the outcome. In a sense, the experiment is akin to a *treatment plan* rather than a one-off treatment, as is traditional in applied micro.

Note that in the specification of the LP-DiD estimator, we condition on future values of treatment (between  $t + 1$  and  $t + h$ ). This removes the effect of subsequent treatment effects on future values of the outcome. It should be obvious that one can recover the impulse response by the convolution of the treatment plan with the single-treatment effect measured with the LP-DiD estimator. That is, our estimator computes the treatment effect of a one-off intervention. If the intervention itself then generates subsequent interventions, the overall effect—the impulse response—is the result of combining one-off treatment effects with the treatment plan itself.

Is one approach more correct than the other? As [Alloza, Gonzalo, and Sanz \(2019\)](#) show, not really. The impulse response captures the effect of an intervention on an outcome that is the most likely to be seen directly in the data, allowing for the path of future treatments. The researcher is less interested in the sequence of individual treatment effects on the outcome generated by the treatment plan. Rather the goal is to understand the overall effect on the outcome over time. In the applied-micro setting for which our LP-DiD estimator is constructed, we are instead careful to parse out the one-off effect. This object is of equal value scientifically as it would permit the researcher to craft an alternative treatment plan than that usually observed (though in that case, deviations from the usual treatment plan can run afoul of the Lucas critique if they are ‘too different’).

There are two important caveats to these statements. First, conditioning on future treatments is not innocuous if treatment assignment is endogenous. Second, even if treatments are exogenous (perhaps conditional on observables), the extent to which the results can be interpreted as measures of one-off treatments when treatment is not absorbing greatly depends on how agents form expectations about future treatments. A one-off treatment will likely be a significant departure from previously observed treatment plans and thus lead forward-looking agents to respond differently.



## 5 Monte Carlo simulation

In this section, we conduct a Monte Carlo simulation to evaluate the performance of the LP-DiD estimator. We consider a binary staggered treatment, with dynamic and heterogeneous treatment effects, and assume unconditional parallel trends.

Unlike conventional TWFE (which in this setting suffers from negative weights bias), LP-DiD tracks well the true effect path even in the presence of heterogeneity. We find that LP-DiD performs in a way similar to the [Sun and Abraham \(2020\)](#), [Callaway and Sant’Anna \(2020\)](#) and [Borusyak, Jaravel, and Spiess \(2021\)](#) estimators, while being computationally faster.

### 5.1 Setting

Our simulated datasets include  $N = 500$  units, observed for  $T = 50$  time periods. The counterfactual outcome  $y_{it}(0)$  that a unit would experience if not treated is given by

$$y_{it}(0) = \rho y_{i,t-1}(0) + \lambda_i + \gamma_t + \epsilon_{it},$$

with  $\rho = 0.5$ , and with  $\lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25)$ .

Treatment is binary, staggered and absorbing. The treatment effect is positive and grows in time for 20 time periods, after which it stabilizes. Moreover, early adopters have larger treatment effects. Specifically, the treatment effect is given by

$$\beta_{it} = \begin{cases} 0 & \text{if } t - p_i < 0, \\ \alpha_0(t - p_i + 1) + \alpha_1(t - p_i + 1)^2 + (1 - \alpha_1) \frac{(t - p_i + 1)^2}{(p_i/p_1)^2} & \text{if } 0 \leq t - p_i \leq 20, \\ \alpha_0 21 + \alpha_1 21^2 + (1 - \alpha_1) \frac{21^2}{(p_i/p_1)^2} & \text{if } t - p_i > 20, \end{cases}$$

where  $p_i$  is the period in which unit  $i$  enters treatment as in the previous sections and  $p_1$  is the treatment period for the ‘earliest adopter’ in the sample. We set  $\alpha_0 = 2$  and  $\alpha_1 = 0.5$ . Observed outcomes  $y_{it}$  are therefore given by  $y_{it} = y_{it}(0) + \beta_{it}$ .

Units are randomly assigned to 10 groups, each of size  $N/10$ . One group never receives treatment; the other nine groups receive treatment respectively at time  $p = 11, 13, 15 \dots, 27$ .

### 5.2 Estimators

We perform 200 replications and evaluate five estimators:

- A conventional event-study TWFE specification ([Equation 4](#)).

- The LP-DiD estimator (Equation 10), using both simple (variance-weighted) LP-DiD and re-weighted LP-DiD, and applying both first-lag differencing and pre-mean differencing.<sup>35</sup>
- The Sun and Abraham (2020) estimator (SA hereafter).
- The Callaway and Sant’Anna (2020) estimator (CS hereafter).
- The Borusyak, Jaravel, and Spiess (2021) estimator (BJS hereafter).

For each estimator, we compare the distribution of the estimated ATE with the (equally-weighted) true ATE.

### 5.3 Results

Results are presented in Figures 1 and 2 and Table 1. Figure 1 displays the estimated effect path in comparison with the true (equally-weighted) average effect path and the full range of heterogeneous group-specific effects. Figure 2 plots the full distribution of the estimates at time-horizons 0,5,10 and -2. Table 1 reports the root mean squared error (RMSE) of each estimator at time horizons between -5 and +10.

Conventional event-study TWFE does a poor job in this setting, due to the heterogeneity of treatment effects. It finds a non-existent decreasing pre-trend and grossly underestimates the dynamic treatment effect. Due to negative weighting, point estimates lie outside the full range of true group-specific effects.<sup>36</sup>

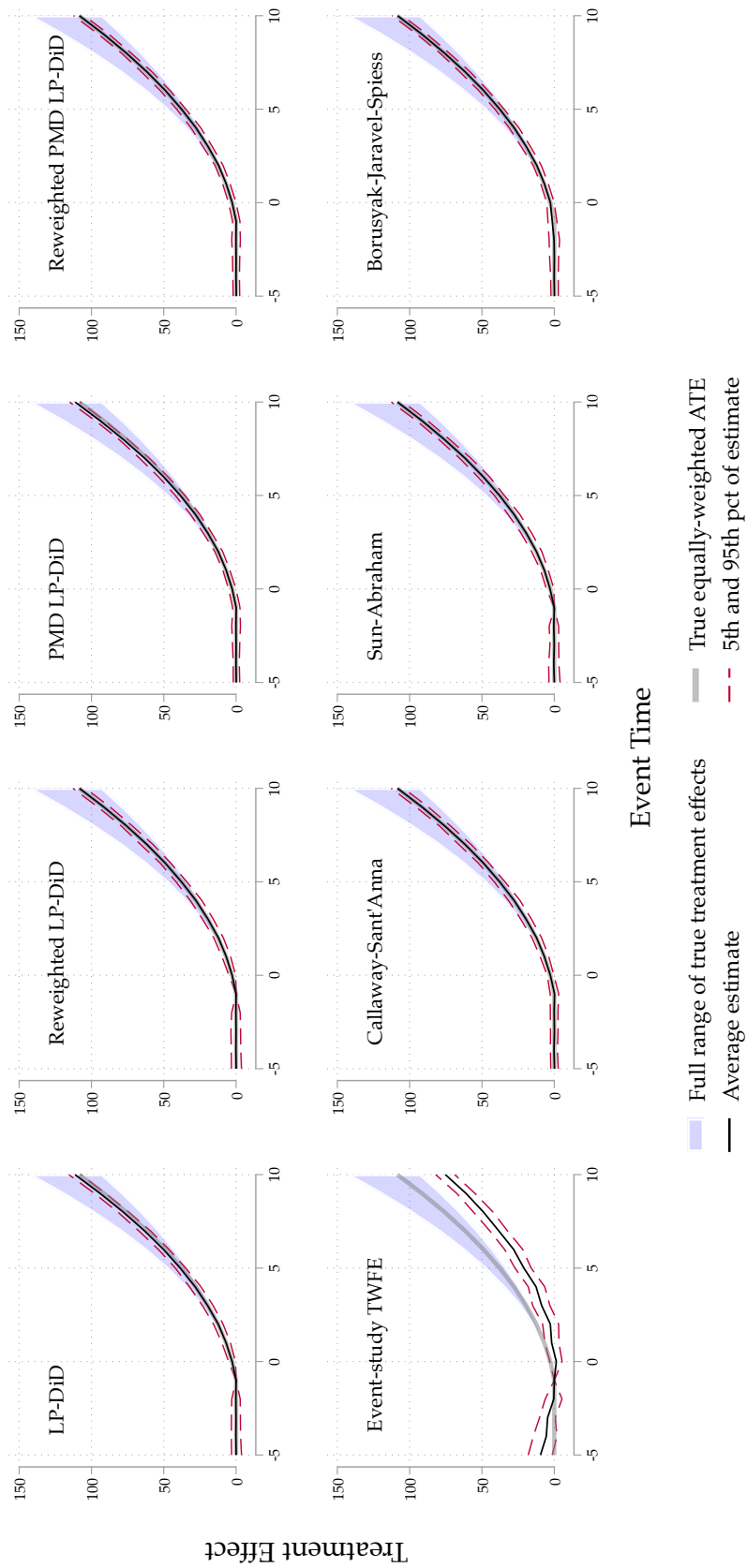
The LP-DiD estimator, instead, tracks well the true average dynamic effect. This is true of both simple (variance-weighted) LP-DiD and re-weighted LP-DiD, and of both pre-mean differencing (PMD) and first-lag differencing.

Results are suggestive of a bias-variance trade-off between equal-weighting and variance-weighting: reweighted LP-DiD (which applies equal weights) is unbiased, while simple (variance-weighted) LP-DiD has a small bias but also lower variance. The bias from variance-weighting tends to become more relevant at longer time-horizons, because in our simulation the true treatment effect variance increases in time after treatment.

<sup>35</sup>As discussed in Section 3.2, the variance-weighted version recovers a convex combination of cohort specific effects, with weights given by Equations 11 and 12. The re-weighted version recovers an equally-weighted ATT; it can be implemented through a weighted least squares regression or using regression adjustment. See Section 3.5 for the discussion of first-lag differencing versus pre-mean differencing.

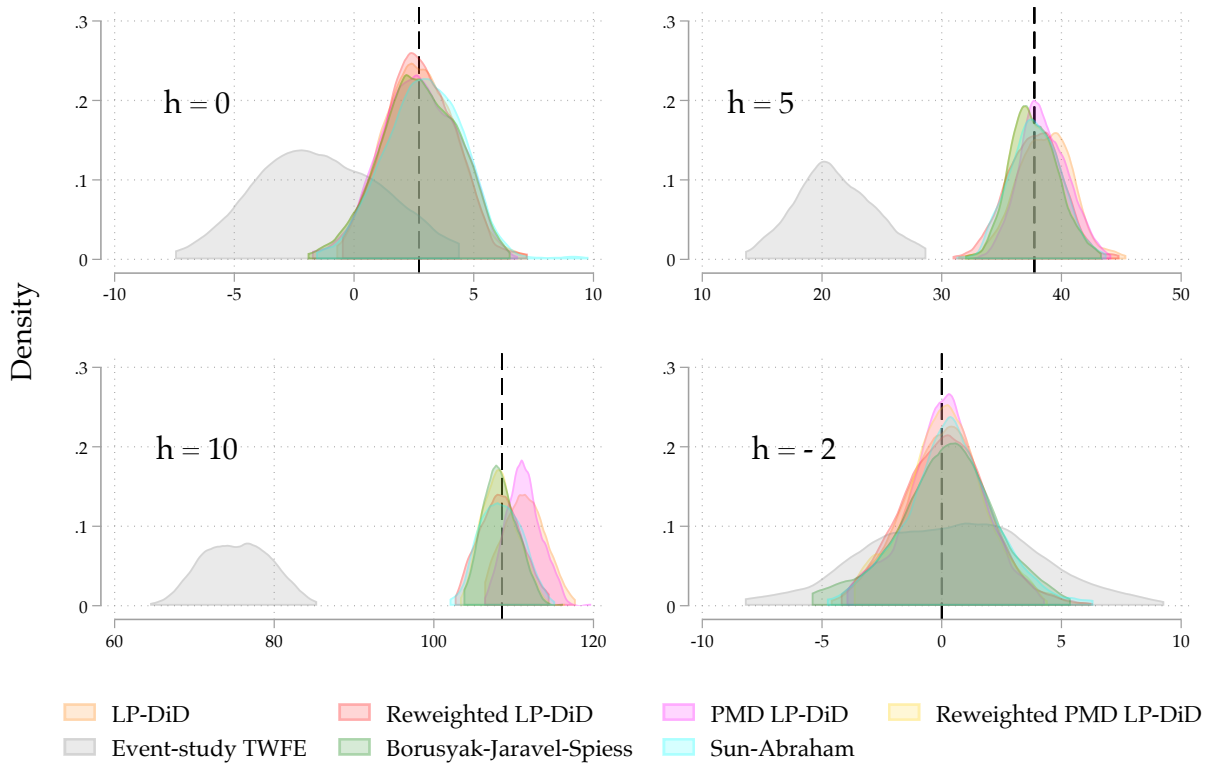
<sup>36</sup>The fact that in our simulated DGP the size of the effect is a function of the date of treatment makes the ‘negative weighting’ problem particularly severe, and therefore the performance of the TWFE specification particularly poor. We choose this DGP in order to test the performance of our estimator in a setting in which the flaws of the conventional estimator are particularly severe.

Figure 1: Simulation results: True effect path and estimates



Notes: Average estimates and 95% and 5% percentiles from 200 replications.

Figure 2: Simulation results: Distribution of estimates



Notes: Distribution of estimates from 200 replications. The black vertical dashed line is the true (equally-weighted) average treatment effect on the treated. The Callaway and Sant’Anna (2020) estimator is not included, because in this setting it is numerically equivalent to Reweighted LP-DiD.

As a result, variance-weighting produces smaller RMSE than equal weighting at short time-horizons ( $h = 1$  to 3), but the opposite is true at longer time horizons (Table 1).

Pre-mean-differencing (PMD) has lower RMSE than first-lag differencing, as expected when parallel trends holds fully and at all time periods (see discussion in Section 3.5).

When comparing LP-DiD with the SA, CS and BJS estimators, we find that it performs similarly. In particular, the reweighted and pre-mean-differenced version of LP-DiD outperforms CS and SA and is very close to BJS. BJS has slightly lower RMSE than reweighted PMD LP-DiD at longer post-treatment time horizons, but the reverse applies

Table 1: Simulation results: Root mean squared error (RMSE)

Event time	ES TWFE	LP-DiD	Rw LP-DiD	PMD LP-DiD	Rw PMD LP-DiD	SA	CS	BJS
-5	10.83	2.05	2.11	1.31	1.35	2.17	1.61	1.54
-4	7.54	1.90	1.97	1.33	1.37	2.19	1.51	1.63
-3	5.70	1.93	1.95	1.47	1.52	1.83	1.54	1.92
-2	3.50	1.79	1.81	1.60	1.62	1.96	1.60	2.08
0	4.86	1.48	1.46	1.58	1.59	1.70	1.46	1.59
1	5.80	1.83	1.85	1.55	1.58	1.68	1.85	1.58
2	10.27	2.09	2.14	1.63	1.68	2.18	2.14	1.67
3	11.22	2.40	2.41	1.89	1.89	2.25	2.41	1.89
4	15.68	2.39	2.36	2.04	1.99	2.42	2.36	1.98
5	17.03	2.47	2.33	2.14	2.00	2.14	2.33	1.99
6	21.22	2.67	2.38	2.33	2.07	2.39	2.38	2.04
7	23.60	2.90	2.45	2.57	2.12	2.37	2.45	2.09
8	27.19	3.12	2.31	2.85	1.98	2.36	2.31	1.92
9	30.65	3.56	2.51	3.28	2.14	2.41	2.51	2.10
10	33.50	3.79	2.57	3.53	2.25	2.62	2.57	2.18

Notes: RMSE from 200 replications. ES TWFE = event-study two-way-fixed-effects; Rw = reweighted; PMD = pre-mean-differenced; CS = Callaway and Sant’Anna (2020); SA = Sun and Abraham (2020); BJS = Borusyak et al. (2021).

to testing for pre-trends (i.e., pre-treatment time horizons).<sup>37</sup> Because of the variance-bias trade-off discussed above and of the advantages of pre-mean differencing in this setting, at shorter time-horizons ( $h = 1$  to  $3$ ) the variance-weighted and PMD version of LP-DiD produces the lowest RMSE among all the estimators employed (Table 1).

<sup>37</sup>As discussed in Section 2, in this setting with binary absorbing treatment, no covariates, and using only not-yet treated observations as controls, the re-weighted LP-DiD estimator with first-lag differencing is numerically equivalent to the CS estimator, and therefore has exactly the same RMSE and empirical standard error. This holds at all post-treatment time-horizons ( $h \geq 0$ ), but not in the estimation of pre-trends ( $h < 0$ ) because pre-trends tests are constructed differently by the CS estimator. As discussed in Section 3.5, reweighted PMD LP-DiD would be numerically equivalent to BJS in the special case of only one treated group, and is very similar in more general settings. The fact that in this simulation reweighted PMD LP-DiD has lower RMSE than BJS at pre-treatment time periods is not inconsistent with BJS’s demonstration of efficiency, because that demonstration assumes no autocorrelation, while in this simulation there is some autocorrelation.

## 5.4 Computational speed

We also find that our LP-DiD estimator provides a substantial computational advantage relative to other recently proposed estimators. To quantify this, we recorded the computation time required for estimating the treatment effect path in our synthetic datasets. For these exercises, the estimations were conducted using STATA software on a laptop with 2.80 GHz Quad-core Intel i7 Processor and 16 GB of RAM. Recorded computation times are reported in [Table 2](#).

Table 2: Computational speed (seconds)

ES TWFE	LP-DiD	PMD LP-DiD	Rw LP-DiD	Rw PMD LP-DiD	CS	SA	BJS
.59	.74	.80	1.59	1.64	79.25	177.71	7.08

Notes: Computation times in a single repetition of the simulated datasets described in Section 5, measured in seconds. Recorded on a laptop with 2.80 GHz Quad-core Intel i7 Processor and 16 GB of RAM, using the STATA software. Rw = reweighted (see Sec 3.3); PMD = pre-mean-differenced (see Sec 3.5); CS = [Callaway and Sant’Anna \(2020\)](#); SA = [Sun and Abraham \(2020\)](#); BJS = [Borusyak, Jaravel, and Spiess \(2021\)](#).

The LP-DiD estimator runs in 0.7 seconds per repetition, similar to the (biased) event-study TWFE estimator, while reweighted LP-DiD (implemented through weighted regression) requires around 1.6 seconds. The BJS estimator requires approximately 7 seconds with the CS and SA demanding respectively 79 and 178 seconds.

## 6 Empirical Applications

We present two empirical applications to illustrate the use of the LP-DiD estimator.

### 6.1 Credit and the labor share

In our first empirical study, we estimate the effect of banking deregulation laws on the labor share in US States, replicating [Leblebicioğlu and Weinberger \(2020\)](#), LW hereafter.

From the late 1970s up to the 1990s, U.S. states lifted restrictions on the ability of out-of-state banks to operate in-state (*interstate banking* deregulation) and on the ability of in-state banks to open new branches (*intra-state branching* deregulation). LW estimate the effects of both inter-state and intra-state banking deregulation laws on the labor share of

value added. They find inter-state banking deregulation had a sizable negative effect on the labor share, but find no effect of intra-state branching deregulation.

The dataset covers the 1970–1996 period. (In 1997, inter-state banking deregulation was imposed in all states by federal law.) Appendix [Figure B.1](#), which reproduces [Figure 1](#) in LW, displays the share of US states with a liberalized banking sector.

### 6.1.1 Conventional TWFE specifications

We first consider the following static TWFE specification for the effect of banking deregulation laws, which replicates LW’s baseline specification:

$$LS_{st} = \beta_{Bank}Bank_{st} + \beta_{Branch}Branch_{st} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st}, \quad (18)$$

where  $s$  indexes states,  $t$  indexes years, and  $LS$  is the labor share.  $Branch_{st}$  and  $Bank_{st}$  are binary indicators equal to one if a state has adopted intrastate branching or interstate banking deregulation.

To assess possible pre-trends and lagged effects, LW also estimate the following event-study TWFE specification:<sup>38</sup>

$$LS_{st} = \sum_{q=-9}^9 \beta_{Bank,t+q} \Delta Bank_{s,t+q} + \sum_{q=-9}^9 \beta_{Branch,t+q} \Delta Branch_{s,t+q} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st}. \quad (19)$$

### 6.1.2 Forbidden comparisons in the TWFE specifications

Given the staggered rollout of banking deregulation laws across US states, the TWFE specifications of [Equations 18](#) and [19](#) might suffer from the negative weights bias highlighted by recent studies and discussed in [Section 2](#). Earlier liberalizers are implicitly used as controls for states that liberalize later on. Specifically, the specifications in [Equations 18](#) and [19](#) produce a weighted average of two types of 2x2 comparisons: (1) ‘clean’ comparisons of newly treated states vs. not-yet treated states and (2) ‘unclean’ comparisons of newly treated states vs. earlier-treated states ([Goodman-Bacon, 2021](#)).

---

<sup>38</sup>The event-study TWFE specification employed by [Lelebicioğlu and Weinberger \(2020\)](#) is not completely standard, since it includes only leads and lags of the differenced treatment indicators but not a last lag in levels. Therefore it is not completely equivalent to the standard event-study TWFE estimator as obtained by estimating [equation 4](#) above. This non-standard specification, however, does not influence results: applying a standard event-study TWFE specification (as in our [equation 4](#)) yields very similar results as those obtained by [Lelebicioğlu and Weinberger \(2020\)](#). This reassures us that any differences between the dynamic TWFE results of [Lelebicioğlu and Weinberger \(2020\)](#) and our results from applying LP-DiD are due to the negative weights bias of TWFE, not to the non-standard specification of the TWFE model used by [Lelebicioğlu and Weinberger \(2020\)](#).

We employ the [Goodman-Bacon \(2021\)](#) diagnostic to decompose the TWFE estimate from [Equation 18](#) into these two types of comparisons.<sup>39</sup> The static TWFE estimator of the effect of interstate banking deregulations assigns a 36% weight to unclean comparisons. For the estimates of the effect of intrastate branching deregulations, the problem is much more severe: unclean comparisons account for as much as 70% of the estimate. [Appendix Figure B.2](#) displays the results of the [Goodman-Bacon \(2021\)](#) decomposition diagnostic, showing how the estimates of the effects of intrastate branching deregulations are driven by a few ‘unclean’ comparisons – those involving states that deregulated before 1970 – that receive a very large weight. Notably, for both types of policies, clean comparisons produce overwhelmingly negative coefficients, while the unclean ones tend to bias the coefficients upwards. The greater bias in the case of intrastate branching—whose adoption is spread over a much longer horizon than interstate banking—is a stark demonstration of the negative weighting problem that arises with staggered treatment.

### 6.1.3 LP-DiD specification

We revisit the effect of banking deregulation laws using the following LP-DiD specification:<sup>40</sup>

$$LS_{s,t+h} - LS_{s,t-1} = \alpha_t^h + \beta_h^{LP-DiD} \Delta Policy_{s,t} + \sum_{m=1}^M \eta_m^h X_{s,t-m} + e_{s,t}^h, \quad (20)$$

restricting the sample to observation that are either:

$$\begin{cases} \text{newly treated} & \Delta Policy_{s,t} = 1, \\ \text{or clean control} & Policy_{s,t+h} = 0, \end{cases} \quad (21)$$

with  $Policy \in \{Bank, Branch\}$ , and where the vector of control variables  $X$  can also include lags of the (differenced) outcome variable,  $\Delta LS_{s,t-m}$ .

### 6.1.4 Results

We begin by presenting some replication results that use the specifications in the original LW study. [Figure 3](#) displays results from the static and event-study TWFE specifications of [Equations 18](#) and [19](#). These replicate the estimates reported in [Table 2](#) and [Figure 2](#)

<sup>39</sup>Specifically, the [Goodman-Bacon \(2021\)](#) diagnostic is applied to a basic static TWFE specification that only includes two-way fixed effects and the treatment indicator for the policy under consideration. This corresponds to the first specification (‘only FEs’) reported in the top panel of [Figure 3](#).

<sup>40</sup>Given that treatment is absorbing in this data, and there is a sufficient number of not-yet treated States at all points in time, we use the version of the clean control condition with only not-yet treated as controls.



of LW. They suggest that the liberalization of inter-state banking has a sizable negative effect on the labor share, although they also show a small pre-treatment trend. Instead, the estimated effects of intra-state branching deregulation on the labor share are positive, small and very imprecise.

Next consider the LP-DiD estimator with clean controls of [Equation 21](#), whose results are displayed in [Figure 4](#). The negative effect of inter-state banking deregulation on the labor share is confirmed, including when controlling for pre-treatment outcome dynamics.

Estimates of the effect of intra-state branching deregulation, instead, change dramatically: while the original TWFE estimates found no effect, we find a sizable negative impact. After addressing the bias of the TWFE estimator by excluding ‘unclean’ comparisons, the estimated effect of inter-state branching deregulation on the labor share is negative and of similar size as that of inter-state banking deregulation. Both types of deregulation are now found to make a difference.

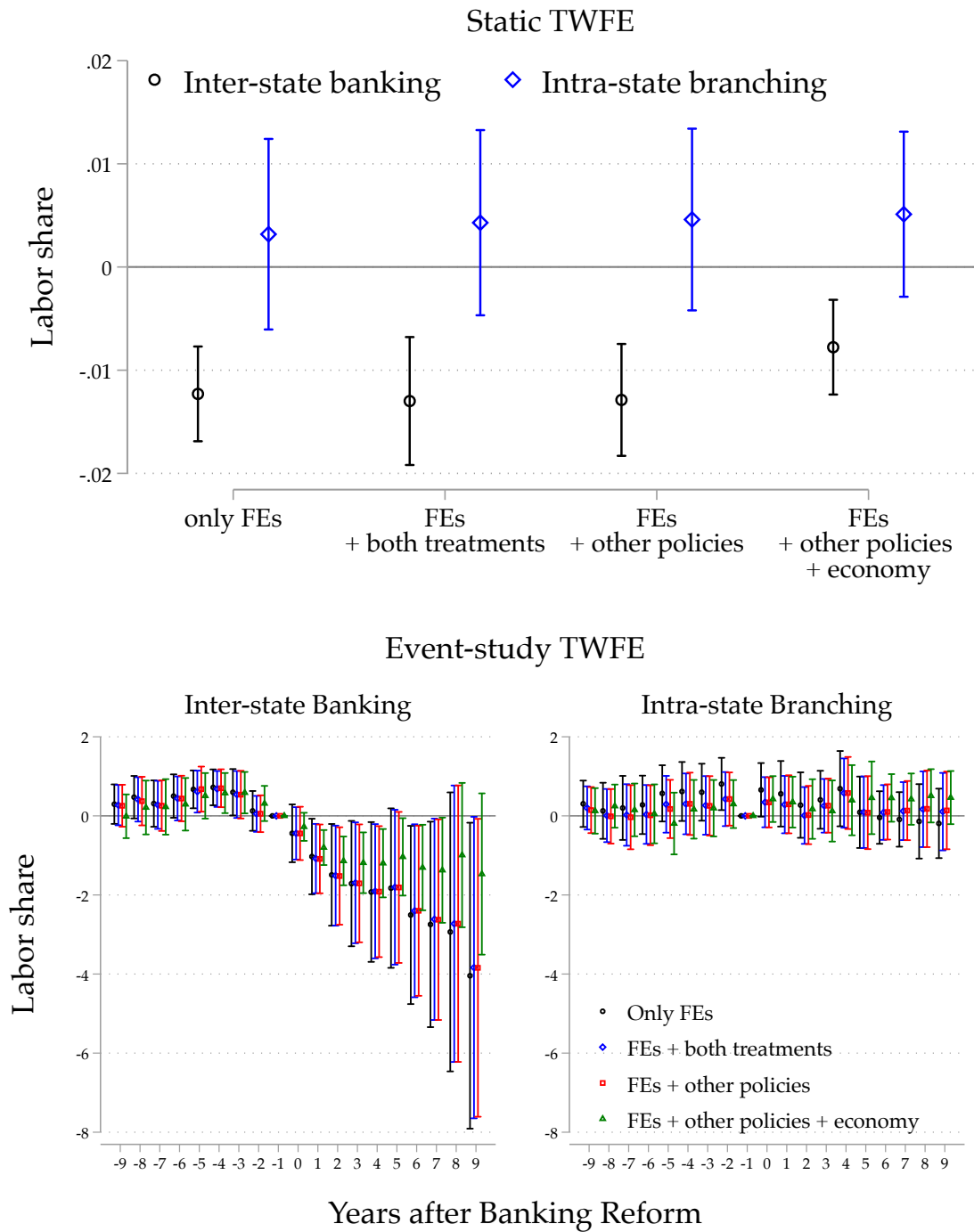
## 6.2 Democracy and economic growth

Our second empirical application estimates the effect of democracy on GDP per capita, replicating the analysis in [Acemoglu, Naidu, Restrepo, and Robinson \(2019\)](#), ANRR hereafter.

The dataset covers 175 countries from 1960 to 2010. The treatment indicator is a binary measure of democracy, which ANRR build from several datasets to mitigate measurement error. The main outcome variable of interest is the log of GDP per capita, obtained from the World Bank Development Indicators.

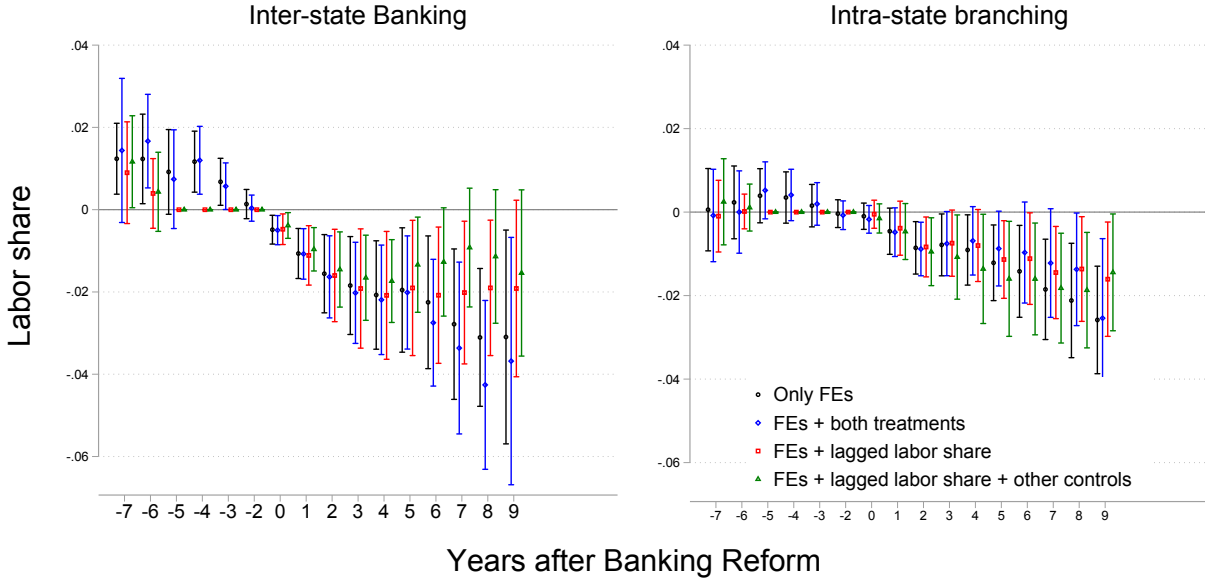
Three features make this application an especially meaningful testing ground for the LP-DiD approach. First, there is potential for negative weighting: fixed-effects regression would implicitly use older democracies as controls for new democracies. Second, treatment is non-absorbing: democracies can slide back into autocracy, and there are indeed multiple instances of reversals in the data. Third, controlling for pre-treatment outcome dynamics is crucial, since there is evidence of selection: ANRR show that democratisation tends to be preceded by a dip in GDP per capita.

Figure 3: Effect of banking deregulation on the labor share: TWFE estimates



Notes: Estimates for the effect of banking deregulation on the labor share, using data from [Lelebicioğlu and Weinberger \(2020\)](#). The top panel uses the static TWFE specification of equation 18. The two bottom panels use the event-study TWFE specification of equation 19. These graphs replicates results from [Lelebicioğlu and Weinberger \(2020\)](#). Following [Lelebicioğlu and Weinberger \(2020\)](#), the additional controls are four lags of real State GDP, average corporate tax rate, and union membership rates.

Figure 4: Effect of banking deregulation on the labor share: LP-DiD estimates



Notes: Estimates for the effect of banking deregulation on the labor share, using data from [Lelebicioğlu and Weinberger \(2020\)](#) and the LP-DiD specification of equations 20 and 21. Following [Lelebicioğlu and Weinberger \(2020\)](#), the additional controls are four lags of real State GDP, average corporate tax rate, and union membership rates.

### 6.2.1 Dynamic panel specifications

The baseline results in ANRR are obtained from the following dynamic fixed effects specification:

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^p \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct}, \quad (22)$$

where  $c$  indexes countries,  $t$  indexes years,  $y$  is the log of GDP per capita and  $D$  is the binary measure of democracy.

Lags of GDP per capita are included to address selection bias, and in particular the pre-democratization decline in GDP per capita.

The estimated coefficients from [Equation 22](#) are then used to build an impulse response function (IRF) for the dynamic effect on GDP. These estimates also allow one to derive the cumulative long-run effect of a permanent transition to democracy, which can be estimated as  $\hat{\beta}(1 - \sum_{j=1}^p \hat{\gamma}_j)^{-1}$ .

Appendix [Figure C.3](#) displays the IRF from the estimation of the dynamic panel model of [Equation 22](#). This reproduces the baseline results in ANRR. The implied long-run effect of democracy on growth is 21 percent with a standard error of 7 percent. This

dynamic fixed effects specification, however, might suffer from the negative weighting bias discussed in Section 2.

### 6.2.2 LP-DiD specifications

Consider an LP-DiD specification for estimating the effect of democracy on growth,

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP \text{ DiD}} \Delta D_{ct} + \delta_t^h + \sum_{j=1}^p \gamma_j^h y_{c,t-j} + \epsilon_{ct}^h, \quad (23)$$

restricting the estimation sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1; D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L, \\ \text{clean controls} & D_{i,t-j} = 0 \text{ for } 0 \leq j \leq L. \end{cases} \quad (24)$$

In words, in each year  $t$  treated units are countries that democratize at  $t$  and have experienced no other change in treatment status in the previous  $L$  years; clean controls are countries that have been non-democracies continually for at least  $L$  years.

This is an example of how the LP-DiD framework can be applied in a setting in which treatment is not absorbing, and the clean control condition can (and should) be tailored to the specific application. For example, this specification does not condition inclusion in the estimation sample on treatment status between time  $t+1$  and  $t+h$ . This is to take into account the concern that, under endogenous selection into treatment, constraining future treatment status might introduce bias (see ANRR, pp. 54–55, for a discussion).

Moreover, similar to ANRR, only non-democracies are included in the control group, although in principle under Assumption 9 also countries that are continually democracies from  $t-L$  to  $t$  could be included. This reflects the concern that established democracies might not be a good comparison for new democracies, therefore a control group composed only of continuing autocracies is more likely to satisfy the parallel trends assumption.

In a section of their analysis, ANRR employ a semiparametric LP specification that can be seen as a special version of the LP-DiD estimator above. Specifically, they estimate Equations 23–24 with  $L = 1$ , which implies a time-window for defining clean controls of just 1 year.

Seeing ANRR’s semiparametric specification as a version of LP-DiD provides a useful novel perspective on their analysis and suggests possible deviations from their specification. Our formal analysis in Section 4 makes clear that their choice relies on an implicit (and unintended) assumption that treatment effects stabilize after 1 year, which is clearly

too strong in this setting.<sup>41</sup>

We thus estimate the LP-DiD specification of Equations 23–24 with a time-window of 20 years ( $L = 20$ ) for defining clean controls, thus excluding observations that have experienced some transition in the previous 20 years.

We also test robustness to excluding countries that democratize between  $t + 1$  and  $t + h$  from the control group. To do this, we adopt a second version of the clean control condition, in which treated units are defined as in condition 24 but clean controls are observations with  $D_{i,t-j} = 0$  for  $-h \leq j \leq L$ .<sup>42</sup> This test, however, should be interpreted with caution: in this setting conditioning on future treatment status might introduce bias.

### 6.2.3 Results

Figure 5 displays results from LP-DiD specifications (Equations 23–24). We present four specifications: The first (top left panel) follows ANRR and sets a time-window of just one year for defining clean controls ( $L = 1$ ). The second (top right) uses a time-window of 20 years ( $L = 20$ ). The third (bottom left) adds the additional requirement that control units remain non-democracies between  $t + 1$  and  $t + h$ . The fourth (bottom right) estimates the LP-DiD specification using regression adjustment (RA) to obtain an equally-weighted (rather than a variance-weighted) ATT, as discussed in Section 3.3.

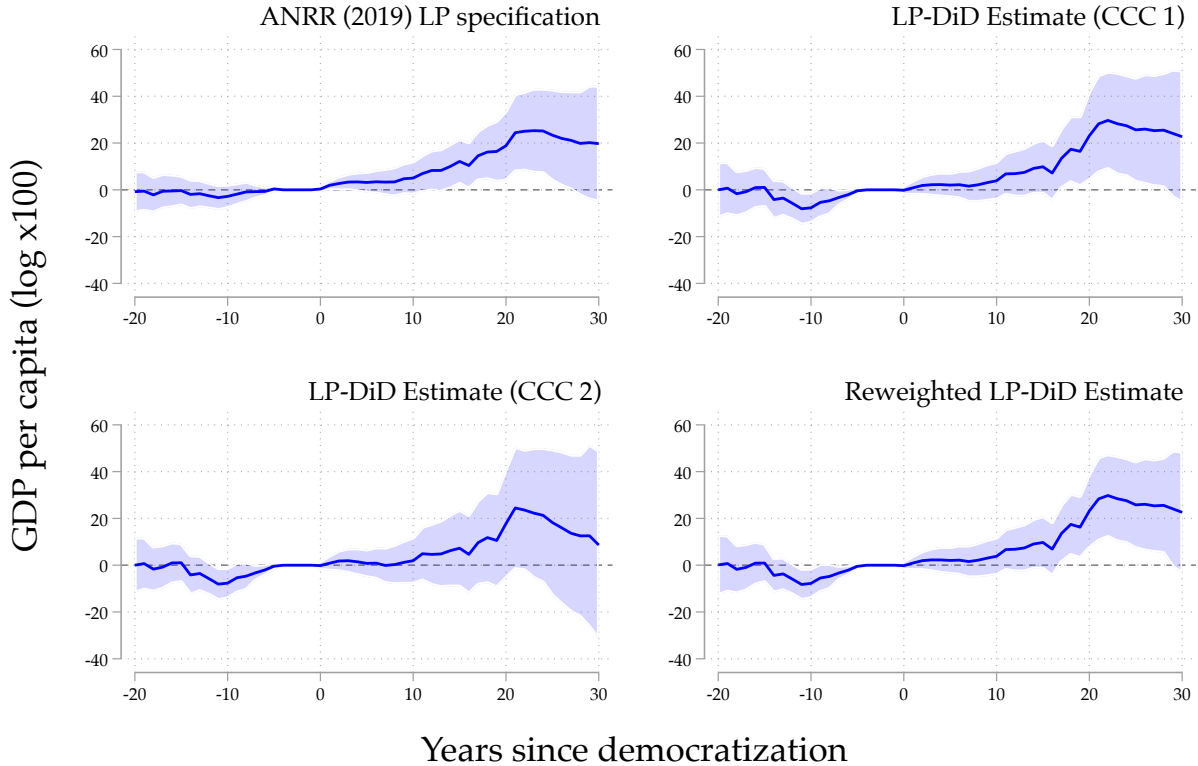
Overall, the result of a positive and large effect of democracy on GDP per capita appears robust to stricter definitions of the control group. The partial exception is the specification that excludes countries that democratize between  $t + 1$  and  $t + h$  from the control group, which finds similar positive short- and medium-term effects, but much smaller and very imprecise long-term effects. However, that specification is to be interpreted with caution, since constraining treatment status between  $t$  and  $t + h$  could introduce a form of selection bias.

---

<sup>41</sup>For example, consider Argentina, which democratized in 1973 and became a dictatorship again in 1976. The ANRR approach means that Argentina contributes to the counterfactual for measuring the effect of the 1978 democratization of Spain (among others). It seems natural to consider an alternative specification that excludes Argentina from the counterfactual for countries that (like Spain) democratize shortly after 1973–1976, reflecting the concern that the country might have experienced prolonged dynamic effects from the 1973–1976 transitions in and out of democracy.

<sup>42</sup>For example, in measuring the effect of the 1978 democratization of Spain on subsequent GDP dynamics, ANRR allow Ecuador, which was a nondemocracy in 1977 and 1978 but democratized in 1979, to be part of the control group. Instead, using the above version of the clean control condition we exclude Ecuador from the control group for the 1978 democratization of Spain.

Figure 5: Effect of democracy on growth: LP-DiD estimates



Notes: LP-DiD estimates for the effect of democracy on GDP per capita, using the specification of Equations 23-24. Top left panel ('ANRR (2019) LP specification') replicates results in Section IV of [Acemoglu et al. \(2019\)](#), which set  $L = 1$  year in Equations 23-24. The other three panels set  $L = 20$  years. CCC 1 is a clean control condition that defines treated units as countries that democratize in year  $t$  and experienced no transition between  $t-20$  and  $t-1$ , and clean controls as countries that are continually non-democracies between  $t-20$  and  $t$ . CCC 2 defines treated units in the same way, but clean controls are continually non-democracies between  $t-20$  and  $t+h$ . Right bottom panel uses reweighting to obtain an equally-weighted effect, using CCC 1. See main text for more details.

## 7 Conclusion

We propose a simple, transparent, easy, and fast technique for difference-in-differences estimation with dynamic heterogeneous treatment effects. Our LP-DiD estimator has several advantages and provides an encompassing framework, which can be flexibly adapted to address a variety of settings. It does not suffer from the negative weighting problem, and indeed can be implemented with any weighting scheme the investigator desires. Simulations demonstrate the good performance of the LP-DiD estimator and empirical exercises illustrate its use.

## References

- Daron Acemoglu, Suresh Naidu, Pascual Restrepo, and James A. Robinson. Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100, 2019.
- Mario Alloza, Jesus Gonzalo, and Carlos Sanz. Dynamic effects of persistent shocks. Banco de España Working Paper 1944, 2019.
- Joshua Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, N.J., 2009.
- Andrew C. Baker, David F. Larcker, and Charles C.Y. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022. doi: <https://doi.org/10.1016/j.jfineco.2022.01.004>. URL <https://www.sciencedirect.com/science/article/pii/S0304405X22000204>.
- Iavor Bojinov, Ashesh Rambachan, and Neil Shephard. Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics*, 12(4):1171–1196, 2021. doi: <https://doi.org/10.3982/QE1744>.
- Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. <https://arxiv.org/abs/2108.12419>, August 2021.
- Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2020.
- Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3):1405–1454, 2019.
- Clément de Chaisemartin and Xavier D'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, 2020.
- Clément de Chaisemartin, Xavier D'Haultfœuille, Félix Pasquier, and Gonzalo Vazquez-Bare. Difference-in-differences estimators for treatments continuously distributed at every period. <https://arxiv.org/abs/2201.06898>, January 2022.
- Clément de Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects, March 2022. URL <http://www.nber.org/papers/w29873>.
- Clément de Chaisemartin and Xavier D'Haultfœuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, 06 2022. doi: 10.1093/ectj/utac017.
- John Gardner. Two-stage differences in differences. [https://jrgcmu.github.io/2sdd\\_current.pdf](https://jrgcmu.github.io/2sdd_current.pdf), April 2021.
- Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.
- Nikolaj A Harmon. Difference-in-differences and efficient estimation of treatment effects. Technical report, Working paper, 2022.

- Òscar Jordà. Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182, 2005.
- Aslı Leblebicioğlu and Ariel Weinberger. Credit and the labour share: Evidence from US states. *Economic Journal*, 130(630):1782–1816, 2020.
- Michelle Marcus and Pedro H. C. Sant’Anna. The role of parallel trends in event study settings: An application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, 8(2):235–275, 2021.
- James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986. doi: [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6).
- Jonathan Roth. Interpreting event-studies from recent difference-in-differences methods, 2024.
- Jonathan Roth, Pedro H.C. Sant’Anna, Alyssa Bilinski, and John Poe. What’s trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023. doi: <https://doi.org/10.1016/j.jeconom.2023.03.008>. URL <https://www.sciencedirect.com/science/article/pii/S0304407623001318>.
- Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2020.
- Davide Viviano and Jelena Bradic. Dynamic covariate balancing: estimating treatment effects over time with potential local projections, 2023.
- Jeffrey M. Wooldridge. Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. <http://dx.doi.org/10.2139/ssrn.3906345>, August 2021.



# Appendix

## A Weights of the LP-DiD estimator

This appendix derives the weights assigned to each cohort-specific ATET by the LP-DiD estimator, first in a baseline version without control variables (Equations 11 and 12 in the main text) and then in more general specifications with control variables.

### A.1 Baseline version without control variables

#### Assumptions about the DGP

Consider the general setup and notation introduced in Section 2.1 in the main text. Treatment is binary, staggered and absorbing; parallel trends and no anticipation hold unconditionally (Assumptions 1 and 2); potential outcomes without treatment are determined according to Equation 2. As in Section 2.5, treatment effects can be dynamic and heterogeneous across treatment cohorts.

We can write the observed long-difference  $y_{i,t+h} - y_{i,t-1}$  as follows,

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \tau_{i,t+h} D_{i,t+h} - \tau_{i,t-1} D_{i,t-1} + e_{i,t}^h, \quad (\text{A.1})$$

where  $\delta_t^h = \delta_{t+h} - \delta_{t-1}$  and  $e_{i,t}^h = e_{i,t+h} - e_{i,t-1}$ .

#### LP-DiD specification

Consider the following LP-DiD specification with clean controls,

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP-DiD} \Delta D_{it} + \epsilon_{it}^h, \quad (\text{A.2})$$

restricting the sample to observations that are either

$$\begin{cases} \text{newly treated} & \Delta D_{it} = 1, \\ \text{or clean control} & D_{i,t+h} = 0. \end{cases} \quad (\text{A.3})$$

$\beta_h^{LP-DiD}$  is the LP-DiD estimate of the dynamic ATET,  $h$  periods after entering treatment.

#### Derivation of the weights

First, we need to define a clean control sample (CCS) for each treatment group. Consider a treatment group (or cohort)  $g > 0$ , as defined in Section 2.1. Define the clean control sample (CCS) for group  $g$  at time horizon  $h$  (denoted as  $CCS_{g,h}$ ) as the set of observations for time  $t = p_g$  that satisfy condition A.3. Therefore  $CCS_{g,h}$  includes the observations at time  $p_g$  for all units that either enter treatment at  $p_g$  or are still untreated at  $p_g + h$ . In other words,  $CCS_{g,h}$  includes observations at time  $t = p_g$  for group  $g$  and its *clean controls*.

By definition of groups and CCSs, each observation that satisfies condition A.3 enters into one and only one CCS. Therefore, the unbalanced panel dataset defined by the clean control condition in A.3 can always be reordered as a ‘stacked’ dataset, in which observations are grouped into consecutive and non-overlapping CCSs. The equivalence between the estimation sample defined

by the clean control condition of [Equation A.3](#) and the stacked dataset we just described implies that, in this baseline setting, LP-DiD is equivalent to the stacked approach of [Cengiz et al. \(2019\)](#).

Moreover, for any observation  $\{i, t\} \in CCS_{g,h}$ , we have  $\Delta D_{i,t} = \Delta D_{i,p_g} = D_{i,p_g}$ . This follows from the fact that for any  $\{i, t\} \in CCS_{g,h}$ , we have  $D_{i,t-1} = D_{i,p_{g-1}} = 0$  by virtue of the clean control condition.

Define event indicators as a set of  $G$  binary variables that identify the CCS that an observation belongs to. For each treatment group  $g > 0$ , the corresponding event indicator is equal to 1 if  $\{i, t\} \in CCS_{g,h}$  and 0 otherwise. By definition of treatment groups and CCCs, these event indicators are fully collinear with time indicators.

By the Frisch-Waugh-Lovell theorem,

$$E\left(\hat{\beta}_h^{LP-DiD}\right) = \frac{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E\left(y_{i,p_j+h} - y_{i,p_j-1}\right) \right]}{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \Delta \tilde{D}_{i,p_j}^2}, \quad (\text{A.4})$$

where  $\Delta \tilde{D}_{i,p_g}$  is the residual from a regression of  $\Delta D$  on time indicators in the sample defined by condition [A.3](#).

This residualized treatment dummy for unit  $i$  at time  $p_g$  is equal to

$$\Delta \tilde{D}_{i,p_g} = \Delta D_{i,p_g} - \frac{\sum_{i \in CCS_{g,h}} \Delta D_{i,p_g}}{N_{CCS_{g,h}}} = D_{i,p_g} - \frac{\sum_{i \in CCS_{g,h}} D_{i,p_g}}{N_{CCS_{g,h}}} = D_{i,p_g} - \frac{N_g}{N_{CCS_{g,h}}}, \quad (\text{A.5})$$

where  $N_{CCS_{g,h}}$  is the number of observations belonging to  $CCS_{g,h}$ , and  $N_g$  is the number of observations belonging to group  $g$ . For all observations belonging to the same group  $g > 0$ , we have  $\Delta \tilde{D}_{i,p_g} = \Delta \tilde{D}_{g,p_g} = 1 - \frac{N_g}{N_{CCS_{g,h}}}$ .

The first equality in [Equation A.5](#) follows from the full collinearity between time indicators and event indicators (defined as above); the second and third equalities follow from the definitions of groups and CCCs.

Given the parallel trends assumption ([Assumption 2](#)), we have

$$\begin{aligned} E\left(\hat{\beta}_h^{LP-DiD}\right) &= \frac{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E\left(y_{i,p_j+h} - y_{i,p_j-1}\right) \right]}{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \Delta \tilde{D}_{i,p_j}^2} \\ &= \frac{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E\left(\tau_{i,p_j+h} D_{i,p_j+h}\right) \right]}{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \Delta \tilde{D}_{i,p_j}^2} \\ &= \frac{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E\left(\tau_{i,p_j+h} D_{i,p_j}\right) \right]}{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \Delta \tilde{D}_{i,p_j}^2} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \frac{\Delta \tilde{D}_{i,p_j}}{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \Delta \tilde{D}_{i,p_j}^2} E \left( \tau_{i,p_j+h} D_{i,p_j} \right) \\
&= \sum_{j=1}^G \sum_{i \in j} \frac{\Delta \tilde{D}_{i,p_j}}{\sum_{j=1}^G \sum_{i \in j} \Delta \tilde{D}_{i,p_j}^2} \tau_{i,p_j+h} \\
&= \sum_{g \neq 0} \frac{N_g \Delta \tilde{D}_{g,p_g}}{\sum_{g \neq 0} N_g \Delta \tilde{D}_{g,p_g}^2} \tau_{g,p_g+h} \\
&= \sum_{g \neq 0} \omega_{g,h}^{LP-DiD} \tau_h^g,
\end{aligned}$$

where the weights are given by

$$\omega_{g,h}^{LP-DiD} = \frac{N_g \Delta \tilde{D}_{g,p_g}}{\sum_{g \neq 0} N_g \Delta \tilde{D}_{g,p_g}^2} = \frac{N_g \left( 1 - \frac{N_g}{N_{\text{CCS}_{g,h}}} \right)}{\sum_{g \neq 0} N_g \left( 1 - \frac{N_g}{N_{\text{CCS}_{g,h}}} \right)} = \frac{N_{\text{CCS}_{g,h}} [n_{gh}(n_{c,g,h})]}{\sum_{g \neq 0} N_{\text{CCS}_{g,h}} [n_{g,h}(n_{c,g,h})]}, \quad (\text{A.6})$$

where  $n_{g,h} = N_g/N_{\text{CCS}_{g,h}}$  is the share of treated units in the  $\text{CCS}_{g,h}$  subsample; and  $n_{c,g,h} = N_{c,g,h}/N_{\text{CCS}_{g,h}}$  is the share of control units in the  $\text{CCS}_{g,h}$  subsample. Recall that  $\tau_h^g$  was defined in the main text as the dynamic ATET for group  $g$  at time-horizon  $h$  (Equation 1).

## A.2 Weights with control variables

What are the weights of the LP-DiD estimator in a specification that includes control variables? If covariates have a linear and homogenous effect on the outcome, and parallel trends holds conditional on covariates, it is possible to show that the weights assigned to each group-specific effect by the LP-DiD estimator are unchanged by the inclusion of covariates. In more general settings, however, the weights are proportional to the residuals of a regression of the treatment indicator on time effects and the covariates, and it is not possible to ensure that they are always positive. (In these settings, using the regression-adjustment specification with covariates presented in the main text, which unbiasedly estimates an equally-weighted ATT, is preferable.)

To explore the role of covariates, we now assume that no anticipation and parallel trends hold after conditioning on a set of covariates (Assumptions 3 and 4 in the main text).

### A.2.1 Covariates with linear and homogeneous effects

**The DGP** Assume that covariates have a linear and homogeneous effect on the outcome. Specifically, assume the following DGP,

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \gamma_h x_i + \tau_{i,t+h} D_{i,t+h} - \tau_{i,t-1} D_{i,t-1} + \epsilon_{i,t}^h, \quad (\text{A.7})$$

**LP-DiD specification with covariates** The LP-DiD estimating equation with clean controls and control variables is

$$\begin{aligned}
y_{i,t+h} - y_{i,t-1} = & \beta_h^{LP-DiD} \Delta D_{it} && \text{treatment indicator} \\
& + \gamma_h \mathbf{x}_i && \text{covariates} \\
& + \delta_t^h && \text{time effects} \\
& + e_{it}^h; && \text{for } h = 0, \dots, H,
\end{aligned} \tag{A.8}$$

restricting the sample to observations that respect condition A.3.

**Weights derivation** All the definitions of clean control subsamples and indicators, and the results related to those, that have been described in Section A.1 above, still hold.

The LP-DiD specification of Equation A.8 can be rewritten as

$$y_{i,t+h} - y_{i,t-1} - \gamma_h \mathbf{x}_i = \beta_h^{LP-DiD} \Delta D_{it} + \delta_t^h + e_{it}^h.$$

Therefore, by the Frisch-Waugh-Lovell theorem, we have

$$E\left(\hat{\beta}_h^{LP-DiD}\right) = \frac{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E(y_{i,t+h} - y_{i,t-1} - \hat{\gamma}_h \mathbf{x}_i) \right]}{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \Delta \tilde{D}_{i,p_j}^2}, \tag{A.9}$$

where  $\Delta \tilde{D}_{i,p_g}$  is the residual from a regression of  $\Delta D$  on time indicators in the sample defined by condition A.3.

The equivalence of Equation A.5 above still holds; therefore, for all observations belonging to the same group  $g > 0$ , we have  $\Delta \tilde{D}_{i,p_g} = \Delta \tilde{D}_{g,p_g} = 1 - N_g / N_{\text{CCS}_{g,h}}$

Given the assumptions about the DGP, we have

$$\begin{aligned}
E\left(\hat{\beta}_h^{LP-DiD}\right) &= \frac{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E(y_{i,t+h} - y_{i,t-1} - \hat{\gamma}_h \mathbf{x}_i) \right]}{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \Delta \tilde{D}_{i,p_j}^2} \\
&= \frac{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \left[ \Delta \tilde{D}_{i,p_j} E(\tau_{i,p_j+h} D_{i,p_j+h}) \right]}{\sum_{j=1}^G \sum_{i \in \text{CCS}_{j,h}} \Delta \tilde{D}_{i,p_j}^2}.
\end{aligned}$$

This is the same expression as in the case of unconditional parallel trends and no covariates analyzed above, and it therefore leads to the same result,

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

where the weights are given by Equation A.6 above.

### A.2.2 More general setting

Now consider a more general setting, in which Assumptions 3 and 4 hold, but we do not restrict the effect of covariates to be linear or homogeneous. In this more general setting, the

Frisch-Waugh-Lovell theorem implies

$$E\left(\beta_h^{LP-DiD}\right) = \frac{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left[ \Delta \tilde{D}_{i,p_j}^c E\left(y_{i,p_j+h} - y_{i,p_j-1}\right) \right]}{\sum_{j=1}^G \sum_{i \in CCS_{j,h}} \left( \Delta \tilde{D}_{i,p_j}^c \right)^2}, \quad (\text{A.10})$$

where  $\Delta \tilde{D}_{i,p_g}^c = \Delta \tilde{D}_{g,p_g}^c$  is the residual from a regression of  $\Delta D$  on time indicators and the control variables  $\mathbf{x}_{it}$  in the sample defined by condition A.3.

The weights are thus given by

$$\omega_{g,h}^{c LP-DiD} = \frac{N_g \Delta \tilde{D}_{g,p_g}^c}{\sum_{g \neq 0} N_g \left( \Delta \tilde{D}_{g,p_g}^c \right)^2}. \quad (\text{A.11})$$

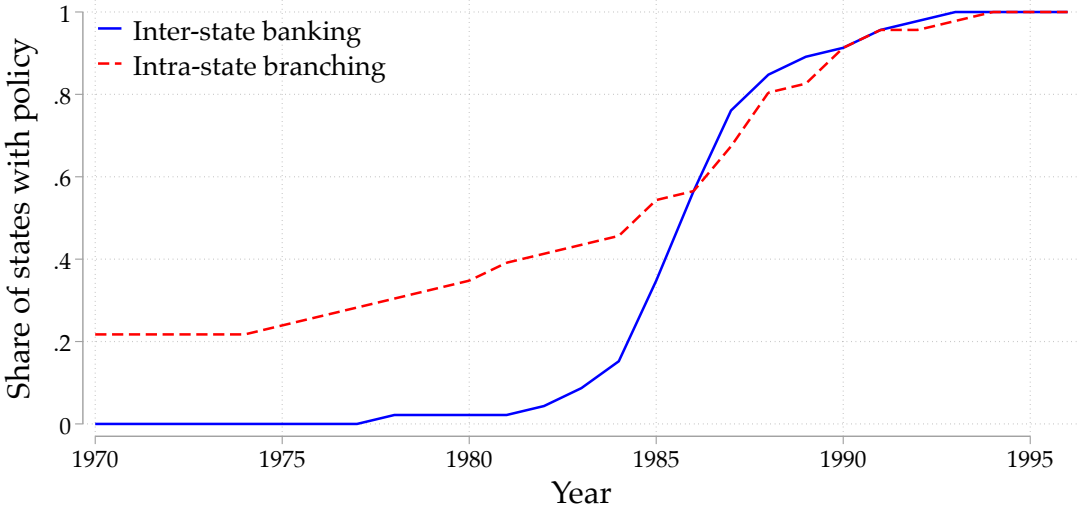
As noted in the main text (Section 4), it is always possible to preserve non-negative variance-weighted weights by employing semi-parametric propensity-score methods, or obtain an equally-weighted effect using regression adjustment.

# B Credit and the labor share: Additional statistics and results

This Appendix presents additional descriptive statistics and results concerning our first empirical application, which estimates the effect of banking deregulation on the labor share, replicating [Lelebicioğlu and Weinberger \(2020\)](#) (Section 6.1).

[Figure B.1](#) displays the share of US States which have adopted interstate banking and intrastate branching deregulation (similar to [Figure 1](#) in [Lelebicioğlu and Weinberger \(2020\)](#)).

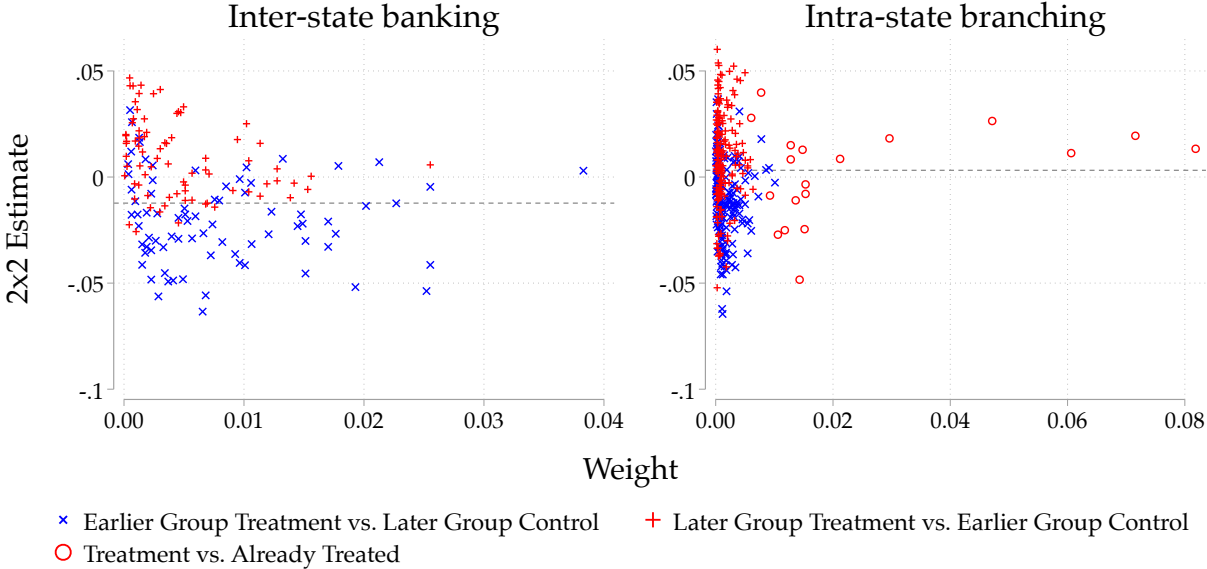
Figure B.1: Banking deregulation in US States



Notes: Data from [Lelebicioğlu and Weinberger \(2020\)](#).

[Figure B.2](#) displays the results of the [Goodman-Bacon \(2021\)](#) decomposition diagnostic. The figure plots each constituent 2x2 comparison that contributes to the static TWFE estimates, with its weight on the horizontal axis and its estimate on the vertical axis. The graph suggests that the estimates of the effects of branching deregulations are driven by a few ‘unclean’ comparisons – those involving states that deregulated before 1970 – that receive a very large weight. Notably, for both types of policies, clean comparisons produce overwhelmingly negative coefficients, while the unclean ones tend to bias the coefficients upwards.

Figure B.2: Goodman-Bacon (2021) decomposition diagnostic for the static TWFE specification of equation 18

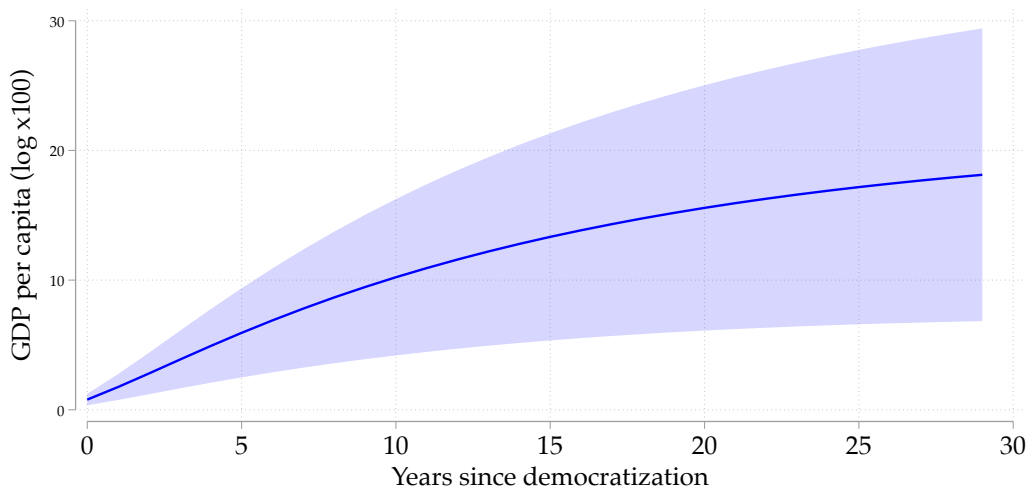


## C Democracy and growth: Additional statistics and results

This Appendix presents additional descriptive statistics and results concerning our second empirical application, which estimates the effect of democratization on economic growth, replicating [Acemoglu et al. \(2019\)](#) (Section 6.2).

[Figure C.3](#) displays the impulse-response function (IRF) from the estimation of the dynamic panel model of [Equation 22](#), reproducing the baseline results in [Acemoglu et al. \(2019\)](#).

Figure C.3: Effect of democracy on growth: dynamic panel estimates



Notes: Extrapolated impulse response function for the effect of democracy on GDP per capita, using the dataset of [Acemoglu et al. \(2019\)](#) and the dynamic fixed effects specification of [Equation 22](#). This graph replicates the baseline results from [Acemoglu et al. \(2019\)](#).