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

Arindrajit Dube† Daniele Girardi‡ Òscar Jordà§ Alan M. Taylor¶

May 2023

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, accommodates controls for pre-treatment values of the outcome and of other time-varying covariates, and is simple and fast to implement. Simulations and two empirical applications demonstrate that the LP-DiD estimator performs well in common applied settings.

Keywords: difference-in-differences, two-way fixed effects, event study, negative weights, local projections, clean controls

JEL codes: C01, C10, C21, C22, C23, C31, C33

*We thank 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, Asia School of Business (ASB), Sant'Anna School of Advanced Studies, UMass Amherst, and University of Pisa for helpful comments and exchanges. STATA example files illustrating the implementation of the LP-DiD estimators discussed in this paper can be found 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 or the Federal Reserve System.

†Department of Economics, University of Massachusetts, Amherst; NBER; and IZA (adube@umass.edu).

‡Department of Economics, University of Massachusetts, Amherst; and Department of Political Economy, King's College London (dgirardi@umass.edu).

§Federal Reserve Bank of San Francisco; Department of Economics, University of California, Davis; and CEPR (oscar.jorda@sf.frb.org; ojorda@ucdavis.edu).

¶Columbia University; NBER; and CEPR (amt2314@columbia.edu).

1 Introduction

Difference-in-differences (DiD) has become a widely used method 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 (e.g., no anticipation and parallel trends), the DiD estimator properly identifies 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 timing of treatment adoption can differ across groups (i.e., is *staggered*) and treatment effects can differ across groups (i.e., are *heterogeneous*) and occur gradually over time (i.e., are *dynamic*).[1] In such a 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. 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. What was once a seemingly simple tool of general application increasingly appears to need bespoke adjustments to suit each specific situation.

In this paper we propose a framework for DiD estimation that exploits an important link to a *local projection* (or LP), a statistical technique introduced in a time-series context in Jordà (2005). The local projection approach has been widely used to estimate dynamic impulse responses in time series or panel data. By its very design, and as used in applied macroeconomics, the LP approach was set up to estimate average treatment responses that are heterogeneous and dynamic. 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 criterion, such as past treatment history. By further developing these LP techniques from macroeconometrics in conjunction with the potential outcomes approach of microeconometrics to derive results for a wide range of DiD settings, we provide a more general toolkit for implementing the DiD method.

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

[1] See for example Callaway and Sant'Anna (2020); de Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2020); Goodman-Bacon (2021); Borusyak, Jaravel, and Spiess (2021); Gardner (2021); Wooldridge (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).

treatment events, or the choice of a base period for constructing the local projection. We also show that the LP approach can easily accommodate extensions to the basic case such as the inclusion of covariates or the generalization of the nature of the treatment.

Specifically, 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.[2] 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 clean control condition of our LP-DiD estimator 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, the LP-DiD estimator identifies a convex weighted average of potentially heterogeneous cohort-specific treatment effects. We characterize explicitly 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.

A skeptical reaction we can imagine hearing at this point is: why do we need yet another expanded DiD technique? Indeed, several alternative DiD estimators have recently been proposed to address the different settings that can arise in empirical applications without incurring in the 'negative weights' bias of two-way fixed-effects regression.[3]

However, rather like a Swiss Army knife, the LP-DiD approach offers a single, universal multi-tool with advantages over the array of methods for estimation, inference, and robustness currently employed in practice. The regression-based formulation of LP-DiD allows simple coefficient tests to estimate models for cases with heterogeneous or homogeneous effects, dynamic or non-dynamic responses, and staggered or non-staggered treatments, as well as to perform the standard validation check for pre-trends. Other advantages of LP-DiD are the simplicity of implementation, the ability to control for pre-treatment values of the outcome and other covariates, and the flexibility in how one

[2]The possible bias of two-ways fixed-effects regression when treatment is staggered is studied for example in Borusyak, Jaravel, and Spiess (2021); de Chaisemartin and D'Haultfœuille (2020); Goodman-Bacon (2021); Callaway and Sant'Anna (2020); Sun and Abraham (2020). See Section 2 for more detailed discussion.

[3]Alternative estimators 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).

defines the sets of treated and control units, in addition to the ability to handle arbitrary weighting schemes. Because the LP-DiD estimator is not specific to a particular setting, but can be applied in a variety of situations, it provides an encompassing framework, and 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.

As we noted above, LP-DiD also offers a unified approach that encompasses many of the recent estimators as specific sub-cases. In a baseline version without covariates and with binary absorbing treatment, and in which only not-yet-treated observations are used as controls, it will become clear that 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 two Monte Carlo simulations suggests that the LP-DiD estimator performs well in staggered difference-in-differences settings, also in comparison with other estimators that have recently been proposed. Our simulations consider a binary staggered treatment with dynamic and heterogeneous effects. In the first simulation treatment timing is exogenous. Under this scenario, 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. In our second simulation, the probability of entering treatment depends on lagged outcome dynamics. In this second scenario, the ability of LP-DiD to match on pre-treatment outcomes allows it to outperform other estimators. The purpose of these simulations is not mainly that of performing a horse race between LP-DiD and other estimators, but to show that LP-DiD performs well in plausible scenarios and that there is a class of settings—those in which matching on pre-treatment outcome dynamics or other pre-determined covariates is appropriate and important—in which LP-DiD could become the 'go-to' approach.

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 examples of important

empirical settings where conventional estimates are potentially subject to bias because of previously treated units being effectively used as controls, and where matching on pre-treatment outcomes and other covariates is likely to be important. These applications demonstrate the value of the LP-DiD approach in providing unbiased estimates both when the parallel trends assumption holds unconditionally, and when it holds only conditional on pre-treatment outcome dynamics.

The rest of this paper is organized as follows. In sections 2 and 3 we draw a connection between DiD methods and the LP estimator, and then present our proposed LP-DiD specification. In Section 4 we use simulations to assess the performance of our LP-DiD approach, with comparisons to other methods in the recent literature. In Section 5 we apply the LP-DiD estimator in two empirical applications. Section 6 concludes.

2 DiD local projection implementation

In this section we clarify the connection between the difference-in-differences (DiD) and local projection (LP) methods. The main aim is to show how LPs can be used to implement DiD in several different settings and with high flexibility. The presentation gradually increases the complexity of the scenarios introduced to more clearly establish the links between the issues raised in the literature and how LPs help sort them out.

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 the staggered setting, the conventional two-way fixed-effects (TWFE) implementation of DiD, both in the static and event-study version, can suffer from 'negative weights' bias as uncovered by recent important studies (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak, Jaravel, and Spiess, 2021; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020). We show that an LP approach can successfully address this problem.

The main result is that in this setting a LP regression with properly defined treated and control units following the 'clean control approach' of Cengiz et al. (2019) is able to identify a convex weighted average treatment effect without incurring in the negative weights problem. 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. We also discuss how a simple re-weighted LP regression can recover an equally-weighted average treatment effect on the treated.

2.1 General setup and notation

We consider the following general setup. An outcome y_{it} is observed for $i = 1, ..., N$ units over $t = 1, ..., T$ 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) $g \in \{0, 1, \ldots, G\}$ 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 $y_{it} = y_{it}(0) + \sum_{p=1}^{T} \left(y_{it}(p) - y_{it}(0)\right) \times \mathbf{1}\{p_i = p\}$.[4]

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 then define the (group-specific and dynamic) average treatment effect on the treated (ATT) at time horizon h for group g which enters treatment at time p_g as

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

In other words, τ_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.[5]

Throughout the discussion which follows, we will make use of the assumptions of parallel trends and no anticipation, the two essential assumptions that underpin the DiD approach. We state these assumptions unconditionally though later we will condition on covariates.

[4] Similar notation is used, for example, in Callaway and Sant'Anna (2020) and Sun and Abraham (2020).

[5] This object is analogous to the cohort-specific treatment effect on the treated (CATT) defined in Sun and Abraham (2020).

Assumption 1. No anticipation

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

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

Assumption 2. Parallel trends

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

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. This assumption is made to ensure that the measured treatment effect cannot be explained by differences in time trends between treated and control units.

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

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

where α_i is a unit-specific fixed effect, and δ_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). We will discuss, compare and evaluate these specifications throughout our discussion.

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

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

where the α are unit-specific intercepts and the δ are common time-specific fixed effects, and we denote with e the error term.

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

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

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 ($0 \leq h \leq H$). Fixed effects α and δ are again included, and an error term e. Also note that $\beta_{-h}^{ETWFE} = -\sum_{j=-h}^{-1} \gamma_j^{ETWFE}$ 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 ΔD instead of its level D, except for the H–th lag, which is taken in level.[6]

Specification 3. Local Projections regression (LP)

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

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, in Equation 5, a different regression is needed for each time horizon h, in contrast to the previous two specifications. Given our assumptions on a staggered treatment that is absorbing, 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 β 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, two time periods, and a binary treatment. 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 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.

[6] Yet another specification of event study TWFE, often used in applications, uses the first difference of the treatment indicator ΔD and normalizes estimates by subtracting the coefficient on the first lag.

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,

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

In the second line we simply add and subtract $y_{i1}(0)$; in the third line we use Assumption 2 (parallel trends); and in the last line we make use of Assumption 1 (no anticipation). β^{2x2} is thus 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 β^{2x2}, which (given no-anticipation and parallel trends) equals the ATT. As is well known, in this setting also the estimand β^{TWFE} from the static TWFE regression of Equation 3 is equivalent to the coefficient from a first-difference regression and corresponds to β^{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. Therefore, in pre-treatment periods $t < s$ no unit is treated. In

post-treatment periods $t \geq s$, units in the treatment group are treated, while units in the control group are not. 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 $\tau_h = E[y_{i,s+h}(s) - y_{i,s+h}(0)|p_i = s]$.

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

$$\begin{aligned}
\tau_h &\equiv E[y_{i,s+h}(s) - y_{i,s+h}(0)|p_i = s] \\
&= E[(y_{i,s+h}(s) - y_{i,s-1}(0)) - (y_{i,s+h}(0) - y_{i,s-1}(0))|p_i = s] \\
&= E[y_{i,s+h}(s) - y_{i,s-1}(0)|p_i = s] - E[y_{i,s+h}(0) - y_{i,s-1}(0)|p_i = \infty] \\
&= E[y_{i,s+h} - y_{i,s-1}|p_i = s] - E[y_{i,s+h} - y_{i,s-1}|p_i = \infty] \qquad \equiv \beta_h^{DiD},
\end{aligned}$$

where β_h^{DiD} is the DiD estimand for the dynamic ATT h periods after treatment and we have applied Assumptions 1 and 2 in lines 3 and 4 as we did previously.

The population coefficient β_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 ΔD_{it}. Hence, observations with $t \neq s$ do not contribute to the estimated coefficient β_h^{LP}, and the coefficient β_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 τ_h estimands.[7] Moreover, the β^{STWFE} estimand from the static TWFE

[7]This can be seen using the decomposition of the event-study TWFE coefficients (β_h^{ETWFE} in our notation) provided by Sun and Abraham (2020). This shows that β_h^{ETWFE} is equal to τ_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 $E[\tau_{it}|D_{it} = 1]$.[8]

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

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

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,[9]

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

Therefore the dynamic ATT τ_h corresponds to the β_h^{LP} population coefficient in the

[8] 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, β^{STWFE} corresponds to an equally-weighted average of all the cell-specific ATTs.

[9] 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,

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

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.[10]

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, is able to recover 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 Sun and Abraham, 2020, in particular Proposition 4 and Equation 19). Intuitively, the lagged treatment indicators control for the lagged dynamic effects of previous treatments, which in this setting are the same (in expectation) for all units.

2.5 Staggered treatment adoption with dynamic and heterogeneous treatment effects

For greater generality, 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 $g' \neq g$.

[10] Heterogeneous effects, which we consider below in Section 2.5, would make this problem worse.

This more general case, because of the problems it can cause, has been the main focus of a growing recent literature (e.g., de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2020; Callaway and Sant'Anna, 2020; Goodman-Bacon, 2021; Borusyak, Jaravel, and Spiess, 2021; Wooldridge, 2021; Gardner, 2021).

With heterogeneous treatment effects of this form, it is now well understood that the static TWFE estimator (Equation 3) is biased both because of dynamic lagged effects and because of heterogeneity. Let us define a cell as a given treatment group g in a given period t. de Chaisemartin and D'Haultfœuille (2020) show that under no-anticipation and parallel trends β^{STWFE} in Equation 3 equals in expectation a weighted average of all cell-specific ATTs, but with weights that can be negative. Negative weights introduce bias: for example, positive cell-specific effects can enter the formula for the TWFE cofficient with a negative sign.

Another way to see this problem is through the lens of the Goodman-Bacon (2021) decomposition theorem, which shows that the static TWFE estimator in Equation 3 is an average of all potential 2x2 comparisons in the data, with weights based on subsample shares and treatment variances.

The problem is that some of these 2x2 comparisons are 'unclean' comparisons in which previously treated units are used as controls for newly-treated units. These 'unclean comparisons' are the source of the 'negative weights' bias of static TWFE.[11]

Furthermore, in the case of heterogeneous treatment effects across cohorts, it is also known that the event-study TWFE specification (Equation 4) is generally biased (Sun and Abraham, 2020). 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.

In short, the two most common DiD methods fail to work in this general type of setting, which has led to the search for, and proliferation of, so many alternative estimators.

To understand the relation between LP and DiD in this setting, we can start by noting

[11] Goodman-Bacon (2021) also shows that under parallel trends and no anticipation (assumptions 1 and 2), $\text{plim}_{N\to\infty}\hat{\beta}^{STWFE} = VWATT - \Delta ATT$, where $VWATT$ is a convex variance-weighted average of ATTs from all possible 2x2 comparisons in the data, and $-\Delta ATT$ is bias coming from dynamic and heterogeneous effects. Seen in this way, the bias term is equal to a weighted sum of changes in treatment effects within each group.

that here $E[y_{i,t+h}]$ is determined as follows,

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

Subtracting $E[y_{i,t-1}]$ from both sides, we obtain,[12]

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

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.[13] 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.

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-

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

[13] 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 2.6).

horizon $h+j$, we have $\tau^g_{h+j} \neq \tau^g_{j-1}$, 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 above and noted in the recent literature on DiD (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020; Borusyak, Jaravel, and Spiess, 2021).

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 $\Delta D_{it} = 0$ but $\Delta D_{i,t+j} \neq 0$ for some j in $1 \leq j \leq h$.[14]

Contribution A properly specified LP regression which we call *LP-DiD* solves all of these problems and identifies a convex combination of cohort-specific effects.

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

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

[14] As Equation 8 shows, one solution would be a LP regression that identifies 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.

by restricting the estimation sample to observations that are either

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

By removing previously treated observations and observations treated between $t+1$ and $t+h$ from the control group, β_h^{LP-DiD} from Equation 10 provides a convex combination of all group-specific effects τ_h^g. Before analysing the general case (in Section 2.6 below), it is instructive to consider again the special case with $\delta_t^h = \delta^h$. In this special case, we have

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

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 3 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 LP-DiD approach 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.

2.6 Weights of the LP-DiD estimator

We can explicitly characterize the weights assigned to each cohort-specific effect τ_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 estimator without covariates identifies 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 and an extension to the case with covariates are 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 β_h^{LP-DiD} identifies the following weighted average effect,

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

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

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

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.[15]

Thus, in short, the LP-DiD estimator β_h^{LP-DiD} identifies a variance-weighted ATT (*VWATT* in the terminology of Goodman-Bacon, 2021).

2.7 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 $1/(\omega_{g,h}^{LP-DiD}/N_g)$, identifies the equally-weighted ATT.

In practical applications, the weight $(\omega_{g,h}^{LP-DiD}/N_g)$ can be obtained by computing sub-samples sizes and shares of treated and control units in the sample and using Equation 12,

[15] The derivation of these weights is in Appendix A.

or through an auxiliary regression. Specifically, consider an auxiliary regression of ΔD on time indicators in the sample defined by Equation 10. Define $\Delta\widetilde{D}_{g,p_g}$ as the residual at time p_g for a unit belonging to group g.[16] The Frisch-Waugh-Lovell theorem implies that

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

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 $y_{i,t+h} - y_{i,t-1}$ on time effects 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 $\widehat{y_{i,t+h} - y_{i,t-1}}$ for each treated unit. The ATT is then estimated as

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

where TR is the set of treated observations (i.e., the set of observations with $\Delta D_{i,t} = 1$).[17] This regression-adjustment implementation of LP-DiD constitutes an imputation estimator, in the same sense as in Borusyak, Jaravel, and Spiess (2021).

2.8 Alternative pre-treatment base periods

We now discuss how, 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 $y_{i,t+h} - y_{i,t-1}$

[16] Note that $\Delta\widetilde{D}_{g,p_g}$ will be identical for all units belonging to the same group.

[17] 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.

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, and produce more noisy estimates than necessary.

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,

PMD LP-DiD regression

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

restricting the estimation sample to observations that are either:

$$\begin{cases} \text{newly treated:} & \Delta D_{it} = 1\,, \\ \text{or clean control:} & D_{i,t+h} = 0\,. \end{cases} \tag{13}$$

(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 $\beta_h^{PMD\ LP\text{–}DiD}$ identifies a convex weighted average of cohort-specific effects, with the same weights $\omega^{LP\text{–}DiD}$ discussed in Section 2.6. Also in this case, weighted regression or regression adjustment can be employed to obtain an equally-weighted ATE (Section 2.7).

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. 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.

2.9 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 β_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, $h \in 0, ..., H$. 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 β_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 β_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).

2.10 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 $p_g > T - H$).

2.11 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 β_h^{LP-DiD} is numerically equivalent to the estimate from a stacked regression approach as implemented in Cengiz et al. (2019).[18] Moreover, the re-weighted version of LP-DiD (discussed in Section 2.7) is numerically equivalent to the estimator proposed by Callaway and Sant'Anna (2020).[19]

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.[20] 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

[18] See Appendix A for more details about this equivalence.

[19] In terms of our notation, the Callaway and Sant'Anna (2020) estimator of τ_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 a 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.

[20] With one single treated group g that enters treatment in period p_g, the BJS estimator has a closed form

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 4 below) show that with more than one treated group, when using 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.[21] Movever, unlike the BJS estimator, PMD LP-DiD can be implemented using simple OLS (or weighted least squares) regression, using commonly used and well understood methods for statistical inference. Third, as we will discuss in Section 3.1 and as our simulations and empirical application will illustrate, the structure of the LP-DiD specification allows one to control for *pre-treatment* lags of the outcome and of other 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, our first simulation (Section 4 below), in which both estimators are unbiased, shows that 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, which might be seen as implausible in most panel data applications (de Chaisemartin and D'Haultfœuille, 2022, p. 18). Moreover, if there is little heterogeneity in treatment effects between treatment cohorts or a variance-weighted effect is as satisfactory as an equally-weighted effect, variance-weighting (as done by LP-DiD or

(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.

[21] With one single treated group, the PMD LP-DiD estimator (like the BJS estimator) is equal to the expression in foonote 20. With more than one treated group, the PMD LP-DiD estimator is equal to a weighted average of the expression in footnote 20 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 20 across all treated units.

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 easier to generalize, for example by conditioning on pre-treatment values of the outcome or using other covariates.

3 Extensions

In this section we extend the LP-DiD approach introduced in the previous section to a variety of settings encountered in empirical research. We begin by discussing the inclusion of covariates. Covariates can help in two ways: (1) as a control for variation in treatment assignment; and (2) as a way to improve the efficiency of the estimated ATT. Next we consider settings in which treatment is not absorbing, that is, units can enter and exit treatment multiple times. In addition, we briefly comment on non-binary treatments. Because the LP-DiD approach is based on simple regression, it can naturally normalize the treatment effect depending on treatment intensity (or *dose*), though we leave a more thorough development of this topic for another paper. However, 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.

3.1 Inclusion of covariates

We will now go on to show how the LP-DiD specification of Equation 10 can easily be augmented to include both time-invariant and time-varying covariates. Including covariates might be necessary for identification, if the parallel trends assumption only holds conditional on (constant or time-varying) variables that determine selection, or to increase precision.

A distinctive feature of the LP-DiD approach is that it allows to control for *pre-treatment* values of time-varying covariates, including lagged outcome dynamics. This is made possible by the structure of the LP specification. Unlike a standard event-study TWFE specification or most alternative estimators proposed in the recent literature, the LP specification implies that any lagged variable included in the estimating equation is measured before treatment.[22]

[22]Of course, controlling for pre-treatment outcome dynamics (as well as any other exogenous or pre-

Formally, in the special case with binary absorbing treatment, an LP-DiD specification which controls for K lags of the outcome dynamics and also controls for contemporaneous and lagged values of a vector of exogenous or pre-determined covariates, $\boldsymbol{x}_{it}$ can be written as follows.[23]

LP-DiD regression with exogenous covariates and lagged outcome dynamics

$$\begin{aligned} y_{i,t+h} - y_{i,t-1} = \; & \beta_h^{LP-DiD} \Delta D_{it} && \text{treatment indicator} \\ & + \textstyle\sum_{k=1}^{K} \gamma_k^h \Delta y_{i,t-k} && \text{outcome lags} \\ & + \boldsymbol{\gamma}^h \boldsymbol{x}_{it} && \text{covariates} \\ & + \delta_t^h && \text{time effects} \\ & + e_{it}^h \; ; && \text{for } h = 0, \ldots, H \,, \end{aligned} \tag{14}$$

restricting the sample to observations that are either

$$\begin{cases} \text{newly treated} & \Delta D_{it} = 1 \,, \\ \text{or clean control} & D_{i,t+h} = 0 \,. \end{cases} \tag{15}$$

This setup requires variants of Assumptions 1 and 2, labeled 3 and 4, where the statement of the assumptions is made conditional on outcome lags and other covariates $\boldsymbol{x}_{it}$.[24]

Appendix A.2 discusses the weights assigned to each group-specific effect in the specification that includes control variables. The main result is that the weights are guaranteed to remain the same as in Equation 12 if covariates have linear and homogeneous effects. In more general settings, the presence of covariates will alter the weighting scheme in

determined covariate) will be appropriate in some applications but not in others. A discussion of the conditions under which it is appropriate or necessary to control for lagged outcome dynamics and other covariates in the DiD setting is outside the scope of this paper (see for example Chabé-Ferret 2015; Caetano, Callaway, Payne, and Rodrigues 2022). What matters here is that the LP-DiD estimator offers flexibility in this respect: the researcher can decide whether to control for lagged outcomes and other covariates based on the application.

[23] Of course, the vector of covariates $\boldsymbol{x}_{it}$ can include not only variables taken in levels and measured at time t, but also differenced or lagged variables.

[24] Formally, conditional versions of Assumptions 1 and 2 can be written as follows:

Assumption 3. Conditional no anticipation

$$E\left[y_{it}(p) - y_{it}(0) | \Delta y_{i,t-1}, ..., \Delta y_{i,t-K}; \mathbf{x}_{it}\right] = 0, \text{ for all } p \text{ and } t \text{ such that } t < p.$$

Assumption 4. Conditional parallel trends

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

ways that are difficult to characterize analytically. In applications, the weights can still be computed empirically through the auxiliary regression described in Section 2.6.

If one wants to preserve the variance-weighting scheme of the baseline specification without covariates, or to avoid other possible drawbacks from the inclusion of covariates in linear regression in the DiD setting (discussed by Sant'Anna and Zhao, 2020), it is easy to control for covariates semi-parametrically using propensity-score based methods in the spirit of Sant'Anna and Zhao (2020). Moreover, an equally-weighted average effect can still be estimated by using regression adjustment, as described in Section 2.6. Jordà and Taylor (2016) discuss the implementation of propensity-score based and doubly-robust methods in the LP setting and apply them to estimate the effects of fiscal consolidation.

When the specification includes lags of the outcome, Nickell (1981) bias can arise from the presence of $y_{i,t-1}$ both as a regressor and in the error term, which is equal to $e_{i,t+k} - e_{i,t-1}$. However, two conditions must both be met for this bias to be problematic. First, the autoregressive coefficient on the lagged outcome variable must be high. Second, the time dimension of the dataset must be relatively small. If either of these two conditions fails, the bias is negligible as Álvarez and Arellano (2003) show. In applications in which 'Nickell bias' is a concern, the researcher can nevertheless correct for it by using a simple split-sample correction, following Chen, Chernozhukov, and Fernández-Val (2019).

3.2 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 definitions of the causal effect of interest and the different identification assumptions that might be appropriate under non-absorbing treatment.

An appropriate modification of the 'clean control' approach of Equation 10 will generally be necessary to implement LP-DiD in this setting. One can recover, for example, the effect of entering treatment for the first time and staying treated, relative to a counterfactual of remaining untreated, by using the LP-DiD specification of Equation 10 but then modifying the 'clean control' sample restriction as follows,

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } j \geq 1)\,, \\ \text{or clean control} & D_{i,t-j} = 0 \text{ for } j \geq -h\,. \end{cases} \tag{16}$$

However, in numerous settings of practical importance the 'clean control' condition in Equation 16 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 $t+h$. This case can be dealt with in a simple way in the LP-DiD framework, under the additional assumption that dynamic treatment effects stabilize after a finite number of periods.

Formally, we introduce the following assumption.

Assumption 5. Dynamic effects stabilize after L periods:

$$\tau_L^g = \tau_{L+l}^g \quad \text{for } l \geq 0\,, \text{ and all groups } g = 1, ..., G\,.$$

Assumption 5 implies that, for any time horizon h and for any $j \geq L+1$, we have $\tau_{h+j}^g = \tau_{j-1}^g$. Therefore, under Assumptions 1, 2 and 5, Equation 9 becomes[25]

$$\begin{aligned} E[\hat{\beta}_h^{LP}] \;=\; & E\left(\sum_g^G \left[\tau_h^g \times \mathbf{1}\{t = p_g\}\right] |\Delta D_{it} = 1\right) \\ & - E\left(\sum_{g=1}^G \left[\sum_{j=1}^L \left((\tau_{h+j}^g - \tau_{j-1}^g) \times \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\}\right)\right] |\Delta D_{it} = 0\right) \\ & - E\left(\sum_{g=1}^G \left[\sum_{j=1}^h \left(\tau_{h-j}^g \times \Delta D_{i,t+j} \times \mathbf{1}\{t = p_g - j\}\right)\right] |\Delta D_{it} = 0\right) \end{aligned} \tag{17}$$

Equation 17 implies that bias only comes from observations that experience a change in treatment status between time $t-L$ and $t-1$ or between $t+1$ and $t+h$.

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

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L)\,, \\ \text{or clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L\,. \end{cases} \tag{18}$$

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 $t+h$.[26]

Later, in Section 5.2, we illustrate the use of LP-DiD under non-absorbing treatment in an empirical application which estimates the effect of democracy on economic growth.

[25] Recall that Equation 9 isolates the role of 'unclean controls' in creating bias by focusing on the special case with $\delta_t^h = \delta^h$. Equation 17 does the same but adds Assumption 5.

[26] 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.

3.3 Continuous treatment

A detailed formal discussion of the issues that can arise under continuous treatment is outside the scope of this paper (see for example de Chaisemartin et al. 2022). However, it is clear that the LP-DiD framework also offers flexibility to accommodate the different definitions of the causal effect of interest and the different identification assumptions that might be appropriate with continuous treatment. For example, the clean control condition can be adapted to define clean controls as 'stayers' (or alternatively 'quasi-stayers'), in the terminology of de Chaisemartin et al. (2022).

3.4 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 framework for applied microeconomic research and the ways LPs are used in macroeconomics.

First, we reiterate the identification assumptions behind LP-DiD (and the DiD literature in general). As Ghanem, Sant'Anna, and Wüthrich (2022) show, as long as selection into treatment is driven by static covariates or covariates that change over time randomly (so that they are unpredictable), then we can rely on the no-anticipation and parallel trends assumptions to eliminate selection bias. When identification further requires conditioning on covariates (especially in dynamic settings where the parallel trends assumption may only hold conditionally), these assumptions can be made conditionally, as the literature has shown. In practice, we expect that these should be the default assumptions.

We think it is helpful to compare 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, in which case, an instrumental variable approach would be advisable. 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.

4 Simulations

In this section, we conduct two Monte Carlo simulations to evaluate the performance of the LP-DiD estimator. We consider a binary staggered treatment, with dynamic and heterogeneous treatment effects. In the first simulation, treatment is exogenous; the parallel trends assumption holds and the conventional TWFE model only fails because of heterogeneous dynamic effects, which lead to negative weighting (as discussed in Section 2.5). In the second simulation, treatment is endogenous; specifically, the probability of receiving treatment depends on previous outcome dynamics. We compare the performance of LP-DiD with a conventional event-study TWFE specification and other recently proposed estimators.

Results suggest that, unlike the conventional TWFE specification, LP-DiD tracks well the true effect path even in the presence of heterogeneity. With exogenous treatment, 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. When the probability of

treatment depends on lagged outcome dynamics, the ability of LP-DiD to match on pre-treatment outcomes makes it outperform other estimators.

4.1 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 and staggered and treatment is an absorbing state. 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

$$y_{it} = y_{it}(0) + \beta_{it}\,.$$

Simulation 1: Exogenous treatment timing In simulation 1, we assume that treatment is exogenous. Specifically, 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 $p = 11, 13, 15 \ldots, 27$.

Simulation 2: Endogenous treatment timing In simulation 2, treatment timing is endogenous: the probability of receiving treatment depends on past outcome dynamics. Specifically, unit i enters treatment in the first period that satisfies that following condition:

$$\psi \Delta y_{i,t-1} + (1-\psi)u_i \leq \theta \quad \text{and} \quad 11 \leq t \leq 30\,,$$

with $\psi = 0.6$, $u_i \sim N(0, 25)$ and $\theta = -\sigma_{\Delta y_{it}(0)}$. The probability of entering treatment is therefore higher for untreated units that experience a large negative change in the outcome variable.

4.2 Results

We perform 200 replications of each of the two simulations and evaluate five estimators:

- A conventional event-study TWFE specification with leads and lags of the treatment indicator (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.[27]
- 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.

4.2.1 Results of Simulation 1: Exogenous treatment scenario

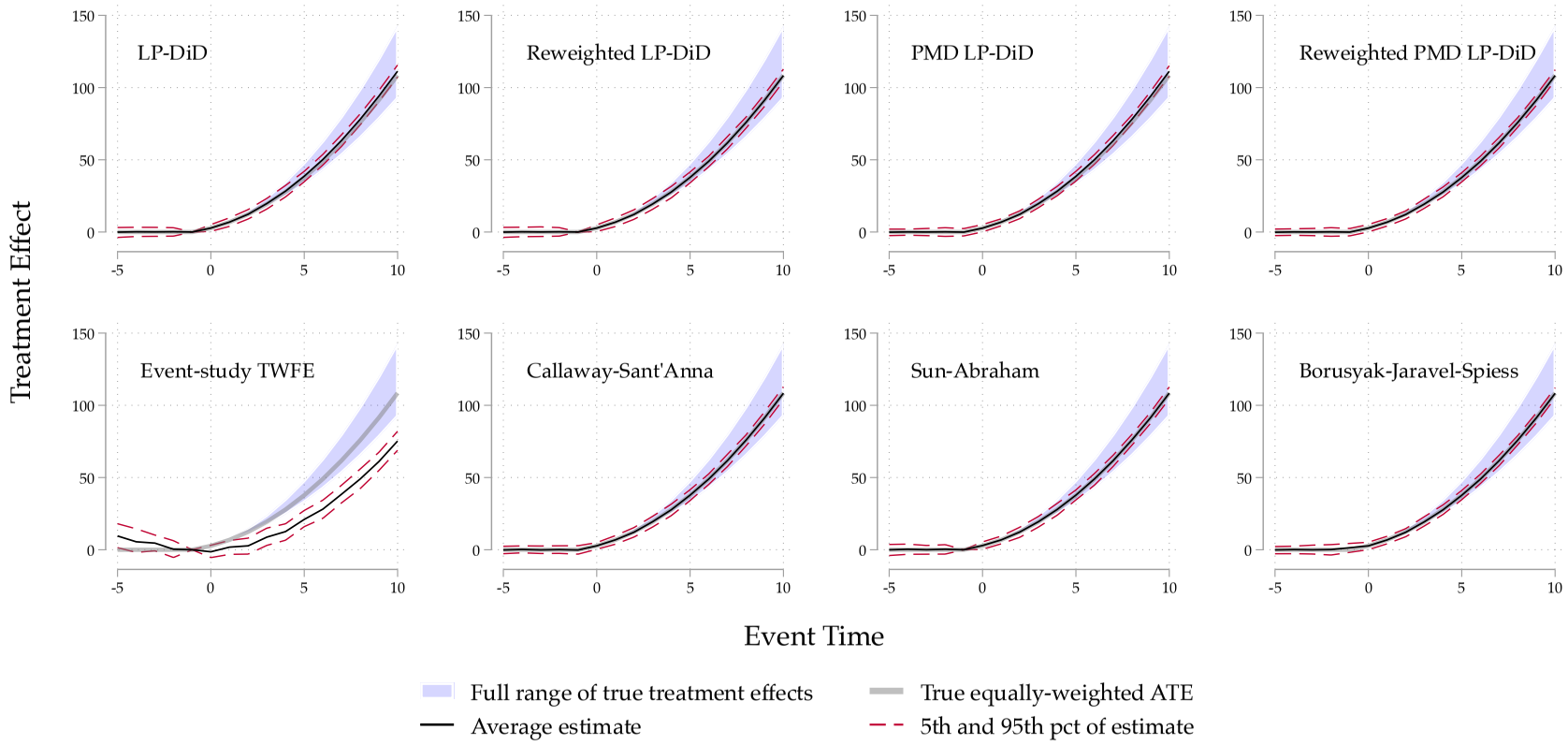
Results from Simulation 1, with exogenous treatment timing, are presented in Figures 1 and 2 and Tables 1 and 2.

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. Tables 1 and 2 report the root mean squared error (RMSE) and the empirical standard error of each estimator at time horizons between –5 and +10.

The conventional event-study TWFE specification does a poor job in our setting, due to the heterogeneity of treatment effects. It finds a non-existent decreasing pre-trend, and

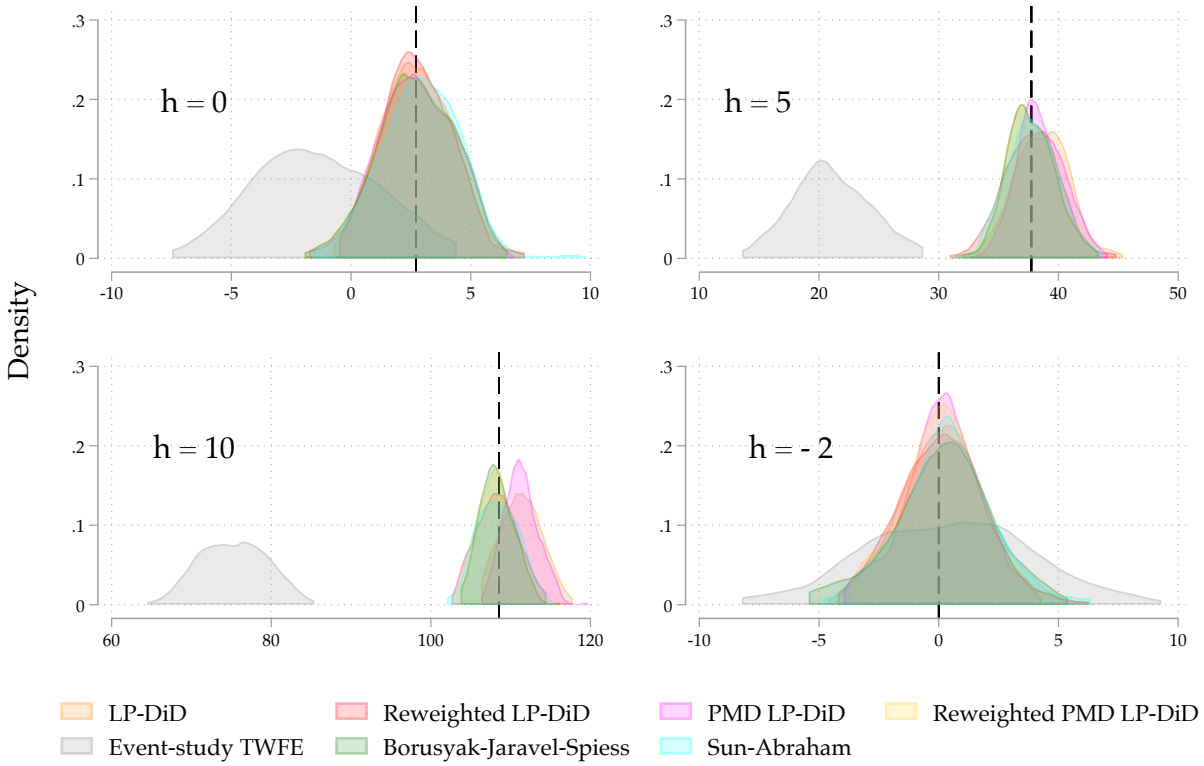
[27] As discussed in Section 2.6, the variance-weighted version identifies a convex combination of all group specific effects, with weights given by Equation 11 and Equation 12. The re-weighted version identifies an equally-weighted ATT; it can be implemented through a weighted least squares regression or using regression adjustment. See Section 2.9 for the discussion of first-lag differencing versus pre-mean differencing.

Figure 1: Simulation 1: Exogenous treatment scenario: True effect path and estimates



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

Figure 2: Simulation 1: Exogenous treatment scenario: 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.

it grossly underestimates the dynamic treatment effect. Due to negative weighting, point estimates lie outside the full range of true group-specific effects.[28]

Our LP-DiD estimator, instead, tracks well the average true effect. This is true of both simple (variance-weighted) LP-DiD and re-weighted LP-DiD. As expected, there is a bias-variance trade-off between the two. The variance-weighted version has a slightly smaller empirical standard error (Table 2), but also a small bias due to differences between the variance-weighted and equally-weighted ATTs (Figure 1 and Figure 2). This bias tends to become more relevant at longer time-horizons, because in our simulation the true treatment effect variance increases in time after treatment.

Overall, variance-weighted LP-DiD has smaller RMSE than re-weighted LP-DiD at short time-horizons, but the opposite is true at longer time horizons (Table 1). Pre-mean-

[28]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.

Table 1: Simulation 1: Exogenous treatment scenario: 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).

differencing (PMD) has a lower RMSE than first-lag differencing, as expected in a setting in which parallel trends holds fully and at all time periods (see the discussion in Section 2.9).

In this exogenous treatment setting, also the SA, CS and BJS estimators perform very well. Overall, LP-DiD performs similarly to these (computationally more demanding) estimators. 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 to testing for pre-trends (i.e., pre-treatment time horizons). [29]

[29] 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 2.9, 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 demostration

Table 2: Simulation 1: Exogenous treatment scenario: Empirical standard errors

Event time	ES TWFE	LP-DiD	Rw LP-DiD	PMD LP-DiD	Rw PMD LP-DiD	SA	CS	BJS
-5	5.07	2.06	2.11	1.31	1.34	2.18	1.61	1.54
-4	5.12	1.89	1.97	1.33	1.37	2.17	1.50	1.63
-3	3.35	1.93	1.96	1.48	1.52	1.83	1.54	1.93
-2	3.49	1.79	1.80	1.60	1.62	1.94	1.60	2.08
0	2.61	1.48	1.46	1.58	1.59	1.68	1.46	1.59
1	2.89	1.83	1.85	1.55	1.58	1.68	1.85	1.58
2	3.39	2.09	2.13	1.63	1.65	2.18	2.13	1.64
3	3.43	2.39	2.41	1.88	1.90	2.25	2.41	1.89
4	3.53	2.35	2.36	2.01	1.99	2.42	2.36	1.98
5	3.25	2.32	2.33	2.00	2.01	2.14	2.33	1.99
6	3.76	2.40	2.38	2.07	2.07	2.39	2.38	2.04
7	3.76	2.39	2.45	2.06	2.13	2.38	2.45	2.09
8	4.14	2.25	2.32	1.94	1.98	2.37	2.32	1.93
9	3.94	2.42	2.52	2.07	2.15	2.41	2.52	2.10
10	4.22	2.53	2.57	2.18	2.25	2.63	2.57	2.18

Notes: Empirical standard errors 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).

4.2.2 Results of Simulation 2: Endogenous treatment scenario

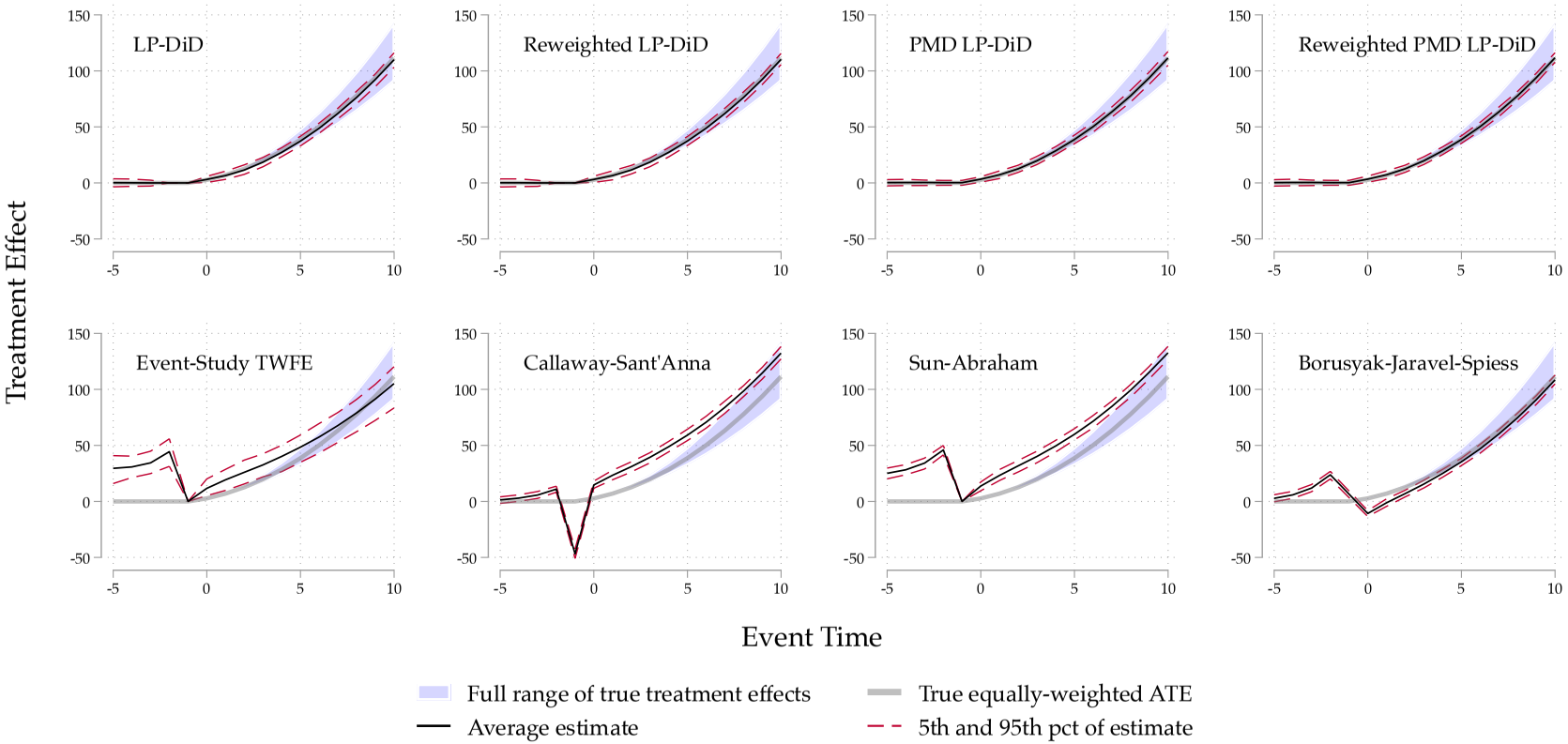
Results from the simulation with endogenous treatment timing (Simulation 2) are reported in Figures 3 and 4 and Tables 3 and 4.

In applying our LP-DiD estimator in this setting, we include one lag of the change in the outcome variable as a control. This warrants some discussion in comparison to other estimators. While the SA, CS and BJS estimators do allow the inclusion of time-invariant, and in some cases time-varying, control variables, there is no straightforward way to control for *pre-treatment lags* of the outcome (or of other time-varying covariates) in their specifications.

LP-DiD outperforms other estimators in the presence of this particular failure of the parallel-trends assumption, due to its ability to match on pre-treatment outcome dynamics. Indeed, the LP-DiD estimator tracks quite well the true dynamic effect also in this second simulation (Figure 3) and it has the lowest RMSE (Table 3). The SA, CS

of efficiency, because that demostration assumes no autocorrelation, while in this simulation there is some autocorrelation.

Figure 3: Simulation 2: Endogenous treatment scenario: True effect path and estimates



Notes: Average estimates and 95% and 5% percentiles from 200 replications. To filter out variation in estimates due to variation in the true treatment effect across replications, we subtract from each estimate the true effect, and then add back the average true effect across all replications. This adjustment is in order because in the 'endogenous treatment' setting, the average treatment effect is not deterministic.

Table 3: Simulation 2: Endogenous treatment scenario: 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	30.66	2.12	2.18	1.70	1.73	25.29	2.29	3.38
-4	31.43	2.05	2.09	1.69	1.78	28.41	3.46	6.20
-3	35.00	1.57	1.60	1.52	1.63	34.36	5.99	12.09
0	10.13	1.71	1.72	1.73	1.78	11.28	12.11	13.63
1	13.77	2.24	2.35	2.13	2.19	16.52	16.59	8.44
2	14.78	2.53	2.61	2.04	2.07	19.40	18.61	6.00
3	14.19	2.76	2.75	2.18	2.17	20.60	19.77	4.76
4	13.47	2.63	2.58	2.19	2.16	21.27	20.54	3.99
5	12.15	2.94	2.76	2.33	2.12	21.58	20.74	3.82
6	10.48	3.27	3.06	2.68	2.44	21.64	20.75	4.09
7	9.14	3.36	3.04	2.89	2.47	21.73	20.93	3.94
8	8.58	3.73	3.12	3.24	2.43	21.67	20.9	3.93
9	10.03	3.95	3.00	3.60	2.50	21.67	20.98	3.87
10	13.18	4.24	3.09	3.92	2.61	21.57	21.18	3.85

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).

and BJS estimators are biased in this setting in which selection into treatment is based on lagged outcome dynamics.

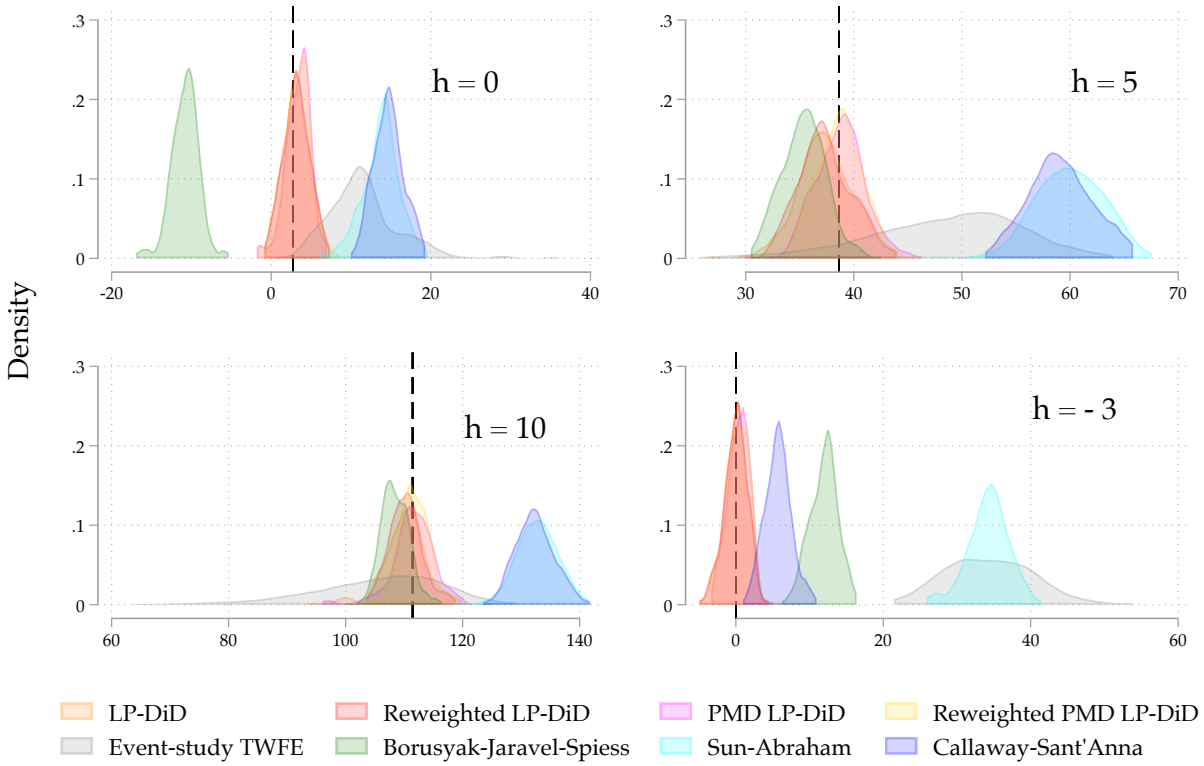
Also in this endogenous treatment setting, there is a bias-variance tradeoff between simple (variance-weighted) LP-DiD and reweighted DiD. Reweighted LP-DiD is unbiased, while variance-weighted LP-DiD has a small bias but also smaller variance. Overall, the two versions of LP-DiD perform similarly in this setting in terms of RMSE.

4.3 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 5.

In the synthetic dataset with exogenous treatment timing, the LP-DiD estimator

Figure 4: Simulation 2: Endogenous treatment scenario: 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. To filter out variation in estimates due to variation in the true treatment effect across replications, we subtract from each estimate the true effect, and then add back the average true effect across all replications. This adjustment is in order because in the 'endogenous treatment' setting, the average treatment effect is not deterministic.

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. With endogenous treatment timing, LP-DiD runs in around 0.7 seconds and re-weighted LP-DiD in around 16 seconds, while the BJS estimator requires approximately 7.5 seconds. The CS and SA estimators are computationally demanding: they require respectively 178 and 903 seconds.[30]

[30]CS and SA are more computationally demanding in the simulation with endogenous treatment timing relative to the one with exogenous treatment because by construction there is a larger number of treatment cohorts. Reweighted LP-DiD is relatively slower in the simulation with endogenous treatment timing than under exogenous treatment because of the presence of covariates. Indeed with control variables other than time effects, reweighted LP-DiD needs to be implemented using regression adjustment, while in the absence of control variables it can be implemented using a (computationally faster) weighted least squares regression (as discussed in Section 2.6).

Table 4: Simulation 2: Endogenous treatment scenario: Empirical standard errors

Event time	ES TWFE	LP-DiD	Rw LP-DiD	PMD LP-DiD	Rw PMD LP-DiD	SA	CS	BJS
-5	8.25	2.11	2.18	1.68	1.72	2.96	1.81	1.93
-4	6.15	2.05	2.09	1.66	1.74	2.86	1.76	1.94
-3	6.12	1.57	1.60	1.48	1.59	2.87	1.80	1.91
0	4.79	1.67	1.67	1.62	1.67	2.32	1.87	1.78
1	6.04	2.23	2.35	2.14	2.19	2.97	2.51	2.14
2	6.24	2.34	2.47	2.11	2.17	2.92	2.74	2.15
3	6.05	2.61	2.70	2.38	2.45	3.15	2.82	2.37
4	6.33	2.58	2.77	2.43	2.61	3.12	2.89	2.58
5	6.57	2.95	3.24	2.69	2.97	3.38	3.24	2.92
6	6.69	3.61	4.10	3.40	3.89	4.04	3.95	3.83
7	6.65	4.06	4.79	3.86	4.55	4.62	4.49	4.45
8	6.68	4.36	5.34	4.18	5.10	5.29	5.07	5.16
9	6.68	4.96	6.24	4.86	6.08	6.24	5.85	6.12
10	7.38	5.74	7.49	5.64	7.30	6.87	7.10	7.35

Notes: Empirical standard errors 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, Jaravel, and Spiess (2021).

Table 5: Computational speed (seconds)

Simulation 1 (exogenous treatment scenario)							
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
Simulation 2 (endogenous treatment scenario)							
ES TWFE	LP-DiD	PMD LP-DiD	Rw LP-DiD	Rw PMD LP-DiD	CS	SA	BJS
.61	.74	.82	16.27	19.03	177.5	902.78	7.48

Notes: Computation times in a single repetition of the simulated datasets described in Section 4, 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 2.7); PMD = pre-mean-differenced (see Sec 2.9); CS = Callaway and Sant'Anna (2020); SA = Sun and Abraham (2020); BJS = Borusyak, Jaravel, and Spiess (2021).

5 Empirical Applications

To illustrate the use of the LP-DiD estimator in practice, we present two empirical applications. In the first, we use the LP-DiD estimator to estimate the effect of banking deregulation laws on the labor share in US States, replicating Leblebicioğlu and Weinberger (2020). In the second, we replicate the Acemoglu, Naidu, Restrepo, and Robinson (2019) country-panel study of the effect of democracy on economic growth.

5.1 Credit and the labor share

In our first empirical study, we replicate the Leblebicioğlu and Weinberger (2020) analysis of the effect of banking deregulation on the labor share in US states.

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). Leblebicioğlu and Weinberger (2020) 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.) Figure 5, which reproduces Figure 1 in Leblebicioğlu and Weinberger (2020), displays the share of US states with a liberalized banking sector.

5.1.1 Conventional TWFE specifications

As a starting point, we consider the following static TWFE specification for the effect of banking deregulation laws, which replicates Leblebicioğlu and Weinberger (2020)'s baseline specification:

$$LS_{st} = \beta_{Bank} Bank_{st} + \beta_{Branch} Branch_{st} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st}\,, \tag{19}$$

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, Leblebicioğlu and Weinberger (2020)

also estimate the following event-study TWFE specification:[31]

$$LS_{st} = \sum_{q=-9}^{9} \beta_{Bank,t+q} \Delta Bank_{s,t+q} + \sum_{q=-9}^{9} \beta_{Branch,t+q} \Delta Branch_{s,t+q} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st} \,. \quad (20)$$

5.1.2 Forbidden comparisons in the TWFE specifications

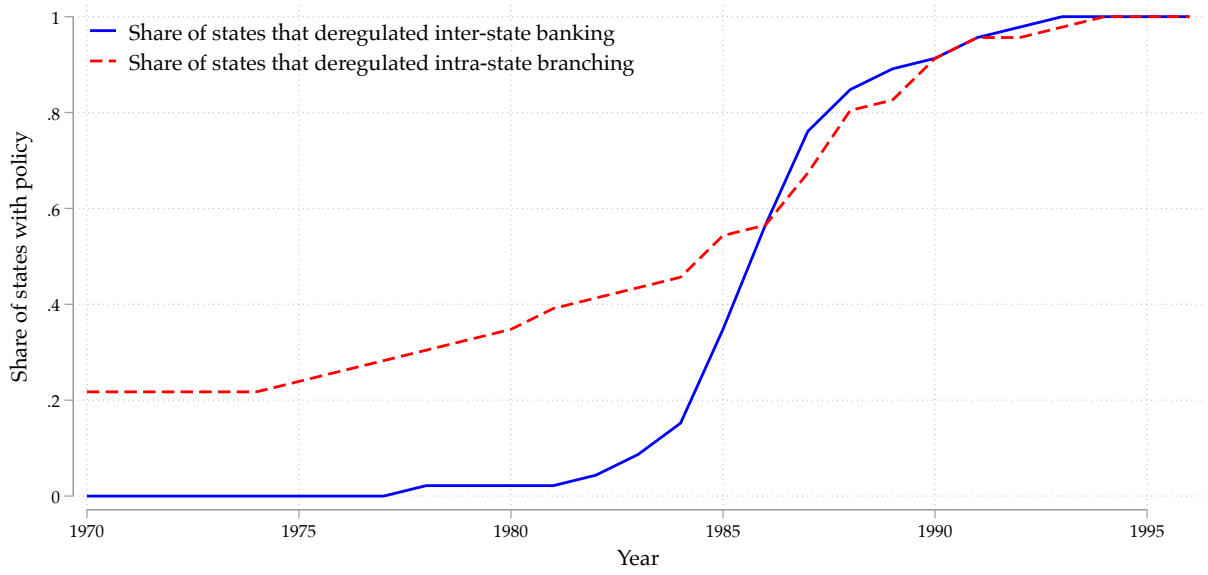
Given the staggered rollout of banking deregulation laws across US states, the TWFE specifications of Equation 19 and Equation 20 suffer from the issues highlighted by recent studies (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). Earlier liberalizers are used as controls for states that liberalize later on. Specifically, the specifications in Equation 19 and Equation 20 produce a weighted average of two types of comparisons: (1) newly treated states vs. not-yet treated states and (2) newly treated states vs. earlier treated states (Goodman-Bacon, 2021).

We employ the Goodman-Bacon (2021) diagnostic to decompose the TWFE estimate from Equation 19 into these two types of comparisons. While 'unclean' 2x2 comparisons with earlier treated units as controls contribute to (and potentially bias) the TWFE estimates of both the policies studied, the estimates of the effect of intrastate branching deregulation are affected most severely. The static TWFE estimator of the effect of interstate banking deregulations assigns a overall weight of 63% to 'clean' comparisons of earlier treated versus not-yet treated states, and 36% to 'unclean' comparisons that use earlier treated units as controls. For the estimates of the effect of intrastate branching deregulations, the problem is much more severe: 'clean' comparisons receive a weight of only 30%. The remaining 70% is accounted for by two types of unclean comparisons: later treated units versus earlier treated units (23%) and treated units versus units that are already treated in the first period of the panel (47%).

Figure 6 displays the results of the Goodman-Bacon (2021) decomposition diagnostic. The figure plots each constituent 2x2 comparison that contributes to the static TWFE estimates of Equation 19, 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

[31] The event-study TWFE specification employed by Leblebicioğlu 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 Leblebicioğlu and Weinberger (2020). This reassures us that any differences between the dynamic TWFE results of Leblebicioğlu 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 Leblebicioğlu and Weinberger (2020).

Figure 5: Banking deregulation in US States



Notes: Data from Leblebicioğlu and Weinberger (2020).

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.

5.1.3 LP-DiD specification

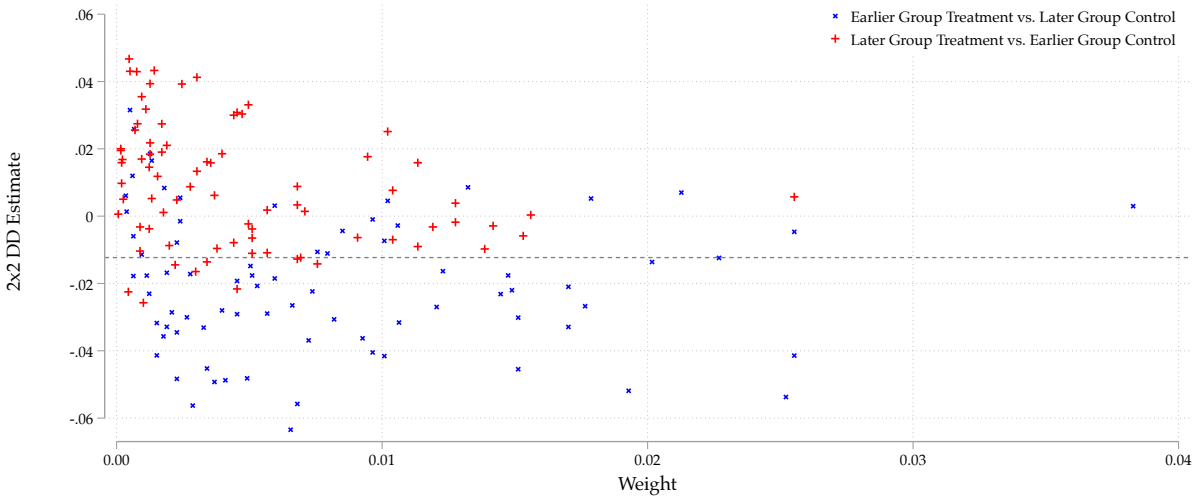
In order to avoid the biases of the conventional TWFE specifications, and to allow for matching based on pre-treatment outcome dynamics, we re-estimate the effect of banking deregulation laws using the following LP-DiD specification:[32]

$$LS_{s,t+h} - LS_{s,t-1} = \alpha_t^h + \beta_h^{LP-DiD} \Delta Bank_{s,t} + \sum_{m=1}^{M} \gamma_m^h \Delta LS_{s,t-m} + \sum_{m=1}^{M} \eta_m^h X_{s,t-m} + e_{s,t}^h , \qquad (21)$$

[32]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 employ the version of the clean control condition which uses only untreated units as controls.

Figure 6: Goodman-Bacon (2021) decomposition diagnostic for the static TWFE specification of equation 19

(a) treatment: interstate banking deregulation



(b) treatment: intrastate branching deregulation

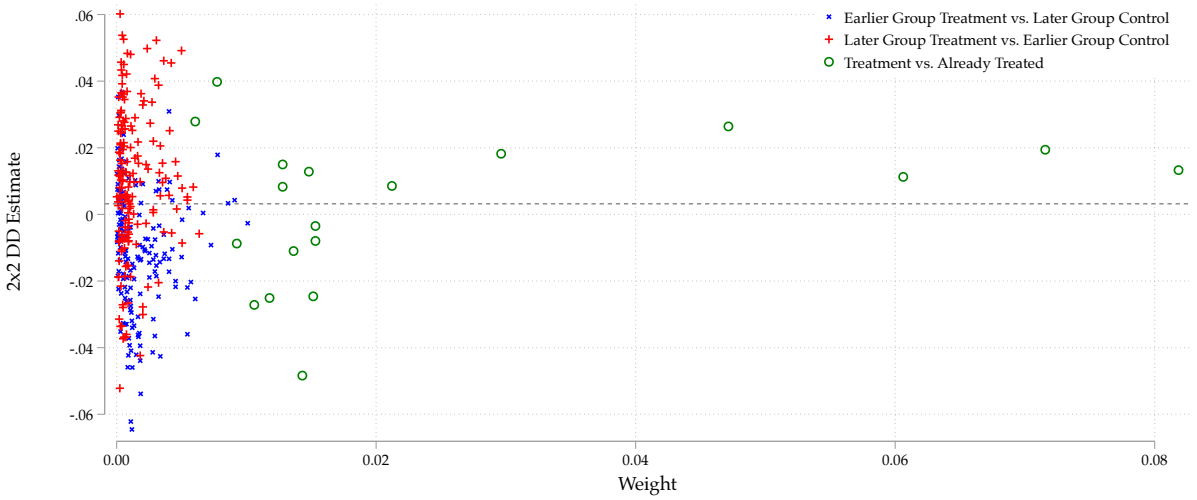
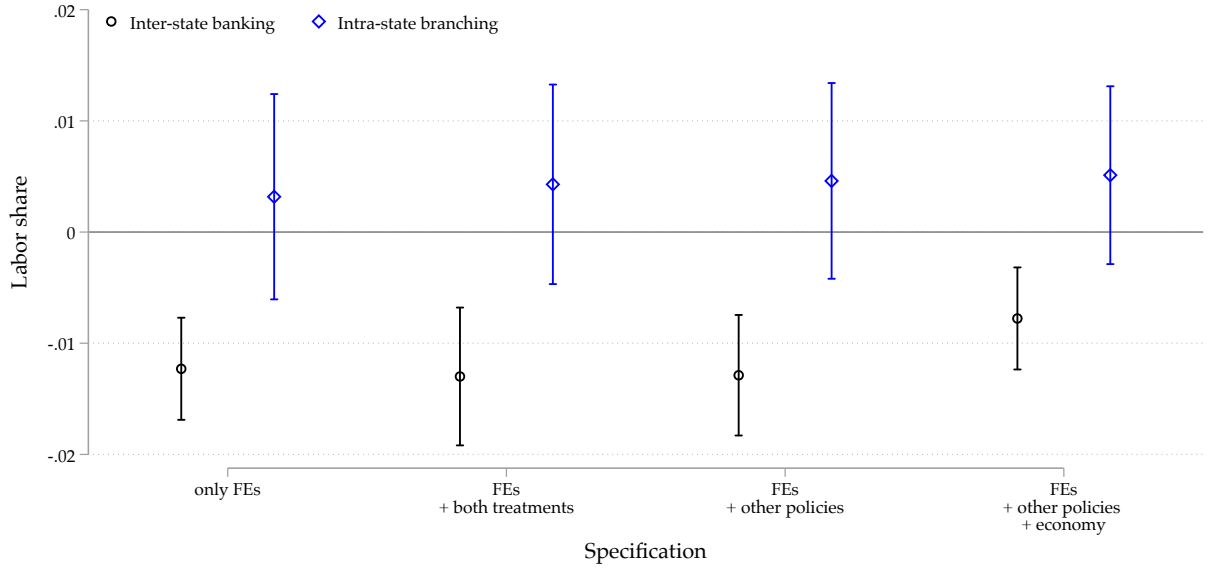


Figure 7: Effect of banking deregulation on the labor share: static TWFE estimates



Notes: Estimates for the effect of banking deregulation on the labor share, using data from Leblebicioğlu and Weinberger (2020) and the static TWFE specification of equation 19. This graph replicates results from Leblebicioğlu and Weinberger (2020).

restricting the sample to observation that are either:

$$\begin{cases} \text{treatment} & \Delta Bank_{s,t} = 1\,, \\ \text{control} & Bank_{s,t+h} = 0\,. \end{cases} \tag{22}$$

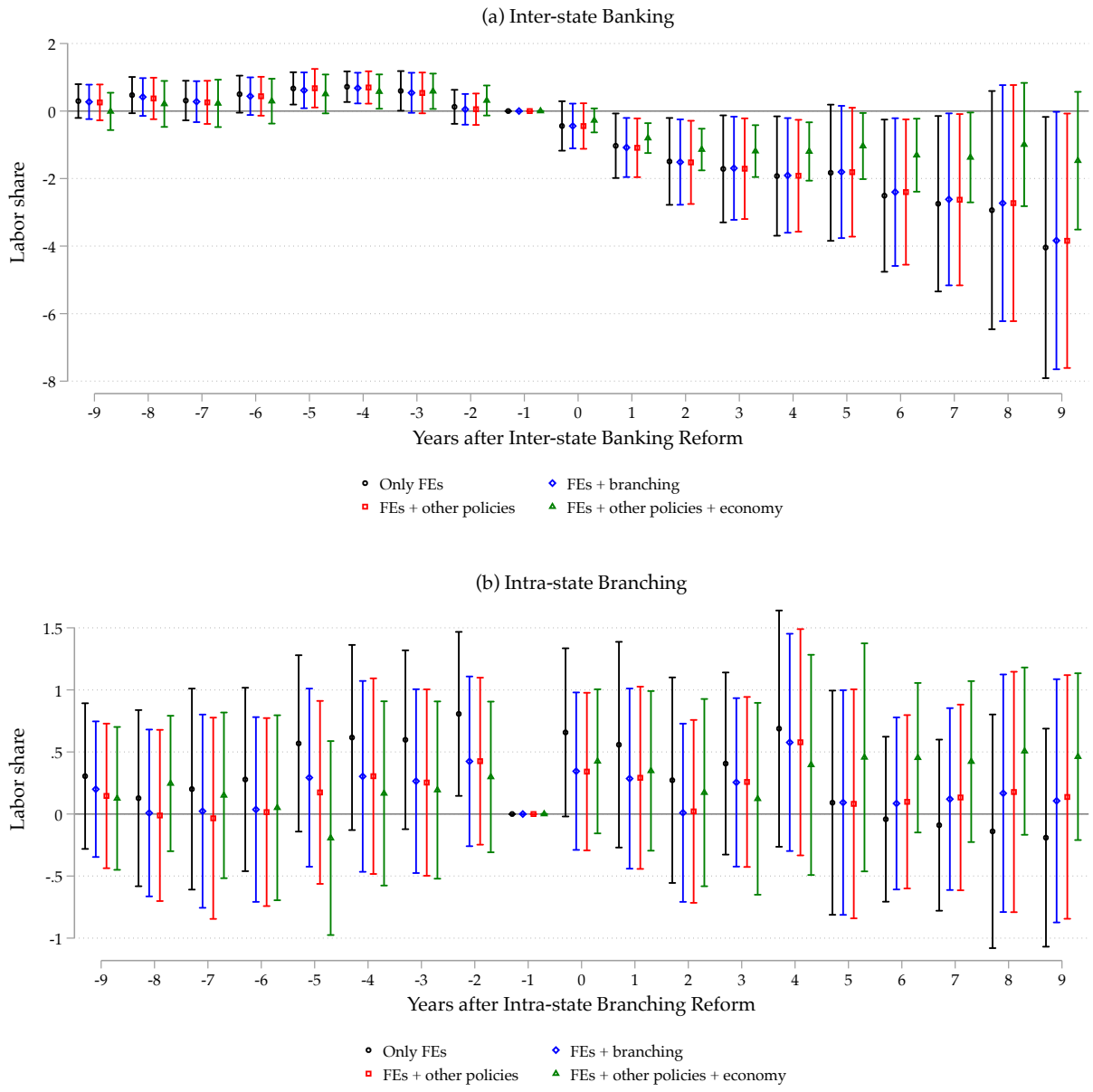
5.1.4 Results

We begin by presenting some replication results where we follow the design of the original study. Figure 7 displays results from the static TWFE specification of Equation 19, while Figure 8 displays results from the event-study TWFE specification of Equation 20. Figure 9, instead, displays results from the LP-DiD estimator with clean controls.

The two sets of TWFE results shown replicate the estimates reported in Table 2 and Figure 2 of Leblebicioğlu and Weinberger (2020). 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 in Figure 9 . The negative effect of inter-state banking deregulation on the labor share is confirmed, including when

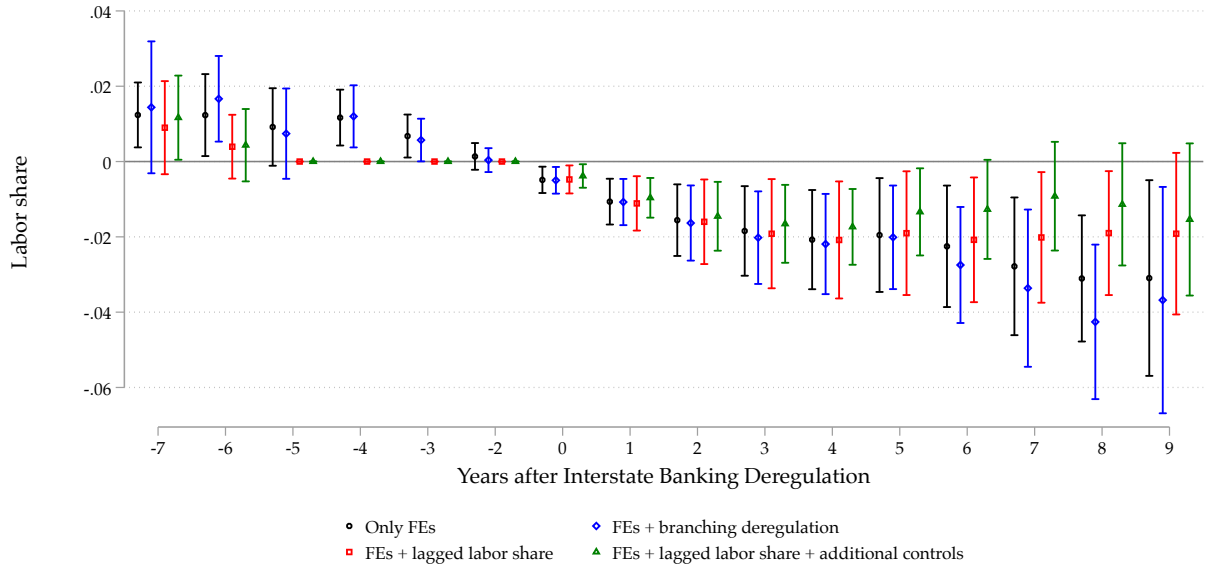
Figure 8: Effect of banking deregulation on the labor share: event-study TWFE estimates



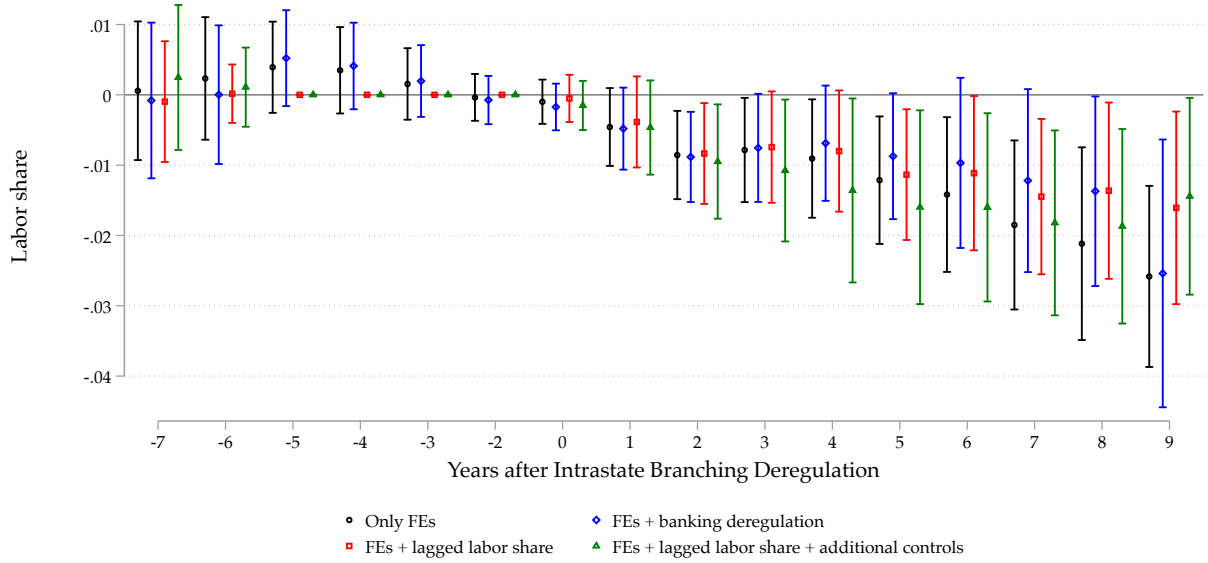
Notes: Estimates for the effect of banking deregulation on the labor share, using data from Leblebicioğlu and Weinberger (2020) and the event-study TWFE specification of equation 20. This graph replicates results from Leblebicioğlu and Weinberger (2020).

Figure 9: Effect of banking deregulation on the labor share: LP-DiD Estimates with clean controls

(a) effect of inter-state banking deregulation



(b) effect of intra-state branching deregulation



Notes: Estimates for the effect of banking deregulation on the labor share, using data from Leblebicioğlu and Weinberger (2020) and the LP-DiD specification of equations 21 and 22. The additional controls are four lags of real State GDP, average corporate tax rate, and union membership rates.

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.

Finally, we can sum up the lessons of this empirical exercise. The results here are not weaker, and in some respects are even stronger, than in the original study. After addressing the downward 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.

5.2 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.

5.2.1 Dynamic panel specifications

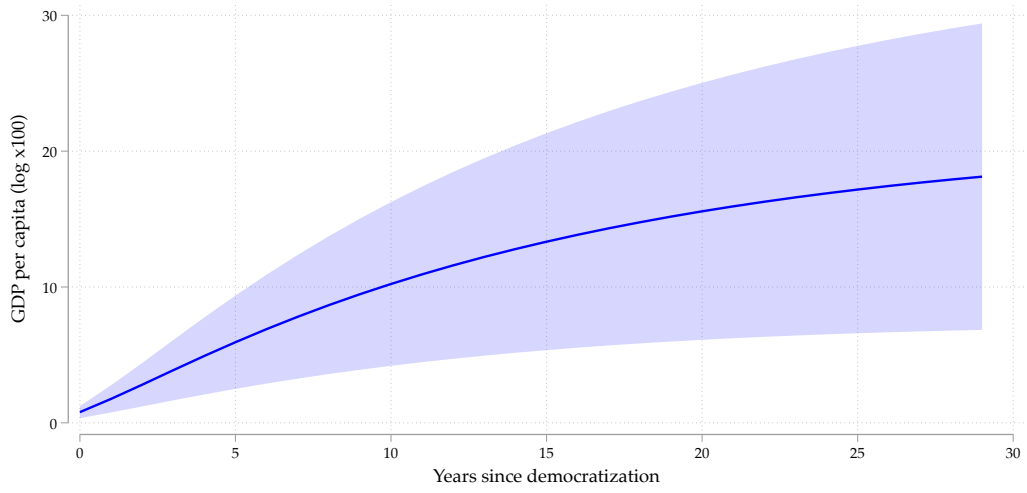
The baseline results in ANRR are obtained from the following dynamic fixed effects specification:

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct}\,, \tag{23}$$

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.

Figure 10: 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 23. This graph replicates the baseline results from Acemoglu et al. (2019).

The estimated coefficients from Equation 23 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 $\hat{\beta}(1-\sum_{j=1}^{p}\hat{\gamma}_j)^{-1}$.

Figure 10 displays the IRF from the estimation of the dynamic panel model of Equation 23. 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 bias if treatment effects are dynamic and heterogeneous, as highlighted in the recent literature.

5.2.2 LP-DiD specifications

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

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP\ DiD}\Delta D_{ct} + \delta_t^h + \sum_{j=1}^{p}\gamma_j^h y_{c,t-j} + \epsilon_{ct}^h, \tag{24}$$

restricting the estimation sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1; D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L\,, \\ \text{clean controls} & D_{i,t-j} = 0 \text{ for } 0 \leq j \leq L\,. \end{cases} \tag{25}$$

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 5 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 control group for new democracies, therefore a control group composed only of continuing autocracies is more likely to satisfy the parallel trends assumption.

In one 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 24–25 with $L = 1$, meaning that their time-window for defining clean controls is 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 2 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.

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 (among others) the 1978 democratization of Spain. 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.

We thus estimate the LP-DiD specification of Equations 24–25 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. This means that Argentina is not part of the control group for the 1978 democratization of Spain.

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 25 but clean controls are defined as observations with $D_{i,t-j} = 0$ for $-h \leq j \leq L$.

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. This test, however, should be interpreted with caution: in this setting conditioning on future treatment status might introduce bias.

5.2.3 Results

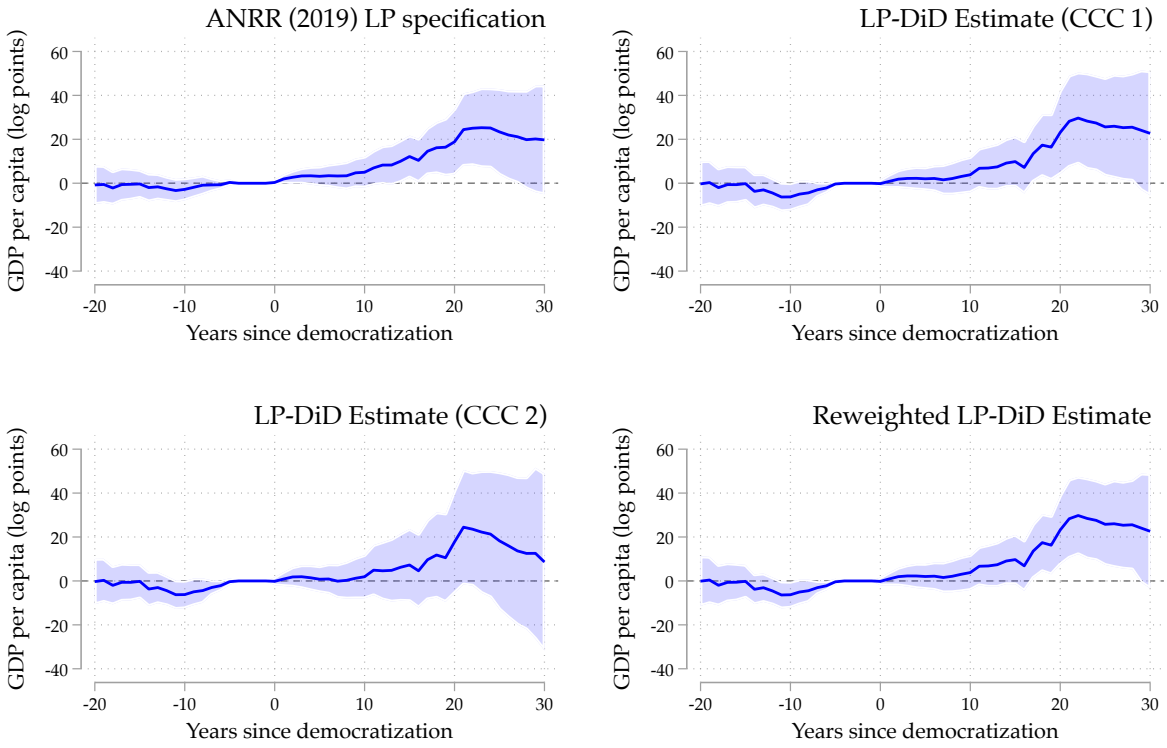
Figure 11 displays results from LP-DiD specifications (Equations 24–25). We present four specifications: The first (top left panel) follows ANRR and sets a time-window of just one year for defining clean controls ($L = 1$). The second (top right) uses a time-window of 20 years ($L = 20$). The third (bottom 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 2.6.

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.

6 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

Figure 11: Effect of democracy on growth: LP-DiD estimates



Notes: LP-DiD estimates for the effect of democracy on GDP per capita, using the dataset of Acemoglu et al. (2019) and the specification of Equations 24-25. The top left panel ('ANRR (2019) LP specification') replicates the results in Section IV of Acemoglu et al. (2019), which use a LP specification and restrict the sample to countries that are either democratizing at year t or non-democracies in both $t-1$ and t. The other three panels use the LP-DiD estimator with a time horizon for defining clean controls of 20 years ($L = 20$). CCC 1 is a clean control condition that defines treated units as countries that democratize in year t and have 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 $t+h$. The right bottom panel uses reweighting to obtain an equally-weighted average effect, adopting CCC 1 as the clean control condition. See main text for more details.

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.

Javier Álvarez and Manuel Arellano. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4):1121–1159, 2003.

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.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. https://arxiv.org/abs/2108.12419, August 2021.

Carolina Caetano, Brantly Callaway, Stroud Payne, and Hugo Sant'Anna Rodrigues. Difference in differences with time-varying covariates. https://arxiv.org/abs/2202.02903, February 2022.

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.

Sylvain Chabé-Ferret. Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1):110–123, 2015.

Shuowen Chen, Victor Chernozhukov, and Iván Fernández-Val. Mastering panel metrics: Causal impact of democracy on growth. *AEA Papers and Proceedings*, 109:77–82, 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'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, April 2021.

Dalia Ghanem, Pedro H. C. Sant'Anna, and Kaspar Wüthrich. Selection and parallel trends. https://arxiv.org/abs/2203.09001, March 2022.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

Òscar Jordà. Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182, 2005.

Òscar Jordà and Alan M. Taylor. The time for austerity: Estimating the average treatment effect of fiscal policy. *Economic Journal*, 126(590):219–255, 2016.

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.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981.

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.

Pedro H. C. Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2020.

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,

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

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

LP-DiD specification

Consider the following LP-DiD specification with clean controls,

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

restricting the sample to observations that are either

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

β_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,

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

where $\Delta\widetilde{D}_{i,p_g}$ is the residual from a regression of ΔD 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

$$\Delta\widetilde{D}_{i,p_g} = \Delta D_{i,p_g} - \frac{\sum_{i\in CCS_{g,h}}\Delta D_{i,p_g}}{N_{CCS_{g,h}}} = D_{i,p_g} - \frac{\sum_{i\in CCS_{g,h}}D_{i,p_g}}{N_{CCS_{g,h}}} = D_{i,p_g} - \frac{N_g}{N_{CCS_{g,h}}}, \tag{A.5}$$

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 $\Delta\widetilde{D}_{i,p_g} = \Delta\widetilde{D}_{g,p_g} = 1 - \frac{N_g}{N_{CCS_{g,h}}}$

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

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

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

where the weights are given by

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

where $n_{g,h} = {}^{N_g}/_{N_{CCS_{g,h}}}$ 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. Recall that τ^g_h 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 more general specification that includes exogenous and pre-determined 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 exogenous or pre-determined covariates. In more general settings, the weights are proportional to the residuals of a regression of the treatment indicator on time effects and the covariates.

To explore the role of covariates, we now assume that no anticipation and parallel trends hold after conditioning on a set of observable exogenous or pre-determined 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,

$$
y_{i,t+h} - y_{i,t-1} = \delta^h_t + \boldsymbol{\rho}_h \Delta \boldsymbol{x}_{it} + \tau_{i,t+h} D_{i,t+h} - \tau_{i,t-1} D_{i,t-1} + e^h_{i,t} \,, \tag{A.7}
$$

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

$$\begin{array}{rll} y_{i,t+h} - y_{i,t-1} = & \beta_h^{LP-DiD} \Delta D_{it} & \text{treatment indicator} \\ & + \boldsymbol{\rho}_h \Delta \boldsymbol{x}_{it} & \text{covariates} \\ & + \delta_t^h & \text{time effects} \\ & + e_{it}^h; & \text{for } h = 0, \ldots, H, \end{array} \tag{A.8}$$

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

$$y_{i,t+h} - y_{i,t-1} - \boldsymbol{\rho}_h \Delta \boldsymbol{x}_{it} = \beta_h^{LP-DiD} \Delta D_{it} + \delta_t^h + e_{it}^h .$$

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

$$E\left(\hat{\beta}_h^{LP-DiD}\right) = \frac{\sum_{j=1}^{G} \sum_{i \in CCS_{j,h}} \left[\Delta \widetilde{D}_{i,p_j} E\left(y_{i,t+h} - y_{i,t-1} - \hat{\rho}_h \Delta \boldsymbol{x}_{it}\right)\right]}{\sum_{j=1}^{G} \sum_{i \in CCS_{j,h}} \Delta \widetilde{D}_{i,p_j}^2}, \tag{A.9}$$

where $\Delta \widetilde{D}_{i,p_g}$ is the residual from a regression of ΔD 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 $\Delta \widetilde{D}_{i,p_g} = \Delta \widetilde{D}_{g,p_g} = 1 - {}^{N_g}/_{N_{CCS_{g,h}}}$

Given the assumptions about the DGP, we have

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

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,

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

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

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

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

The weights are thus given by

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

As noted in the main text (Section 3), it is always possible to preserve non-negative variance-weighted weights by employing semi-parametric propensity-score methods, or obtain a equally-weighted effect using regression adjustment.