NBER WORKING PAPER SERIES

A LOCAL PROJECTIONS APPROACH TO DIFFERENCE-IN-DIFFERENCES

Arindrajit Dube
Daniele Girardi
Òscar Jordà
Alan M. Taylor

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2023, revised May 2025

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

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

A Local Projections Approach to Difference-in-Differences
Arindrajit Dube, Daniele Girardi, Òscar Jordà, and Alan M. Taylor
NBER Working Paper No. 31184
April 2023, revised May 2025
JEL No. C1, C23, C5

ABSTRACT

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

Arindrajit Dube
University of Massachusetts, Amherst
Department of Economics
and NBER
adube@econs.umass.edu

Daniele Girardi
King's College London
daniele.l.girardi@gmail.com

Òscar Jordà
Federal Reserve Bank of San Francisco
and University of California, Davis
and also CEPR
oscar.jorda@gmail.com

Alan M. Taylor
Columbia University
and CEPR
and also NBER
amt2314@columbia.edu

A STATA example files illustrating the implementation of the LP-DiD estimators discussed in this paper can be found here: is available at https://github.com/danielegirardi/lpdid/

1 Introduction

Difference-in-differences (DiD) designs are commonly used to estimate treatment effects with observational data. In canonical form, with two periods and two groups (of which one is treated), and under no anticipation and parallel trends assumptions, the DiD comparison is an unbiased estimator of the average treatment effect on the treated (ATT). In this basic two-periods/two-groups setting, conventional two-way fixed-effects (TWFE) or first-difference regression can be used to implement the DiD method.

However, a recent literature has shown that conventional TWFE implementations can be biased in more general multi-period settings, where the dates of treatment can differ across groups (staggered treatment) and treatment effects can be dynamic and heterogeneous. In this setting, the conventional TWFE implementation has been shown to estimate a weighted average of group-specific effects where the weights may be negative, leading to potentially severe bias. Because of this 'negative weighting' 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 a proliferation of, several alternative estimators.[1]

In this paper we propose a framework for DiD estimation that exploits an important link to *local projections* (LPs), a statistical technique introduced in a time-series context in Jordà (2005). By its very design, and as used in applied macroeconomics, the LP approach was set up to estimate dynamic average effects. 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. Leveraging this feature, we further develop the LP framework to derive a simple to implement, general and flexible regression-based framework for DiD.

Our LP-DiD approach employs LPs to estimate dynamic effects alongside a flexible 'clean control' condition in the spirit of Cengiz et al. (2019) to avoid the negative weighting bias of conventional fixed-effects estimators. Intuitively, negative weighting 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 ones. The LP-DiD clean control condition avoids this bias by restricting the estimation sample so that 'unclean' observations, whose outcome dynamics are still potentially influenced by previous changes in treatment status, are not in the control group.

Under the usual DiD assumptions, LP-DiD provides an unbiased estimate of a convex

[1] See for example Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), Sun and Abraham (2020), Callaway and Sant'Anna (2020), Borusyak, Jaravel, and Spiess (2024), Chaisemartin, D'Haultfœuille, et al. (2022), Gardner et al. (2024), Baker, Larcker, and Wang (2022), and Wooldridge (2021). Surveys of this 'new DiD' literature are provided by Roth et al. (2023) and Chaisemartin and D'Haultfœuille (2022)

weighted average of cohort-specific effects. We explicitly characterize the weights assigned to each cohort-specific effect, and show that they are always positive and depend on treatment variance and subsample size. As we will explain, however, it is easy to implement different weighting schemes to obtain alternative estimands – including the equally-weighted ATT. It is also easy to obtain a 'pooled' estimate, averaged over the full post-treatment window.

We then extend the LP-DiD approach to settings with covariates or non-absorbing treatment. Under conditional parallel trends, a regression-adjustment LP-DiD specification yields an unbiased estimate of the ATT. We also clarify the additional assumptions under which covariates can be added directly to an OLS LP-DiD specification. To illustrate how our framework can accommodate non-absorbing treatment, we derive (i) a LP-DiD estimator for the effect of first-time treatment entry, and (ii) a LP-DiD estimator for the average effect of a treatment event under an additional *effect stabilization* assumption.

Importantly, several recent DiD estimators can be reproduced as specific sub-cases of our general approach based on either weights assigned to treatment events, or the choice of a base period for constructing LPs. In a baseline variance-weighted version, the LP-DiD estimate is numerically equivalent to the estimate from a stacked regression approach as implemented in Cengiz et al. (2019). 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). Yet another version of LP-DiD, reweighted and with an alternative pre-treatment base period, is very close to the Borusyak, Jaravel, and Spiess (2024) imputation estimator. However, the LP-DiD implementation is simpler and computationally faster, and can be more easily generalized for non-absorbing treatment.

Moreover, all the estimators proposed in this paper allow for standard statistical inference using well understood techniques, already incorporated in common statistical software. For this reason, we don't discuss statistical inference here.

Evidence from a Monte Carlo simulation suggests that LP-DiD performs well in staggered DiD settings, also in comparison with other recent estimators. We consider a binary staggered treatment with dynamic and heterogeneous effects, where the parallel trends assumption holds but conventional TWFE performs poorly because of negative weighting bias. We show that LP-DiD performs similarly to the Sun and Abraham (2020), Callaway and Sant'Anna (2020) and Borusyak, Jaravel, and Spiess (2024) estimators, while being computationally faster.

We 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 et al. 2019), two important empirical applications where conventional TWFE estimates are potentially subject to negative weighting bias. These applications demonstrate the applicability and simplicity of implementation of the LP-DiD approach both under absorbing and under non-absorbing treatment, and both when the parallel trends assumption holds unconditionally, and when it holds only conditional on pre-determined covariates – including, when appropriate, pre-treatment outcomes.

Our proposed LP-DiD estimator is, in our view, a useful addition to the growing toolkit of DiD techniques available to researchers. First, the LP-DiD framework is especially simple and computationally fast to implement. Its regression-based formulation allows estimation and inference using well-understood standard methods. Second, the clean control condition employed by LP-DiD defines the appropriate set of treated and control observations in a way that is transparent and therefore easy to understand, communicate, and evaluate. Third, the LP-DiD framework is especially flexible: it is easy to flexibly adapt the clean control condition to specific empirical settings; to implement different weighting schemes; to employ alternative pre-treatment base periods; and to pool estimates over a desired pre- or post-treatment horizon. Moreover, in estimating pre-treatment trends, LP-DiD does not suffer from the potential difficulties of interpretation highlighted in Roth (2024), since pre- and post-treatment coefficients are estimated symmetrically.

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

2 Local Projections and Difference-in-Differences

In this section we clarify the connection between difference-in-differences (DiD) designs and the local projections (LPs) method. While we start from simpler settings (Sections 2.2 to 2.3), the core of this section is the discussion of the case of binary staggered treatment with dynamic and heterogeneous treatment effects (Section 2.4). In this setting, negative weighting bias analogous to that incurred by conventional two-way fixed effects regression would apply to a naive LP implementation. Section 3 will then show how a properly specified LP approach (which we call LP-DiD) successfully addresses these problems and yields a simple to implement, general and flexible regression-based framework for DiD.

2.1 General setup and notation

Assume that an outcome y_{it} is observed for $i = 1, ..., N$ units over $t = 1, ..., T$ periods. Units can receive a binary treatment, denoted by $D_{it} \in \{0, 1\}$. For now, treatment is permanent (or *absorbing*), that is, $D_{is} \leq D_{it}$ for $s < t$. We let p_i denote the period in which unit i enters treatment, 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 it 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 $y_{it} = y_{it}(0) + \sum_{p=1}^{T} \left(y_{it}(p) - y_{it}(0) \right) \times \mathbf{1}\{p_i = p\}$. The treatment effect at time t for unit i which enters treatment at time $p_i \neq \infty$ is defined as $\tau_{it} \equiv E[y_{it}(p_i) - y_{it}(0)]$.

We are interested in estimating some convex average of treatment effects across treated units. To this end, we define the group-specific ATT at time horizon h for group $g \neq 0$ as $\tau_h^g \equiv E\left[y_{i,p_g+h}(p_g) - y_{i,p_g+h}(0) | p_i = p_g \right]$. In other words, τ_h^g represents the average dynamic effect, h periods after entering treatment, across all units that enter treatment at time p_g.[2]

Given our focus on DiD designs, we adopt a model-based framework, where causal identification hinges on adoption of a model for untreated potential outcomes. In particular, we will make use of the assumptions of parallel trends and no anticipation, the two essential assumptions that underpin the DiD approach.

Assumption 1. No anticipation

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

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

Assumption 2. Parallel trends

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

[2] Similar notation is used, for example, in Callaway and Sant'Anna (2020) and Sun and Abraham (2020). The τ_h^g object is analogous to the cohort-specific ATT (CATT) defined in Sun and Abraham (2020).

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. See Ghanem, Sant'Anna, and Wüthrich (2024) and Marx, Tamer, and Tang (2024) for discussions of the underlying selection mechanisms that are compatible with Assumption 2.

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. Specifically, we assume

$$E[y_{it}(0)|i,t] = \alpha_i + \delta_t \,, \tag{1}$$

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 the following three regression specifications of interest.

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

$$y_{it} = \alpha_i^{STWFE} + \delta_t^{STWFE} + \beta^{STWFE} D_{it} + e_{it}^{STWFE} \,, \tag{2}$$

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

Specification 2. Dynamic two-way fixed-effects regression (dynamic TWFE)

$$y_{it} = \alpha_i^{ETWFE} + \delta_t^{ETWFE} + \sum_{h=-Q}^{H} \gamma_h^{ETWFE} D_{i,t-h} + e_{it}^{ETWFE} \,; \quad Q, H \geq 0 \,, \tag{3}$$

where $\beta_h^{ETWFE} = \sum_{j=0}^{h} \gamma_j^{ETWFE}$ provides the dynamic TWFE estimate for the effect at horizon h after treatment ($0 \leq h \leq H$), and $\beta_{-h}^{ETWFE} = -\sum_{j=-h}^{-1} \gamma_j^{ETWFE}$ is likewise an estimate of possible pre-trends at horizon h before treatment ($-Q \leq h \leq -1$).[3]

Specification 3. Local Projections regression (LP)

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP} \Delta D_{it} + e_{it}^h \,; \quad \text{for } h = -Q, \dots, 0, \dots, H; \quad Q, H \geq 0; \quad h \neq -1 \,. \tag{4}$$

[3]An equivalent specification of the dynamic 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. Yet another specification of dynamic TWFE, often used in applications, uses the first difference of the treatment indicator ΔD and normalizes estimates by subtracting the coefficient on the first lead.

As a result of the differencing, the LP specification does not include unit fixed effects. Moreover, a separate regression is estimated for each time horizon h.

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

In a basic 2-groups/2-periods (2x2) setting, which in terms of our notation implies $G = 1$ and $T = 2$, the link between LP and DiD is trivial. In this setting, the LP specification of Equation 4 at horizon $h = 0$ is equivalent to a first-difference regression or a static TWFE regression, both widely-used DiD implementations, and it unbiasedly estimates the ATT under Assumptions 1 and 2. Similarly, in a setting with two groups and multiple time periods ($G = 1$, $T \geq 2$), in which all treated units enter treatment in the same time period, it is straightforward to show that the LP regression in Equation 4 is a way to implement the DiD method and recover the dynamic ATT. We don't delve into these points in the main text because they are straightforward and well known, but a detailed discussion is provided in Online Appendix A.

2.3 Staggered treatment adoption with homogeneous treatment effects

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

In this setting, we still have that a LP specification, if augmented with an adequate number of lags and leads of the treatment indicator, is able to consistently estimate the average treatment effect path under Assumptions 1 and 2. Specifically, the dynamic ATT τ_h is consistently estimated by the β_h^{LP} coefficient in the following LP regression,

$$y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP} \Delta D_{it} + \sum_{\substack{j=-h \\ j \neq 0}}^{\infty} \theta_j^h \Delta D_{i,t-j} + e_{it}^h . \tag{5}$$

A detailed formal discussion supporting this statement is in Online Appendix A.

2.4 Staggered treatment adoption with heterogeneous treatment effects

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

It is now well understood that in this setting conventional (static or dynamic) TWFE specifications can be biased (Roth et al., 2023; Chaisemartin and D'Haultfœuille, 2022).

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

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

Subtracting $E[y_{i,t-1}|i,t]$ from both sides and defining $\delta_t^h = \delta_{t+h}-\delta_{t-1}$, we have

$$
\begin{aligned}
E[y_{i,t+h}-y_{i,t-1}|i,t,h] &= \delta_t^h &&+ \textstyle\sum_{g=1}^{G}\left[\tau_h^g \times \Delta D_{i,t}\times \mathbf{1}\{t=p_g\}\right] \\
& &&+ \textstyle\sum_{g=1}^{G}\left[\sum_{j=1}^{\infty}\left((\tau_{h+j}^g-\tau_{j-1}^g)\times \Delta D_{i,t-j}\times \mathbf{1}\{t=p_g+j\}\right)\right] \\
& &&+ \textstyle\sum_{g=1}^{G}\left[\sum_{j=1}^{h}\left(\tau_{h-j}^g \times \Delta D_{i,t+j}\times \mathbf{1}\{t=p_g-j\}\right)\right].
\end{aligned}
\tag{6}
$$

Without appropriate adjustment to take into account the last two sums on the right-hand side of Equation 6, the LP regression of Equation 4 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. In this special case, we have

$$
\begin{aligned}
E[\hat{\beta}_h^{LP}] &= E[y_{i,t+h}-y_{i,t-1}|\Delta D_{it}=1]-E[y_{i,t+h}-y_{i,t-1}|\Delta D_{it}=0] \\
&= E\left[\textstyle\sum_g^G(\tau_h^g \times \mathbf{1}\{t=p_g\})|\Delta D_{it}=1\right] \\
&\quad - E\left[\textstyle\sum_{g=1}^{G}\left[\sum_{j=1}^{\infty}\left((\tau_{h+j}^g-\tau_{j-1}^g)\times \Delta D_{i,t-j}\times \mathbf{1}\{t=p_g+j\}\right)\right]|\Delta D_{it}=0\right] \\
&\quad - E\left[\textstyle\sum_{g=1}^{G}\left[\sum_{j=1}^{h}\left(\tau_{h-j}^g \times \Delta D_{i,t+j}\times \mathbf{1}\{t=p_g-j\}\right)\right]|\Delta D_{it}=0\right].
\end{aligned}
\tag{7}
$$

Equation 7 shows that, without appropriate adjustment, the LP regression of Equation 4 suffers from two sources of bias.[4]

The first source of bias is the presence of previously treated units in the control group, i.e., observations such that $\Delta D_{it} = 0$ but $\Delta D_{i,t-j} \neq 0$ for some $j \geq 1$. These previously treated units contribute to the estimated counterfactual for units entering treatment at time t, as if they were untreated, although they might in fact be experiencing dynamic treatment effects. This bias exists as long as, for some treatment cohort g at some time-horizon $h+j$, we have $\tau^g_{h+j} \neq \tau^g_{j-1}$, meaning that treatment effects evolve gradually over time. Any dynamic changes in treatment effects that these previously treated units might be experiencing enter Equation 7 with a *negative* sign. This is a manifestation of the negative weighting bias discussed in the recent literature on DiD.

Moreover, in the LP setting, a second potential source of bias is the presence in the control group of units that are treated between $t+1$ and $t+h$, i.e., observations such that $\Delta D_{it} = 0$ but $\Delta D_{i,t+j} \neq 0$ for some j in $1 \leq j \leq h$.[5]

To summarize: with staggered treatment, a naive LP regression might suffer from the 'negative weights' problem highlighted by recent studies, arising from unclean comparisons using previously treated units as controls.

3 LP-DiD estimator

This section presents our main contribution: a properly specified LP regression, which we call *LP-DiD*, consistently estimates a convex weighted average treatment effect without incurring in the negative weighting problem. The key is to restrict comparisons to 'clean' treated and control units. After presenting the LP-DiD specification (Section 3.1), we explicitly characterize the weights assigned to each cohort-specific effect, and show that they are non-negative and proportional to group size and treatment variance (Section 3.2). We then show how a simple re-weighted LP regression yield an unbiased estimate of the equally-weighted ATT (Section 3.3). Moreover, we show how to employ alternative

[4]Chaisemartin and D'Haultfœuille (2024, pp. 33-34) present similar negative results about a naive application of LP with panel data, although they consider a LP specification in levels, with $y_{i,t+h}$ (rather than $y_{i,t+h} - y_{i,t-1}$) as the dependent variable and with D_{it} (rather than ΔD_{it}) on the right-hand side. The LP specification in levels studied by Chaisemartin and D'Haultfœuille (2024) has sometimes been used in applied work and suffers from additional problems, which they analyse in detail.

[5]As Equation 6 suggests, one solution would be a LP regression that estimates separately the effect for each group by interacting group indicators with ΔD_{it}, while controlling for interaction terms between group indicators and the leads and lags of ΔD_{it}. One could then obtain the ATT as an average of the estimated group-specific effects. This solution could be fruitful in some settings and is similar to the Sun and Abraham (2020) estimator. However, it involves estimating a potentially very large number of interaction terms, which coefficients are of no economic interest. Our aim in this paper is to show that it is possible to directly estimate a convex combination of cohort-specific effects, without having to first estimate them separately and then aggregate.

pre-treatment base periods (Section 3.4), pool estimates over a post-treatment horizon (Section 3.4), and avoid composition effects (Section 3.6). Finally, we lay out the relation with other recent DiD estimators (3.7).

We consider the setup, notation and identification assumptions introduced in the previous section, and in particular in Section 2.4: treatment is binary, staggered and absorbing; the no anticipation and parallel trends assumptions hold unconditionally; and treatment effects can be dynamic and heterogeneous.[6]

3.1 LP-DiD specification

LP-DiD consists, in essential form, of estimating the LP specification of Equation 4 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 6 and Equation 7 illustrate, negative weighting bias comes from unclean comparisons in which previously treated units are used as controls for newly-treated units. Excluding these 'unclean' (or 'forbidden', in the terminology of Roth et al., 2023) observations from the control group eliminates the bias.

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

LP-DiD regression Estimate the regression

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

restricting the estimation sample to observations that are either

$$
\begin{cases} \text{newly treated:} & \Delta D_{it} = 1 \,, \\ \text{or clean control:} & D_{i,t+h} = 0 \,. \end{cases} \tag{8}
$$

[6] The LP-DiD approach presented in this Section can be valuable also in settings in which treatment effects are assumed to be homogeneous. As discussed above (Section 2.3), 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.

The clean control condition in Equation 8 ensures that the estimate is obtained from a set of clean comparisons between units entering treatment at t and units that are not yet treated at $t+h$. As a result, the β_h^{LP-DiD} coefficient consistently estimates a convex combination of all group-specific effects τ_h^g. Before analysing the general case (in Section 3.2 below), it is instructive to consider again the special case with $\delta_t^h = \delta^h$. In this special case, we have

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

3.2 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 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), LP-DiD consistently estimates a weighted average of all cohort-specific treatment effects, with weights that are always positive and depend on treatment variance and subsample size. Here we present this result. A formal derivation based on the Frisch-Waugh-Lovell theorem is provided in Online Appendix B.

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 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 8. 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} consistently estimates the following weighted average effect,

$$
E(\hat{\beta}_h^{LP-DiD}) = \sum_{g \neq 0} \omega_{g,h}^{LP-DiD} \tau_h^g . \tag{9}
$$

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

$$
\omega_{g,h}^{LP-DiD} = \frac{N_{CCS_{g,h}} n_{g,h}(1 - n_{g,h})}{\sum_{g \neq 0} N_{CCS_{g,h}} n_{g,h}(1 - n_{g,h})} , \tag{10}
$$

where $N_{CCS_{g,h}}$ is the number of observations in $CCS_{g,h}$, and $n_{g,h} \equiv N_g / N_{CCS_{g,h}}$ is the share of treated units in $CCS_{g,h}$.[7]

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

3.3 Obtaining an equally-weighted average effect

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

Equations 9-10 imply that estimation of the LP-DiD specification through weighted least squares, assigning to an observation belonging to $CCS_{g,h}$ a weight equal to $1/(\omega_{g,h}^{LP-DiD}/N_g)$, yields unbiased estimation of the equally-weighted ATT.

In practical applications, the weight $(\omega_{g,h}^{LP-DiD}/N_g)$ can be obtained by computing subsamples sizes and shares of treated units and using Equation 10, or through an auxiliary regression. Specifically, consider an auxiliary regression of ΔD on time indicators in the estimation sample defined by Equation 8. Define $\Delta\widetilde{D}_{g,p_g}$ as the residual at time p_g for a unit belonging to group g. (Of course, $\Delta\widetilde{D}_{g,p_g}$ will be identical for all units belonging to the same group.) The Frisch-Waugh-Lovell theorem implies that

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

(Further discussion can be found in Online Appendix B.)

Another equivalent way to estimate the equally-weighted ATT is to estimate the LP-DiD specification through regression adjustment (RA). Here, RA uses clean control units to estimate a counterfactual outcome change for each treated unit, and then computes an average estimated effect assigning equal weight to each treated unit, thus estimating the equally-weighted ATT.

In practice, a RA LP-DiD specification can be implemented as follows. Regress $y_{i,t+h} - y_{i,t-1}$ on time indicators using only clean control observations (i.e., observations with $D_{i,t+h} = 0$). Use the estimated coefficients to get a predicted value in absence of treatment $\hat{E}\left(y_{i,t+h} - y_{i,t-1} \mid D_{i,t+h} = 0\right)$ for each treated unit. The ATT is then estimated by

$$\hat{\beta}_h^{LP-DiD,RA} = N_{TR}^{-1} \sum_{\{i,t\}\in TR} \left[(y_{i,t+h} - y_{i,t-1}) - \hat{E}\left(y_{i,t+h} - y_{i,t-1} \mid D_{i,t+h} = 0\right)\right],$$

[7] The weights in Equation 10 are equivalent to those of the stacked DiD estimator (Gardner et al., 2024, Appendix E).

where TR is the set of newly treated observations (i.e., the set of observations with $\Delta D_{i,t} = 1$).[8] This RA implementation of LP-DiD constitutes an imputation estimator, in the same sense as in Borusyak, Jaravel, and Spiess (2024).

There is, in general, a bias-variance tradeoff between variance-weighting and equal weighting: reweighted LP-DiD applies equal weights and is therefore unbiased, while the variance-weighted version has some bias but also lower variance. This tradeoff will be further discussed and illustrated in the context of the simulation in Section 5.

3.4 Alternative pre-treatment base periods

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

The reason for considering such an alternative specification is a possible efficiency-related concern with the LP-DiD specification of Section 3.1. Typically, a LP uses the long difference $y_{i,t+h} - y_{i,t-1}$ as the dependent variable. Period $t-1$ is thus used as the pre-treatment base period: for a treatment event occurring at time s, the expected value of the outcome in the pre-treatment period in the treated group and its clean controls are estimated from $y_{i,s-1}$. However, the number of time periods available for estimating the expected value of the outcome in the pre-treatment period is larger than just $s-1$: observations for all time periods $t < s$ can potentially be used. For this reason, using a single pre-treatment period as the baseline may be inefficient.

This concern can be accommodated by using the following 'pre-mean-differenced' (PMD hereafter) specification of LP-DiD.

PMD LP-DiD regression

$$y_{i,t+h} - \frac{1}{k}\sum_{\tau=t-k}^{t-1} y_{i,\tau} = \beta_h^{PMD\ LP-DiD} \Delta D_{it} + \delta_t^h + e_{it}^h \quad \text{for } h = -Q, \dots, 0, \dots, H\,,$$

[8] 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 outcome variable; h is the time-horizon of the estimate; `Dhy` $= y_{i,t+h} - y_{i,t-1}$; `dtreat= D.treat` 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.

again restricting the estimation sample to observations that are either newly treated ($\Delta D_{it} = 1$) or clean controls ($D_{i,t+h} = 0$). 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 Online Appendix B imply that $\beta_h^{PMD\ LP\text{–}DiD}$ recovers a convex weighted average of cohort-specific effects, with the same weights $\omega^{LP\text{–}DiD}$ discussed in Section 3.2. Also in this case, weighted regression or regression adjustment can be employed to obtain an equally-weighted ATT (Section 3.3).

The potential advantages and risks of differencing with respect to the pre-treatment average ('pre-mean differencing') relative to differencing over a single lag ('first-lag differencing') have been discussed in the recent literature (a review of these discussions is provided in Chaisemartin and D'Haultfœuille 2022, pp. 18–19). The potential advantage of pre-mean differencing is the efficiency gain discussed above. This advantage is greater the lower the autocorrelation in untreated potential outcomes (Borusyak, Jaravel, and Spiess, 2024; Harmon, 2022). Indeed, Borusyak, Jaravel, and Spiess (2024) prove that in the extreme case of no autocorrelation, differencing with respect to the pre-treatment average is more efficient than differencing over the last pre-treatment lag. However, Harmon (2022) finds that in the opposite polar case where errors follow a random walk (very high autocorrelation), this efficiency ranking reverses, with differencing over the last pre-treatment period being more efficient.

A potential risk is that, under some deviations from the parallel trends assumption, pre-mean differencing can amplify the bias relative to first-lag differencing. If parallel trends holds between periods s and $s + h$ (s being the time of treatment), but not in earlier pre-treatment periods, first-lag differencing will still be unbiased, while pre-mean differencing will be biased. In this sense, first-lag differencing relies on a weaker parallel trends assumption than pre-mean differencing (Marcus and Sant'Anna, 2021). Moreover, if parallel trends does not fully hold at any time period, and the gap in average untreated potential outcomes between treated and controls increases over time, then pre-mean differencing will be more biased than first-lag differencing.

The ability of LP-DiD to allow researchers to flexibly choose the appropriate pre-treatment base period, and to easily test robustness to alternative choices, is an additional advantage of LP-DiD. Other available DiD estimators (for example Callaway and Sant'Anna, 2020, Sun and Abraham, 2020 and Borusyak, Jaravel, and Spiess, 2024) are less flexible in this respect. Of course, as with all specification choices, researchers should commit to a preferred choice of pre-treatment base period ex-ante, based on the features of their application and independent of estimation results, to avoid introducing bias.

3.5 Pooling over post-treatment periods

Instead (or in addition to) an impulse response tracing the average dynamic effect path, 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 estimates an average effect over the post-treatment window.

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: $\frac{1}{H+1}\sum_{h=0}^{H} y_{i,t+h} - \frac{1}{k}\sum_{\tau=t-k}^{t-1} y_{i,\tau}$. In some applications, researchers might want to simply use $\sum_{h=0}^{H} y_{i,t+h} - y_{i,t-1}$ as the dependent variable (without dividing by $H+1$), to cumulate effects over the time horizon.

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).

3.6 Composition effects

In finite samples, the LP-DiD specification of Section 3.1 might suffer from composition effects because the set of treated and clean control units can change across different time horizons h. A 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 8.

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

3.7 Relation to other DiD estimators

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

First, in the baseline variance-weighted version, 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) (see Online Appendix B).

Moreover, the re-weighted version of LP-DiD (discussed in Section 3.3) is numerically equivalent to the estimator proposed by Callaway and Sant'Anna (2020). 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. With absorbing treatment and no control variables, re-weighted LP-DiD equals the same difference in means.

Futhermore, the PMD version of the LP-DiD estimator (Section 3.4) is analogous to the estimator proposed by Borusyak, Jaravel, and Spiess (2024) (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.[9] With more than one treated group, the BJS estimator does not have a closed form expression, and it is therefore not straightforward to assess with precision its relation to our PMD LP-DiD estimator. However, the pre-period mean differencing means that the two estimators use similar information, and indeed our Monte Carlo simulations (presented in Section 5 below) show that with more than one treated group, when using 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

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

LP-DiD even in the case of more than one treated group, unlike for the BJS estimator.[10] Movever, unlike the BJS estimator, PMD LP-DiD can be implemented using simple OLS (or weighted least squares) regression, employing commonly used and well understood methods for statistical inference.

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

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

It also bears noting that efficiency of the BJS estimator is only guaranteed under the Gauss-Markov assumptions. These require, among other things, no auto-correlation in untreated potential outcomes, a polar assumption that might be seen as implausible in most panel data applications (Harmon, 2022; Chaisemartin and D'Haultfœuille, 2022, p. 18). Moreover, if there is limited heterogeneity in treatment effects between treatment cohorts, variance-weighting (as done by LP-DiD or PMD LP-DiD without reweighting) can be more efficient than equal weights, as the simulation in Section 5 will illustrate.

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 a stacked format. Moreover, as we discuss below, the LP-DiD specification is especially easy to generalize to more complicated settings.

4 Extensions

In this section we extend the LP-DiD approach to include covariates (Section 4.1) and to accommodate non-absorbing treatment (4.2). We then briefly discuss the link between LP-DiD and the impulse responses used in macroeconomics (4.3).

[10] With one single treated group, the PMD LP-DiD estimator (like the BJS estimator) is equal to the expression in foonote 9. With more than one treated group, the PMD LP-DiD estimator is equal to a weighted average of the expression in footnote 9 across all treated groups, with weights given by $\omega_{g,h}^{LP-DiD}$ in Equation 10. The reweighted PMD LP-DiD estimator is equal to a simple average of the expression in footnote 9 across all treated units.

4.1 Inclusion of covariates

The LP-DiD framework can be extended to include covariates. To do so, we consider a setting with staggered treatment adoption, dynamic and heterogeneous treatment effects, and in which parallel trends holds only conditional on a vector of covariates $\mathbf{x}_i$.[11]

Using the notation introduced in Section 2.1, we make the following assumptions:

Assumption 3. Conditional no anticipation

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

Assumption 4. Conditional parallel trends

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

Assumption 5. Linear conditional expectation function

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

Importantly, these assumptions do not restrict how the treatment effect might vary depending on the value of the covariates $\mathbf{x}$.

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

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

and the unconditional (equally-weighted) dynamic ATT as

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

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

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

4.1.1 Regression adjustment LP-DiD specification with covariates

In this setting, a regression-adjustment LP-DiD specification with covariates recovers the equally-weighted ATT.[12] The estimator can be written as follows:

$$\hat{\beta}_{h,\mathbf{x}}^{LP-DiD,RA} = N_{TR}^{-1} \sum_{\{i,t\} \in TR} \left[(y_{i,t+h} - y_{i,t-1}) - \hat{E}\left(y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i \right) \right] =$$

$$= N_{TR}^{-1} \sum_{\{i,t\} \in TR} \left[(y_{i,t+h} - y_{i,t-1}) - \hat{\gamma}^h \mathbf{x}_i - \hat{\delta}_t^h \right],$$

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

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

Under the assumptions above, $\hat{\beta}_{h,\mathbf{x}}^{LP-DiD,RA}$ is an unbiased estimator of ATT^h. This can be shown as follows:

$$\begin{aligned}
E(\hat{\beta}_{h,\mathbf{x}}^{LP-DiD,RA}) &= E\left(N_{TR}^{-1} \textstyle\sum_{\{i,t\} \in TR} \left[(y_{i,t+h} - y_{i,t-1}) - \hat{E}\left(y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i \right) \right] \right) \\
&= E\left[E\left(y_{i,t+h} - y_{i,t-1} | \Delta D_{i,t} = 1, \mathbf{x}_i \right) - E\left(y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i \right) | \Delta D_{i,t} = 1 \right] \\
&= E\left[E\left[y_{i,p_i+h}(1) - y_{i,p_i-1}(0) | p_i \neq \infty, \mathbf{x}_i \right] - E\left[y_{i,t+h}(0) - y_{i,t-1}(0) | p_i = \infty, \mathbf{x}_i \right] | p_i \neq \infty \right] \\
&= E\left\{ E\left[y_{i,p_i+h}(p_i) - y_{i,p_i+h}(0) | p_i \neq \infty, \mathbf{x} = \mathbf{x}_i \right] | p_i \neq \infty \right\} \\
&= ATT^h,
\end{aligned} \tag{11}$$

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

Although we assumed linearity here for ease of exposition (Assumption 5), this approach can also accommodate non-linear models for the conditional expectation function. Researchers can allow for non-linearities by obtaining the estimated conditional expectation $\hat{E}[y_{i,t+h} - y_{i,t-1} | D_{i,t+h} = 0, \mathbf{x}_i]$ from a flexible non-parametric or semi-parametric model. Indeed, equation 11 continues to hold under non-linearity, as long as the conditional expectation function is correctly specified in the first-step regression.

[12] It is well known that a similar result holds in the canonical 2x2 setting (Roth et al., 2023, pp.2230-2231).

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

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

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

Assumption 6. Treatment effects are independent of covariates

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

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

Given that the RA specification presented in Section 4.1.1 requires weaker assumptions, it appears generally preferable in applications.

4.2 Non-absorbing treatment

In many applications, treatment is not absorbing: units can enter and exit treatment. Through appropriate modification of the 'clean control' sample restriction of Equation 8, the LP-DiD framework can accommodate non-absorbing treatment settings.

4.2.1 Setup and notation with non-absorbing treatment

To discuss non-absorbing treatment, we extend the framework of Section 2.1 and allow potential outcomes to be a function of the treatment *path*, as in Robins (1986).[13] Treatment is still binary ($D_{it} \in \{0, 1\}$), but we remove the assumption $D_{is} \leq D_{it}$ for $s < t$. Let $\mathbf{D}_i = (D_{i1}, \ldots, D_{iT})$ denote the vector containing the time series of the treatment variable D over the sample period for unit i. The observed outcome can then be written as $y_{it} = y_{it}(\mathbf{D}_i)$. Let $y_{it}(d_1, \ldots, d_T)$ denote the potential outcome for unit i at time t, if its treatments over the sample period were equal to $(D_{i1}, ..., D_{iT}) = (d_1, \ldots, d_T)$.

[13]Recent articles discussing panel estimation with non-absorbing treatment are for example Chaisemartin and D'Haultfœuille (2024), Bojinov, Rambachan, and Shephard (2021), and Viviano and Bradic (2023). Our setup here is similar to Design 2 in Chaisemartin and D'Haultfœuille (2024), and the estimator of the effect of entering treatment for the first time proposed in Section 4.2.2 below constitutes a regression-based method to perform and aggregate the same clean comparisons considered in that article. Unlike Chaisemartin and D'Haultfœuille (2024), however, we also consider an 'effect stabilization' assumption that allows using treatment events subsequent to the first one in estimation.

Given an estimation sample selected by the researcher, which might coincide with the full sample of N units and T periods or be a subset of it, groups $g \in \{-k, \ldots, 1, \ldots, G\}$, with $k \geq -1$, are now defined so that all units within the same group experience the same treatment path over the estimation sample. We let $\mathbf{D}_g = (D_{g1}, \ldots, D_{gT})$ denote the time series of the treatment variable for group g in the given estimation sample, in which $D_{g,t}$ is a missing value if units in group g are excluded from the estimation sample at time t. Groups $-k$ to 0 denote $k+1$ possible groups such that $\Delta D_{gt} \neq 1$ for all t in the estimation sample. We let p_g^n denote the time period in which group g enters treatment for the n-th time in the estimation sample, with $p_g^n = \infty$ if the group enters treatment less than n times in the estimation sample.

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

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

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

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

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

4.2.2 Estimating the effect of entering treatment for the first time

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

Consider an estimation sample that includes observations that are either

$$\begin{cases} \text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } j \geq 1), \\ \text{or clean control} & D_{i,t-j} = 0 \text{ for } j \geq -h. \end{cases} \tag{12}$$

.

We let $\tau_h^{\prime g}$ denote the h-periods horizon effect of entering treatment for the first time and staying treated for group g, which enters treatment for the first time at p_g^1 and then remains treated until at least period $p_g^1 + h$. Given the estimation sample defined by Equation 12 and the corresponding set of groups, this can be written as

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

A variance-weighted average effect $\sum_{g=1}^{G} \omega_{gh}^{LP-DiD'} \tau_h^{\prime g}$, with strictly positive weights

$\omega_{gh}^{LP-DiD'} = \left(N_{CCS_{gh}}[n_{gh}(1-n_{gh})]\right) / \left(\sum_{g=1}^{G} N_{CCS_{gh}}[n_{gh}(1-n_{gh})]\right)$, can be recovered by estimating the LP specification of equation 4 in the estimation sample defined by Equation 12, where $N_{CCS_{gh}}$ and n_{gh} are defined as in Section 3.2. The derivation of this result is the same as for the case of absorbing treatment in Online Appendix B, with $\tau_h'^g$ substituted for τ_h^g and the clean control condition of equation 12 substituted for the one of equation 8.

Estimating the same specification using either a weighted regression or regression adjustment (as discussed in Section 3.3) recovers the equally-weighted effect.

4.2.3 Estimating average treatment effects under an 'effect stabilization' assumption

In numerous settings of practical importance, the estimand $\tau_h'^g$ introduced above and the resulting 'clean control condition' of Equation 12 might not be feasible or appropriate. Consider, for example, the problem of estimating the effect of minimum wage (MW) increases in a panel of regions. For most MW increase events, there will be very few clean control regions that have never experienced any MW increase until period $t+h$. This case can be dealt with in a simple way in the LP-DiD framework, by focusing on the average effect of treatment events and under the additional assumption that dynamic effects stabilize after a finite number of periods.

Let us first define treatment events.[14] Group g has a treatment event at time j if $\Delta D_{g,j} = 1$. We let N_g^T denote the number of treatment events experienced by group g in the estimation sample.

We then introduce the vectors of counterfactual treatment series $\mathbf{D}_{i,j,0}$ and $\mathbf{D}_{i,j,1}$. We let $\mathbf{D}_{i,j,0}$ denote a $(1 \times T)$ vector which is equal to $\mathbf{D}_i$ up to column $j-1$, while its columns j to T are all equal to 0. Similarly, $\mathbf{D}_{i,j,1}$ is equal to $\mathbf{D}_i$ up to column $j-1$, but its columns j to T are all equal to 1. We thus have $\mathbf{D}_{i,j,0} \equiv (D_{i1}, \ldots, D_{i,j-1}, \mathbf{0})$ and $\mathbf{D}_{i,j,1} \equiv (D_{i1}, \ldots, D_{i,j-1}, \mathbf{1})$.

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

$$\tau_h^{g,n} \equiv E\left[y_{i,p_g^n+h}\left(\mathbf{D}_{i,p_g^n,1}\right) - y_{i,p_g^n+h}\left(\mathbf{D}_{i,p_g^n,0}\right) \mid i \in g, p_g^n \neq \infty\right]$$

Furthermore, we introduce the following assumption:

Assumption 9. Dynamic effects stabilize after L periods:

$$\tau_L^{g,n} = \tau_{L+l}^{g,n} \quad \text{for some } L < T\,, \text{ for all } l \geq 0\,, \text{ and for all treatment events (g,n).}$$

[14] 'Exit' events could be defined, and their effect estimated, in an analogous way.

Under assumptions 7, 8 and 9, a convex variance-weighted effect $\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\omega_{gnh}^{LP-DiD''}\tau_h^{g,n}$ is consistently estimated by the LP-DiD specification of Equation 4 with the following modified sample restriction:

$$\begin{cases} \text{treatment} & (D_{i,t+j}=1 \text{ for } 0\leq j\leq h) \text{ and } (D_{i,t-j}=0 \text{ for } 1\leq j\leq L)\,, \\ \text{or clean control} & \Delta D_{i,t-j}=0 \text{ for } -h\leq j\leq L\,. \end{cases} \tag{13}$$

The (non-negative) weight assigned to a treatment event is

$$\omega_{gnh}^{LP-DiD''}=\frac{\bar{n}_{gnh}N_{CCS_{gnh}}[\hat{n}_{gnh}(1-\hat{n}_{gnh})]}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\bar{n}_{gnh}N_{CCS_{gnh}}[\hat{n}_{gnh}(1-\hat{n}_{gnh})]}\,.$$

In this formula, CCS_{gnh} is the set of units that satisfy the sample restriction in equation 13 at time $t=p_g^n$; $N_{CCS_{gnh}}$ is the number of units in CCS_{gnh}; $\hat{n}_{gnh}$ is the share of newly treated units in CCS_{gnh}; $\bar{n}_{gnh}$ is the share of units belonging to group g among all newly treated units in CCS_{gnh}. See Online Appendix C for proof and further discussion.

Intuitively, units that have experienced no change in treatment status in the time-window between $t-L$ and $t+h$ constitute 'clean controls' for units entering treatment at time t, because their outcome dynamics in this time-window are not influenced by previous treatment events by virtue of Assumption 9. Moreover, only units that enter treatment at t and experience no other change in treatment status between $t-L$ and $t+h$ constitute 'clean treated units' to be used in estimation: this restriction ensures that their outcome dynamics are only influenced by the treatment event at time t and not others.[15]

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

Repeated 'one-off' treatments So far, we have assumed that, following a treatment event, treatment status persists (i.e., $D_{it}=1$) until a potential exit or reversal. An important alternative scenario involves 'one-off' treatments—cases in which treatment is by definition confined to a single period, even though its effects may be dynamic and persistent over time. Natural disasters provide a canonical example: $D_{it}=1$ if unit i is

[15] 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.

hit by a hurricane at time t, and $D_{it} = 0$ in all other periods (although effects may be long-lasting). Depending on the modeling approach, minimum wage increases can also be analyzed within this framework.

When units can experience repeated one-off treatments, average treatment effects can be estimated using a modified version of the sample restriction in equation 13. Specifically, 'clean' treated observations are now defined by $D_{it} = 1$ and $D_{i,t-j} = 0$ for $-h \leq j \leq L$, with $j \neq 0$, while the definition of clean controls remains as specified in equation 13.

In this subsection we have provided empirically relevant illustrations of what one can do with the LP-DiD framework in settings with non-absorbing treatment. Of course, one could consider alternative target estimands and assumptions. A comprehensive discussion of non-absorbing treatment would indeed require a whole article in itself. Our point is that the LP-DiD framework and its clean control condition can be flexibly adapted to accommodate non-absorbing treatment settings. For an application with non-absorbing treatment, see Dube and Lindner, 2024 who use the LP-DiD framework to estimate the impact of state-level U.S. minimum wage laws.

4.3 Identification and relation to impulse responses

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

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

Note that in the specification of the LP-DiD estimator, we condition on future values of treatment (between $t+1$ and $t+h$). This removes the effect of subsequent treatment effects on future values of the outcome. Of course, one could 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, as it would permit the researcher to craft an alternative treatment plan than that usually observed (though in that case, deviations from the usual treatment plan can run afoul of the Lucas critique if they are 'too different').

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

5 Monte Carlo simulation

We present a Monte Carlo simulation to evaluate the performance of the LP-DiD estimator in a setting with binary staggered treatment and heterogeneous treatment effects.

5.1 Setting

We calibrate the simulation based on the empirical application that will be presented in Section 6.1, which estimates the effect of banking deregulation on the wage share.

We simulate yearly observations on the wage share for 46 states over 1970-1996. The potential outcome without treatment equals $y_{it}(0) = \lambda_i \gamma_t e_{it}$, with $e_{it} = (1-\rho)\epsilon_{it} + \rho e_{i,t-1}$, where ϵ is a white noise random shock. We set $\rho = 0.63$ based on the corresponding estimated coefficient using wage share data from the Leblebicioğlu and Weinberger (2020) dataset; λ_i, γ_t and ϵ_{it} follow a beta distribution with parameters estimated empirically using the Leblebicioğlu and Weinberger (2020) dataset.

The treatment follows the same time profile as banking deregulation reforms in US states. The treatment effect is negative and grows over time for four years, after which it stabilizes, and is stronger for early adopters. Specifically, define $\phi_{it} = t - p_i + 1$ and $\mu_i = \frac{p_i - 1970}{min\{p_1,...,p_N\} - 1970}$. The treatment effect is $\beta_{it} = \alpha_0 \left(\frac{1}{\phi_{it}}\right)^{1/4} \left[\alpha_1 \left(\frac{1}{\phi_{it}}\right)^{\frac{1}{3\mu_i}} + (1-\alpha_1)\left(\frac{1}{\phi_{it}}\right)^{\frac{1}{3}}\right]$

for $0 \leq t - p_i \leq 3$; $\beta_{it} = \beta_{i,t-1}$ for $t - p_i > 3$; and $\beta_{it} = 1$ for $t < p_i$. We set $\alpha_0 = 0.995$ and $\alpha_1 = 0.75$.

The observed outcome is $y_{it} = y_{it}(0) \times \beta_{it}$. Given the multiplicative structure of the DGP, we take a log transformation of the outcome $ln(y)$ in estimation.

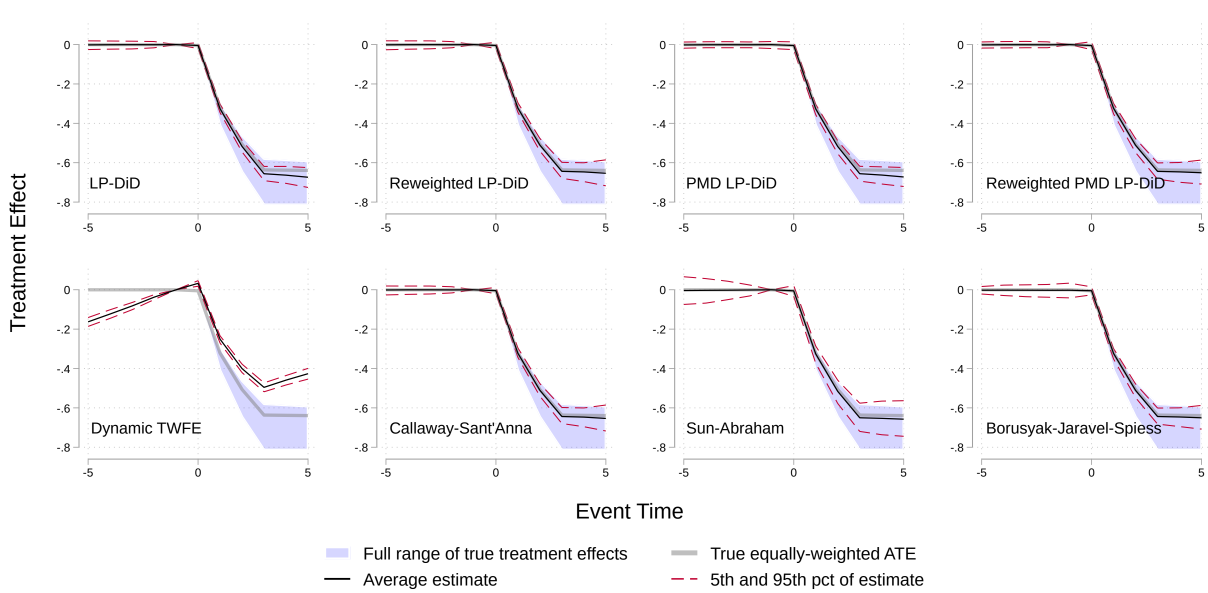
5.2 Estimators

We perform 200 replications and evaluate five estimators: (i) dynamic TWFE; (ii) LP-DiD, using both variance weights and equal weights, and applying both first-lag differencing and pre-mean differencing; (iii) Sun and Abraham (2020) – SA hereafter; (iv) Callaway and Sant'Anna (2020) – CS hereafter; (v) Borusyak, Jaravel, and Spiess (2024) – BJS hereafter.

5.3 Results

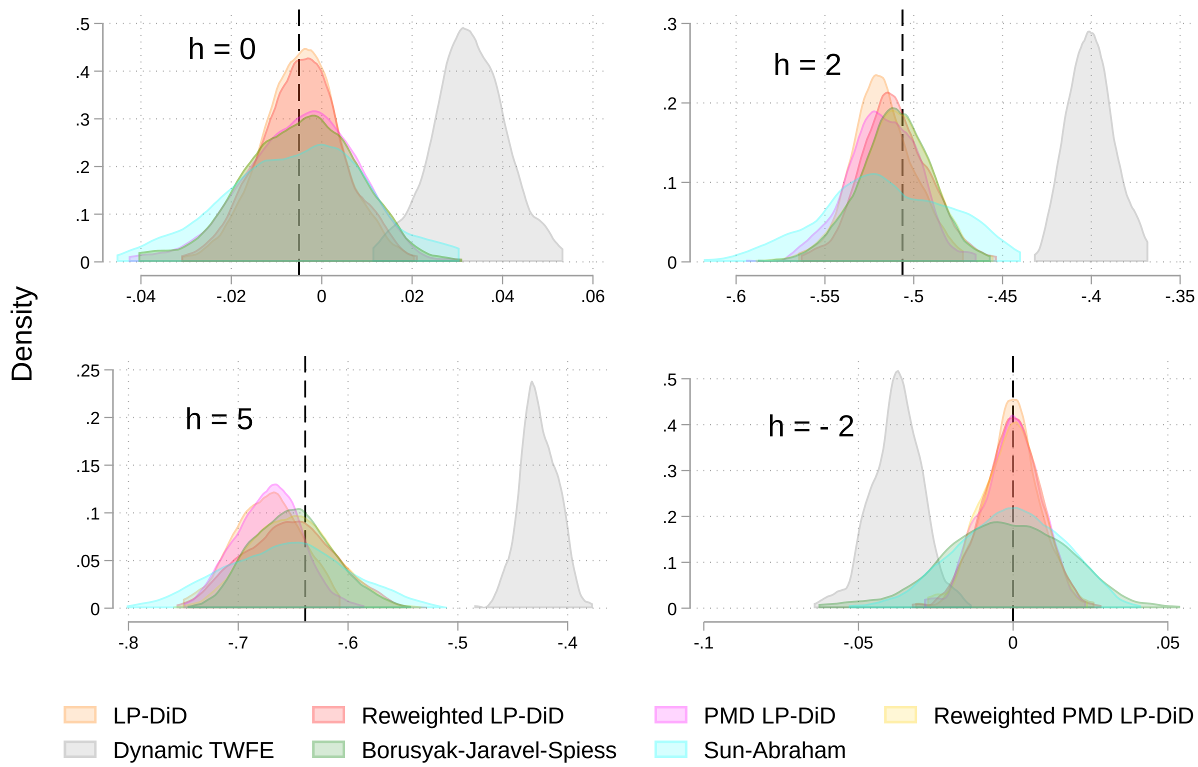
Results are presented in Figures 1 and 2 and Table 1. Figure 1 displays the estimated effect path in comparison with the true (equally-weighted) average effect path and the full range of cohort-specific effects. Figure 2 plots the full distribution of estimates at time-horizons $h = 0, 2, 5$ and -2.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

Figure 1: Simulation results: True effect path and estimates



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

Figure 2: Simulation results: Distribution of estimates



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

Conventional dynamic TWFE does a poor job in this setting, due to the heterogeneity of treatment effects. It finds a spurious, increasing, pre-trend and grossly underestimates the magnitude of the effect. Due to negative weighting, point estimates lie outside the full range of true cohort-specific effects.

The LP-DiD estimator, in contrast, tracks the true average dynamic effect well. This is true for both the simple (variance-weighted) LP-DiD and the re-weighted LP-DiD, and with both pre-mean differencing (PMD) and first-lag differencing. When comparing LP-DiD with the SA, CS and BJS estimators, performance is broadly similar.

The simulation also provides some insights into the possible differences between estimators in this setting. First, the results are suggestive of a bias-variance trade-off between equal-weighting and variance-weighting: reweighted LP-DiD (which applies equal weights) yields an unbiased estimate of the ATT, while the simple (variance-weighted) LP-DiD has a small bias but also lower variance. The bias from variance-weighting tends to become more relevant at longer time-horizons, because in our simulation the true

Table 1: Simulation results: Root mean squared error (RMSE), multipied by 100

Event time	Dynamic TWFE	LP-DiD	Rw LP-DiD	PMD LP-DiD	Rw PMD LP-DiD	SA	CS	BJS
-5	16.34	1.43	1.44	0.99	1.01	4.06	1.44	1.19
-4	12.38	1.31	1.37	0.98	1.02	3.53	1.37	1.52
-3	8.30	1.23	1.25	0.99	0.98	2.81	1.25	1.83
-2	3.85	0.94	0.97	0.96	0.96	1.71	0.97	2.04
0	3.84	0.89	0.94	1.23	1.27	1.56	0.94	1.27
1	7.17	1.36	1.40	1.64	1.67	2.60	1.40	1.65
2	10.67	2.08	1.99	2.32	2.20	3.71	1.99	2.16
3	14.16	2.96	2.65	3.04	2.69	4.63	2.65	2.61
4	17.94	3.59	3.20	3.60	3.11	5.24	3.20	2.99
5	21.33	4.61	4.32	4.44	3.96	5.80	4.32	3.76

Notes: RMSE from 200 replications. Dynamic TWFE = dynamic 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, 2024.

treatment effect variance increases in time after treatment. As a result, variance-weighting produces lower RMSE than equal weighting at short time-horizons ($h = 0$ and 1), but the opposite is true at longer time horizons (Table 1).

Moreover, due to high autocorrelation in untreated potential outcomes, estimators using pre-mean differencing (BJS and PMD LP-DiD) do not systematically outperform those using first-lag differencing, consistent with the theoretical results in Harmon (2022).

Given the variance-bias trade-off discussed above and the advantages of first-lag differencing under high autocorrelation, at shorter time-horizons ($h = 0$ and 1) variance-weighted LP-DiD produces the lowest RMSE across all estimators considered (Table 1).

5.4 Computational speed

To quantify computational speed, we recorded the computation time required for estimating the treatment effect path in one replication of our simulation. To assess how computation time changes with sample size, we also record computation times in another simulated dataset with the same DGP but double the number of units, time periods and treatment events. For these exercises, the estimations were conducted using STATA software on a laptop with Apple M2 Chip processor and 8 GB of RAM.

Table 2, which reports recorded computation times, shows that LP-DiD provides a significant computational advantage relative to other recently proposed estimators. The

Table 2: Computational speed (seconds)

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

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

computational advantage relative to CS and SA is substantial, especially in the second larger dataset, while the advantage relative to BJS is more limited.

There is already some early evidence, from recent empirical studies, for the empirical relevance of the computational advantage of LP-DiD. For example Gilbert, Hoen, and Gagarin (2024) employ LP-DiD in a large dataset of US workers to estimate the labor market impact of wind energy projects and state that the computational advantage of the method is "a crucial benefit in our case" (Gilbert, Hoen, and Gagarin, 2024, p. 278).

6 Empirical Applications

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

6.1 Credit and the labor share

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

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

The dataset covers the 1970–1996 period. (In 1997, inter-state banking deregulation was imposed in all states by federal law.) Online Appendix Figure E.3, which reproduces Figure 1 in LW, displays the share of US states with a liberalized banking sector.

Using static and dynamic TWFE specifications, LW find that inter-state banking deregulation had a sizable negative effect on the labor share, but find no effect of intrastate branching deregulation. Online Appendix E presents replication results using the same TWFE specifications employed in LW. Like the original LW study, these results suggest that the liberalization of interstate 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.

Given the staggered rollout of banking deregulation laws, the TWFE specifications employed in LW might by influenced by unclean comparisons: earlier liberalizers are implicitly used as controls for states that liberalize later on, potentially introducing negative weighting bias. Using the Goodman-Bacon (2021) decomposition diagnostic, we find that a conventional static TWFE estimator for the effect of inter-state banking deregulation assigns a 36% weight to unclean comparisons. For the estimates of the effect of intrastate branching deregulations, the problem is much more severe: unclean comparisons account for as much as 70% of the estimate (details in Online Appendix E). The greater potential 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.

We revisit the effect of banking deregulation using the following LP-DiD specification:

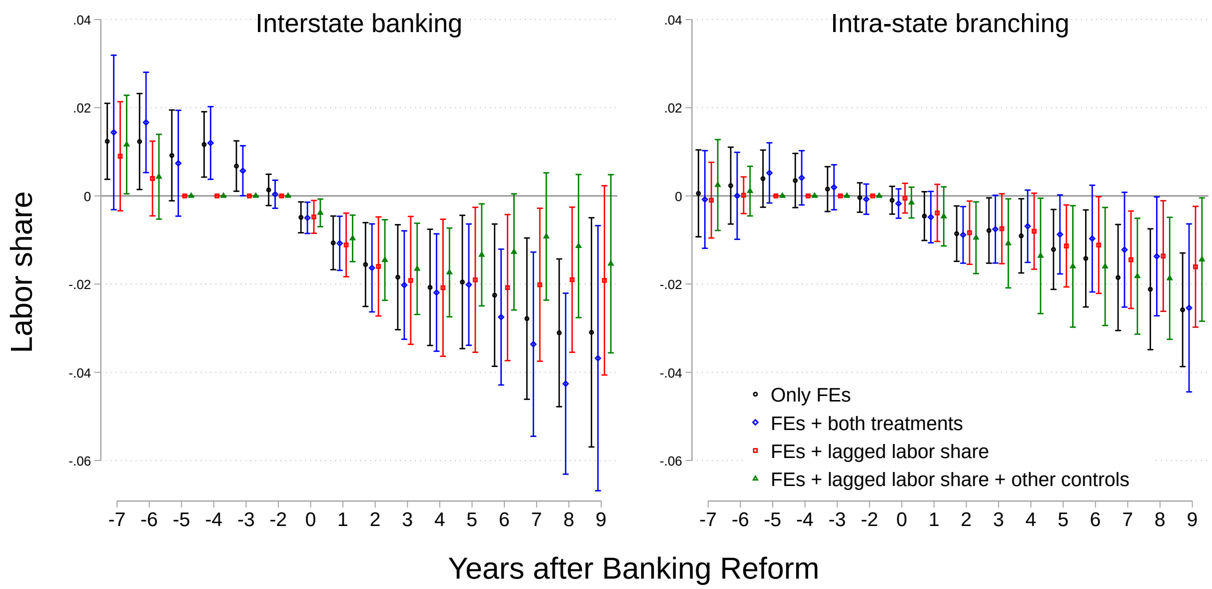
$$LS_{s,t+h} - LS_{s,t-1} = \delta_t^h + \beta_h^{LP-DiD} \Delta Policy_{s,t} + \sum_{m=1}^{M} \eta_m^h X_{s,t-m} + e_{s,t}^h \,, \tag{14}$$

where s indexes states, t indexes years, LS is the labor share, X is a vector of control variables, and $Policy_{st}$ is a binary indicator equal to one if a state has adopted the policy (intrastate branching or interstate banking deregulation). The estimation sample is restricted to observations that are either newly treated ($\Delta Policy_{s,t} = 1$) or clean controls ($Policy_{s,t+h} = 0$).

Results are displayed in Figure 3. The negative effect of *inter-state* banking deregulation is confirmed, including when controlling for pre-treatment outcome dynamics.

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

Figure 3: Effect of banking deregulation on the labor share: LP-DiD estimates



Notes: Estimates for the effect of banking deregulation on the labor share, using data from LW and the LP-DiD specification of equation 14. Following LW, the additional controls are four lags of real State GDP, average corporate tax rate, and union membership rates.

6.2 Democracy and economic growth

Our second empirical application estimates the effect of democratization on GDP per capita, replicating the analysis in Acemoglu et al. (2019), ANRR hereafter.

The ANRR dataset covers 175 countries from 1960 to 2010. The treatment indicator is a binary measure of democracy. The outcome of interest is the log of GDP per capita.

Three features make this application a relevant 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. 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.

First, to clarify the assumptions under which LP-DiD can provide unbiased estimates of the ATT in this setting, Online Appendix D presents a simulation calibrated on this empirical application, where the treatment status is endogenous and responds to past outcome shocks, treatment is non-absorbing, and treatment effects are heterogeneous. The simulation shows that our LP-DiD estimator with controls for lagged outcomes provides reliable estimates, with reasonably small RMSE and centered around the ATT when used with reweighting.[16]

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

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct} \,, \tag{15}$$

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. Estimated coefficients from Equation 15 are used to build an impulse response function (IRF) for the dynamic effect on GDP, as well as the cumulative long-run effect of a permanent transition to democracy, estimated as $\hat{\beta}(1 - \sum_{j=1}^{p} \hat{\gamma}_j)^{-1}$.

Online Appendix Figure F.6 displays the IRF from the estimation of the dynamic panel model of Equation 15. This reproduces the baseline results in ANRR. The implied long-run effect of democracy on growth is 21 percent with a standard error of 7 percent. This dynamic fixed effects specification, however, might suffer from the negative weighting bias discussed in Section 2.

We consider the following LP-DiD specification:

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP\ DiD} \Delta D_{ct} + \delta_t^h + \sum_{j=1}^{p} \gamma_j^h y_{c,t-j} + \epsilon_{ct}^h \,, \tag{16}$$

restricting the estimation sample to democratizations ($D_{it} = 1; D_{i,t-j} = 0$ for $1 \leq j \leq L$) and clean controls ($D_{i,t-j} = 0$ for $0 \leq j \leq L$). 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

[16] The DGP in the simulation in Online Appendix D is such that the parallel trends assumption does not hold unconditionally, but does hold conditional on the lagged outcome change. As already mentioned, some more general and complex dynamic selection settings can be incompatible with (conditional) parallel trends assumptions, and therefore not suitable for a DiD approach. The reader can refer to Ghanem, Sant'Anna, and Wüthrich, 2024 and Marx, Tamer, and Tang, 2024 for comprehensive discussions of the relation between dynamic selection and parallel trends, and to Viviano and Bradic, 2023 for estimators suitable for more general dynamic selection settings in which (conditional) parallel trends does not hold.

treatment status might introduce bias (as argued by ANRR, pp. 54–55). Moreover, similar to ANRR, only non-democracies are included in the control group, although in principle under Assumption 9 also countries that are continually democracies from $t-L$ to t could be included. This reflects the concern that established democracies might not be a good comparison for new democracies, therefore a control group composed only of continuing autocracies is more likely to satisfy the parallel trends assumption.

In a section of their analysis, ANRR employ a semiparametric LP specification that can be seen as a special version of the LP-DiD estimator. Specifically, they estimate Equation 16 with $L = 1$, which implies a time-window for defining clean controls of just 1 year.

Seeing ANRR's semiparametric specification as a version of LP-DiD provides a useful and novel perspective on their analysis and suggests possible deviations from their specification. Our formal analysis in Section 4 makes clear that their choice relies on an implicit (and unintended) assumption that treatment effects stabilize after 1 year, which is clearly too strong in this setting.[17]

We thus estimate the LP-DiD specification of Equation 16 with a time-window of 20 years ($L = 20$) for defining clean controls, thus excluding observations that have experienced some transition in the previous 20 years.

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

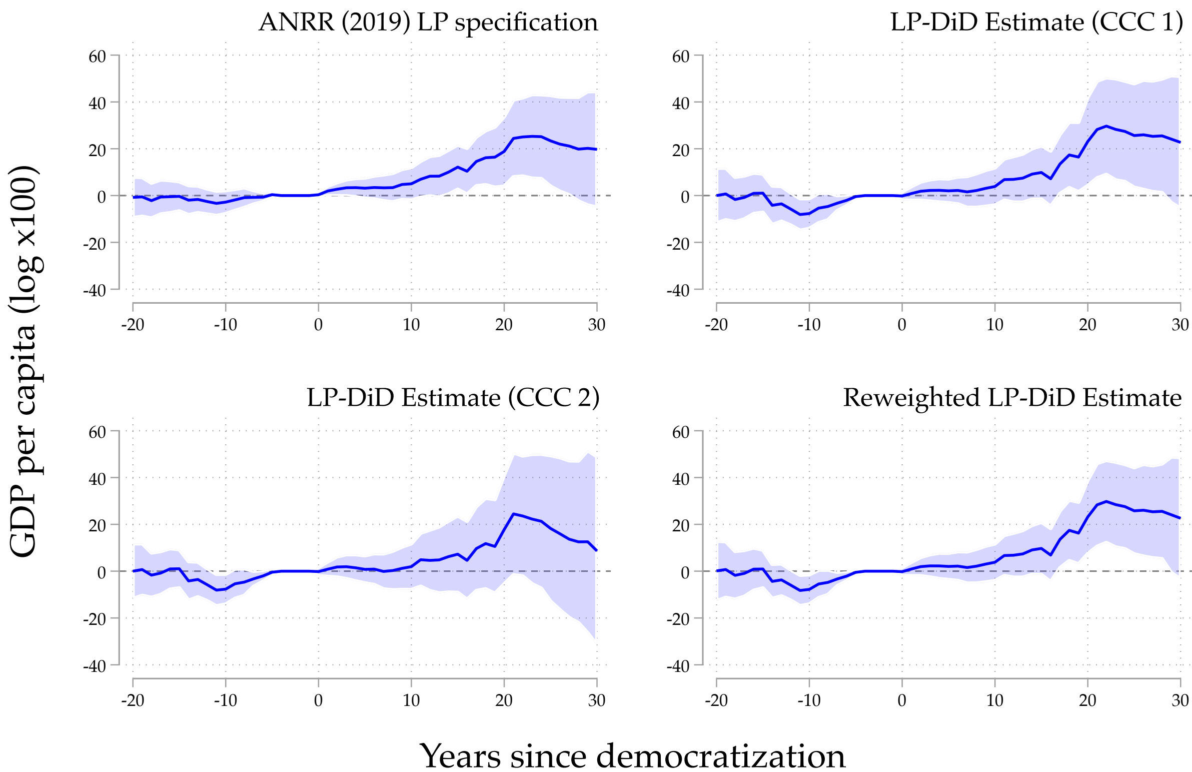
Figure 4 displays results from four LP-DiD specifications. The first (top left panel) follows ANRR and sets a time-window of just one year for defining clean controls ($L = 1$). The second (top right) uses a time-window of 20 years ($L = 20$). The third (bottom left) adds the additional requirement that control units remain non-democracies between $t+1$ and $t+h$. The fourth (bottom right) estimates the LP-DiD specification using regression adjustment (RA) to obtain an equally-weighted (rather than a variance-weighted) ATT, as discussed in Section 3.3.

Overall, the result of a positive and large effect of democracy on GDP per capita appears robust to stricter definitions of the control group. The partial exception is the specification that excludes countries that democratize between $t+1$ and $t+h$ from

[17]For example, Argentina democratized in 1973 and became a dictatorship again in 1976. In the ANRR approach, Argentina contributes to the counterfactual for estimating the effect of Spain's 1978 democratization. It seems natural to consider an alternative specification that excludes Argentina from the counterfactual for countries that (like Spain) democratize shortly after 1973–76, reflecting the concern that the country might have experienced prolonged dynamic effects from its 1973–76 transitions.

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 for the reasons discussed previously.

Figure 4: Effect of democracy on growth: LP-DiD estimates



Notes: LP-DiD estimates for the effect of democracy on GDP per capita, using the specification of Equation 16. Top left panel ('ANRR (2019) LP specification') replicates results in Section IV of ANRR, which set $L = 1$ year in the clean control condition. The other three panels set $L = 20$ years. CCC 1 is a clean control condition that defines treated units as countries that democratize in year t and experienced no transition between $t-20$ and $t-1$, and clean controls as countries that are continually non-democracies between $t-20$ and t. CCC 2 defines treated units in the same way, but clean controls are continually non-democracies between $t-20$ and $t+h$. Right bottom panel uses reweighting to obtain an equally-weighted effect, using CCC 1. See main text for more details.

7 Conclusion

We propose a simple-to-implement, transparent, and computationally fast technique for difference-in-differences estimation with dynamic heterogeneous treatment effects. Our proposed LP-DiD estimator, while based on a simple OLS regression, does not suffer from the negative weighting problem highlighted by recent studies. LP-DiD provides an encompassing framework which can be flexibly adapted to implement different weighting and normalization schemes and target estimands, and can be extended to settings with non-absorbing treatment and covariates. The LP-DiD framework could be further extended in future work to a wider variety of settings, including, for example, continuous treatment variables.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2019). "Democracy Does Cause Growth". In: *Journal of Political Economy* 127.1, pp. 47–100.

Alloza, Mario, Jesus Gonzalo, and Carlos Sanz (2019). *Dynamic effects of persistent shocks*. Banco de España Working Paper 1944.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang (2022). "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2, pp. 370–395. DOI: https://doi.org/10.1016/j.jfineco.2022.01.004.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard (2021). "Panel experiments and dynamic causal effects: A finite population perspective". In: *Quantitative Economics* 12.4, pp. 1171–1196. DOI: https://doi.org/10.3982/QE1744.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (Feb. 2024). "Revisiting Event-Study Designs: Robust and Efficient Estimation". In: *The Review of Economic Studies* 91.6, pp. 3253–3285. DOI: 10.1093/restud/rdae007.

Callaway, Brantly and Pedro H. C. Sant'Anna (2020). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs". In: *Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Chaisemartin, Clément de and Xavier D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110.9, pp. 2964–96.

— (Feb. 2024). "Difference-in-Differences Estimators of Intertemporal Treatment Effects". In: *The Review of Economics and Statistics*, pp. 1–45. DOI: 10.1162/rest_a_01414.

Chaisemartin, Clément de, Xavier D'Haultfœuille, Félix Pasquier, and Gonzalo Vazquez-Bare (Jan. 2022). "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period". https://arxiv.org/abs/2201.06898.

Chaisemartin, Clément de and Xavier D'Haultfœuille (June 2022). "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey". In: *The Econometrics Journal*. DOI: 10.1093/ectj/utac017.

Dube, Arindrajit and Attila Lindner (2024). "Minimum Wages in the 21st Century". In: *Handbook of Labor Economics, volume 5-6*. Elsevier.

Gardner, John, Neil Thakral, Linh T Tô, and Luther Yap (May 2024). "Two-stage differences in differences". https://jrgcmu.github.io/2sdd_gtty.pdf.

Ghanem, Dalia, Pedro H. C. Sant'Anna, and Kaspar Wüthrich (2024). *Selection and parallel trends*. arXiv: 2203.09001 [econ.EM]. URL: https://arxiv.org/abs/2203.09001.

Gilbert, Ben, Ben Hoen, and Hannah Gagarin (2024). "Distributional equity in the employment and wage impacts of energy transitions". In: *Journal of the Association of Environmental and Resource Economists* 11.S1, S261–S298.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Harmon, Nikolaj A (2022). *Difference-in-Differences and Efficient Estimation of Treatment Effects*. Tech. rep. Working paper.

Jordà, Òscar (2005). "Estimation and Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95.1, pp. 161–182.

Leblebicioğlu, Aslı and Ariel Weinberger (2020). "Credit and the Labour Share: Evidence from US States". In: *Economic Journal* 130.630, pp. 1782–1816.

Marcus, Michelle and Pedro H. C. Sant'Anna (2021). "The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics". In: *Journal of the Association of Environmental and Resource Economists* 8.2, pp. 235–275.

Marx, Philip, Elie Tamer, and Xun Tang (2024). "Parallel Trends and Dynamic Choices". In: *Journal of Political Economy Microeconomics* 2.1, pp. 129–171. DOI: 10.1086/727363.

Robins, James (1986). "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". In: *Mathematical Modelling* 7.9, pp. 1393–1512. DOI: https://doi.org/10.1016/0270-0255(86)90088-6.

Roth, Jonathan (2024). *Interpreting Event-Studies from Recent Difference-in-Differences Methods*. arXiv: 2401.12309 [econ.EM].

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature". In: *Journal of Econometrics*. DOI: https://doi.org/10.1016/j.jeconom.2023.03.008.

Rubin, Donald B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of Educational Psychology* 66.5, pp. 688–701.

Sun, Liyang and Sarah Abraham (2020). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

Viviano, Davide and Jelena Bradic (2023). *Dynamic covariate balancing: estimating treatment effects over time with potential local projections*. arXiv: 2103.01280.

Wooldridge, Jeffrey M. (Aug. 2021). "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators". http://dx.doi.org/10.2139/ssrn.3906345.

Online Appendix

A Derivation and further discussion of the results in Sections 2.2 and 2.3

A.1 Two groups and two time periods

The link between LP and DiD is easiest to see in a basic 2-groups/2-periods setting. In the 2x2 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.

Assume two groups of units and two time periods. In the first period (pre-treatment) no unit is treated. In the second period (post-treatment) one group of units is treated while the other remains untreated. In terms of the general setup and notation introduced in the main text, we are setting $T = 2$, and therefore $t \in \{1, 2\}$. Moreover, we have $g \in \{0, 1\}$, where group 0 is the control group and group 1 the treatment group. For units in the treatment group $p_i = p_1 = 2$. For units in the control group $p_i = p_0 = \infty$.

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

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

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

Now consider an LP regression (Equation 4 in the main text) 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 a static TWFE regression (Equation 2 in the main text) 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 $\beta_0^{LP} = \beta^{STWFE} = \beta^{2x2} = ATT$.

A.2 Two groups and multiple time periods

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

Specifically, assume that all units in the treatment group enter treatment at time s, with $1 < s < T$, and remain treated thereafter, while control units are never treated over the sample period. In terms of our general setup and notation, we are setting $g \in \{0, 1\}$, where group 0 is the control group and group 1 the treatment group. For all units in the treatment group, $p_i = p_1 = s$. For all units in the control group, $p_i = p_0 = \infty$. With only one treated cohort, the group-specific dynamic ATT 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 in main text),

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

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

The population coefficient β_h^{LP} from a simple LP regression (as defined in Equation 4 in the main text) corresponds exactly to this estimand. To see this, note that in this setting the LP regression 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 Chaisemartin and D'Haultfœuille 2020; Gardner et al. 2024; Sun and Abraham 2020; Goodman-Bacon 2021), we know that in this setting, with only one treated cohort, and under no anticipation and parallel trends assumptions, the coefficients in the dynamic TWFE regression (Equation 3 in main text) correspond to the τ_h estimands.[18] Moreover, the β^{STWFE} estimand from the static TWFE regression (Equation 2 in main text) equals the ATT, defined as $E[\tau_{it} | D_{it} = 1]$.[19]

A.3 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 a LP regression similar to Equation 4 in main text but 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 in the main text and assuming that treatment effects are homogeneous, mean observed outcomes at time $t + h$

[18] This can be seen using the decomposition of the dynamic 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 $\beta_h^{ETWFE} = \tau_h$.

[19] One way to see this is to use the decomposition of the static TWFE into a weighted average of treatment-cohort specific ATTs (Chaisemartin and D'Haultfœuille 2020, p. 2970; Gardner et al. 2024, 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.

are given by

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

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,[20]

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

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

$$
y_{i,t+h} - y_{i,t-1} = \delta_t^h + \beta_h^{LP} \, \Delta D_{it} + \sum_{\substack{j=-h \\ j\neq 0}}^{\infty} \theta_j^h \, \Delta D_{i,t-j} + e_{it}^h \, . \tag{A.2}
$$

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 static and dynamic TWFE specifications (Equations 2 and 3 in the main text) identify in this setting with staggered treatment and dynamic but homogeneous effects?

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

Intuitively, the bias comes from the fact that previously treated units are effectively used as controls for newly treated units. Since previously treated units might still be experiencing a delayed dynamic response to treatment, these treatment effect dynamics are effectively subtracted from the static TWFE treatment effect estimate (Goodman-Bacon, 2021). That is, delayed dynamic responses to treatment can enter the static TWFE estimate (Equation 2 in main text) with a negative weight (Chaisemartin and D'Haultfœuille, 2020).

Under the assumption of homogeneous treatment effects, however, dynamic TWFE regression does not suffer from this bias and, like the LP regression with lags and leads

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

of treatment discussed above, unbiasedly estimates the average treatment effect path under parallel trends and no anticipation, as long as a sufficient number of lags of the treatment indicator is included (see Proposition 4 and Eq. 19 in Sun and Abraham, 2020). Intuitively, the lagged treatment indicators control for lagged dynamic effects of previous treatments, which in this setting are the same in expectation for all units.

The difference between Equation 3 in the main text and Equation A.2 above arises from how they handle unit fixed effects. The former (TWFE) removes unit fixed effects with mean differencing using the full sample, while the latter (LP) does so by differencing around the treatment times. The latter could be advantageous in reducing bias if the unit effects are not fully fixed (i.e., there are violations in the parallel trends assumption). For an example of this, see Cengiz et al. (2019).

B Weights of the LP-DiD estimator

This appendix derives the weights assigned to each cohort-specific ATT by the LP-DiD estimator, first in a baseline OLS version without control variables and then in OLS specifications with control variables.

B.1 Baseline OLS specification 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 in main text); potential outcomes without treatment are determined according to the fixed-effects model in Equation 1 in the main text. As in Section 2.4, treatment effects are unrestricted, and 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,

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

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:

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

restricting the sample to observations that are either

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

$\hat{\beta}_h^{LP-DiD}$ is the LP-DiD estimate of the dynamic weighted ATT, 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 in the main text. 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 B.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 $p_g + h$. Formally,

$$CCS_{g,h} = \left\{ i, t \mid \left[\Delta D_{i,p_g} = 1 \vee D_{i,p_g+h} = 0 \right] \wedge t = p_g \right\}.$$

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 B.3 enters into one and only one CCS. Therefore, the unbalanced panel dataset defined by the clean control condition in B.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 B.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,

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

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

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

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

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 B.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 Assumptions 1 and 2 in main text, we have

$$\begin{aligned}
E\left(\hat{\beta}_h^{LP-DiD}\right) &= \frac{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\left[\Delta\widetilde{D}_{i,p_j} E\left(y_{i,p_j+h} - y_{i,p_j-1}\right)\right]}{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\Delta\widetilde{D}^2_{i,p_j}} \\
&= \frac{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\left[\Delta\widetilde{D}_{i,p_j} E\left(\tau_{i,p_j+h} D_{i,p_j+h}\right)\right]}{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\Delta\widetilde{D}^2_{i,p_j}} \\
&= \frac{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\left[\Delta\widetilde{D}_{i,p_j} E\left(\tau_{i,p_j+h} D_{i,p_j}\right)\right]}{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\Delta\widetilde{D}^2_{i,p_j}} \\
&= \sum_{j=1}^{G}\sum_{i\in CCS_{j,h}} \frac{\Delta\widetilde{D}_{i,p_j}}{\sum_{j=1}^{G}\sum_{i\in CCS_{j,h}}\Delta\widetilde{D}^2_{i,p_j}} E\left(\tau_{i,p_j+h} D_{i,p_j}\right) \\
&= \sum_{j=1}^{G}\sum_{i\in j} \frac{\Delta\widetilde{D}_{i,p_j}}{\sum_{j=1}^{G}\sum_{i\in j}\Delta\widetilde{D}^2_{i,p_j}} \tau_{i,p_j+h} \\
&= \sum_{g\neq 0} \frac{N_g \Delta\widetilde{D}_{g,p_g}}{\sum_{g\neq 0} N_g \Delta\widetilde{D}^2_{g,p_g}} \tau_{g,p_g+h} \\
&= \sum_{g\neq 0} \omega_{g,h}^{LP-DiD} \tau_h^g,
\end{aligned}$$

where the weights are given by

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

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_h^g was defined in the main text as the dynamic ATET for group g at time-horizon h.

Note that by definition of $CCS_{g,h}$, we have $n_{c,g,h} = 1 - n_{g,h}$. Therefore the weight can be rewritten as $\frac{N_{CCS_{g,h}}[n_{gh}(1-n_{gh})]}{\sum_{g\neq 0} N_{CCS_{g,h}}[n_{g,h}(1-n_{g,h})]}$, as we do in the main text.

B.2 Weights with control variables

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

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

B.2.1 Covariates with linear and homogeneous effects

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

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

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

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

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

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

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

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

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

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

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

The equivalence of Equation B.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

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

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 B.6 above.

B.2.2 More general setting

Now consider a more general setting, in which Assumptions 4 and 5 in the main text 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

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

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 B.3.

The weights are thus given by

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

As noted in the main text (Section 4.1.1), it is always possible to avoid negative weights using a regression adjustment (RA) specification of LP-DiD.

C Derivation and further discussion of the results in Section 4.2.3 in the main text

This Appendix provides further formal discussion and derivations regarding the LP-DiD estimator for the average effect of a treatment event under an effect stabilization assumption in a setting with non-absorbing treatment, presented in the main text in Section 4.2.3.

In addition to the treatment events defined in main text, let us now define exit events. Group g experiences an exit event at time j if and only if $\Delta D_{g,j} = -1$.

We let q_g^n denote the time period in which group g exits treatment for the n-th time in the estimation sample, with $q_g^n = \infty$ if group g exits treatment less than n times. We let N_g^E denote the number of exit events experienced by group g in the estimation sample.

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

$$\eta_h^{g,n} = E\left[y_{i,q_g^n+h}(\mathbf{D}_{i,q_g^n,1}) - y_{i,q_g^n+h}(\mathbf{D}_{i,q_g^n,0}) \mid i \in g, q_g^n \neq \infty\right]$$

Assumption 9 in main text applies also to exit events: namely, $\eta_L^{g,n} = \eta_{L+l}^{g,n}$ for some $L < T$, for all $l \geq 0$, and for all exit events {g,n} .

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

Therefore, in this setting and under assumptions 7, 8 and 9 in main text, in an estimation sample equal to the whole NxT sample we would have

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

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

Equation C.1 differs from its absorbing treatment counterpart (Equation 6 in main text) in two main ways: first, the same group can be under the influence of multiple previous treatment and exit events; second (because of Assumption 9), only events that occur between periods $t-L$ and $t+h$ affect the expected value of the long difference $y_{t+h} - y_{t-1}$.

Equation C.1 thus motivates and justifies the modified clean control condition of equation 13 in Section 4.2.3.

C.1 Derivation of the weight $\omega_{g,n,h}^{LP-DiD''}$ assigned to each treatment event

In what follows we derive the weights that this estimator assigns to treatment events. As in the main text, $\omega_{g,n,h}^{LP-DiD''}$ will denote the weight assigned to $\tau_h^{g,n}$, the h–horizon effect of the n-th treatment event experienced by group g.

In addition to the definitions and notation introduced in the main text, denote the set of time periods in which there is some treatment event as t^τ. This is the set of time periods such that $\Delta D_{i,t} = 1$ for at least some unit i. Formally,

$$t^\tau = \{t \mid \Delta D_{it} = 1 \text{ for some } i\}.$$

Moreover, we denote the h-horizon effect of a treatment event experienced by unit i at time t as $\tau_h^{i,t}$. Using the notation introduced in Section 4.2 in the main text, this is defined as follows

$$\tau_h^{i,t} = E\left[y_{i,t+h}(\mathbf{D}_{i,t,1}) - y_{i,t+h}(\mathbf{D}_{i,t,0}) \mid \Delta D_{i,t} = 1\right]$$

For each time period $t \in t^\tau$, define the clean control sample $CCS_{t,h}$ as the set of observations for time t that satisfy the modified clean control condition in Equation 13 in the main text at time horizon h. Formally,

$$CCS_{t,h} = \left\{i \mid \left[(D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \wedge (D_{i,t-j} = 0 \text{ for } 1 \leq j \leq L)\right] \vee \left[\Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq L\right]\right\}.$$

Moreover, it is also convenient to define $CCS_{g,n,h} = CCS_{p_g^n,h}$. In words, $CCS_{g,n,h}$ is the set of units that satisfy the clean control condition of equation 13 at time $t = p_g^n$. Note that $CCS_{g,n,h}$ might include not only clean control units and units of group g, but also units of other groups that enter treatment at time p_g^n.

We let $\hat{n}_{g,n,h}$ denote the share of newly treated units in $CCS_{p_g^n,h}$. Formally,

$$\hat{n}_{g,n,h} = \frac{\sum_{i \in CCS_{p_g^n,h}} \Delta D_{i,p_g^n}}{N_{CCS_{p_g^n,h}}}$$

We let $\bar{n}_{g,n,h}$ denote the share of units belonging to group g among all newly treated units in $CCS_{p_g^n,h}$. Formally,

$$\bar{n}_{g,n,h} = \frac{\sum_{i \in CCS_{p_g^n,h}} \Delta D_{i,p_g^n} \times \mathbb{1}\{i \in g\}}{\sum_{i \in CCS_{p_g^n,h}} \Delta D_{i,p_g^n}}$$

Finally, let $\hat{\beta}^{LP-DiD''}$ denote the estimated coefficient from a LP regression (equation 4) in the 'clean control' estimation sample defined by the condition in Equation 13 in main text

By the Frisch-Waugh-Lovell theorem,

$$E\left(\hat{\beta}_h^{LP-DiD''}\right) = \frac{\sum_{t \in t^\tau} \sum_{i \in CCS_{t,h}} \left[\Delta\widetilde{D}_{i,t} E\left(y_{i,t+h} - y_{i,t-1}\right)\right]}{\sum_{t \in t^\tau} \sum_{i \in CCS_{t,h}} \Delta\widetilde{D}_{i,t}^2},$$

where $\Delta\widetilde{D}_{i,t}$ is the residual from a regression of ΔD on time indicators in the estimation sample defined by Equation 13 in main text.

Given assumptions 7, 8 and 9 in the main text, we have

$$
\begin{aligned}
E\left(\hat{\beta}_h^{LP-DiD''}\right) &= \frac{\sum_{t\in t^\tau}\sum_{i\in CCS_{t,h}}\left[\Delta\widetilde{D}_{i,t}E(y_{i,t+h}-y_{i,t-1})\right]}{\sum_{t\in t^\tau}\sum_{i\in CCS_{t,h}}\Delta\widetilde{D}_{i,t}^2} \\
&= \frac{\sum_{t\in t^\tau}\sum_{i\in CCS_{t,h}}\left[\Delta\widetilde{D}_{i,t}E\left(\tau_h^{i,t}\Delta D_{i,t}\right)\right]}{\sum_{t\in t^\tau}\sum_{i\in CCS_{t,h}}\Delta\widetilde{D}_{i,t}^2} \\
&= \frac{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\sum_{i\in CCS_{p_g^n,h}}\left(\Delta\widetilde{D}_{i,p_g^n}E\left(\tau_h^{i,p_g^n}\Delta D_{i,p_g^n}\right)\right)}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\sum_{i\in CCS_{p_g^n,h}}\Delta\widetilde{D}_{i,p_g^n}^2} \\
&= \sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\sum_{i\in g}\frac{\Delta\widetilde{D}_{i,p_g^n}}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\sum_{i\in g}\Delta\widetilde{D}_{i,p_g^n}^2}\tau_h^{i,p_g^n} \\
&= \sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\frac{N_g\Delta\widetilde{D}_{g,p_g^n}}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}N_g\Delta\widetilde{D}_{g,p_g^n}^2}\tau_h^{g,n} \\
&= \sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\omega_{g,n,h}^{LP-DiD''}\tau_h^{g,n},
\end{aligned}
$$

where the weights are given by

$$
\begin{aligned}
\omega_{g,n,h}^{LP-DiD''} &= \frac{N_g\Delta\widetilde{D}_{g,p_g^n}}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}N_g\Delta\widetilde{D}_{g,p_g^n}^2} \\
&= \frac{N_g\left(1-\frac{\sum_{i\in CCS_{gnh}}\Delta D_{i,p_g^n}}{N_{CCS_{g,n,h}}}\right)}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}N_g\left(1-\frac{\sum_{i\in CCS_{gnh}}\Delta D_{i,p_g^n}}{N_{CCS_{g,n,h}}}\right)} \\
&= \frac{\bar{n}_{g,n,h}N_{CCS_{g,n,h}}[\hat{n}_{g,n,h}(1-\hat{n}_{g,n,h})]}{\sum_{g=1}^{G}\sum_{n=1}^{N_g^T}\bar{n}_{g,n,h}N_{CCS_{g,n,h}}[\hat{n}_{g,n,h}(1-\hat{n}_{g,n,h})]}.
\end{aligned}
$$

D Additional simulation (dynamic selection into treatment)

In this appendix we test the performance of the LP-DiD estimator in a Monte Carlo simulation which features a simple form of dynamic selection and non-absorbing treatment. Specifically, we assume that the probability of receiving treatment, and of experiencing a treatment reversal, depend on outcome dynamics. This setting matches the empirical application presented in Section 6.2 in main text. In this simulation, the parallel trends assumption does not hold unconditionally but holds conditional on the lagged outcome change.

This simulation thus provides a simple example of a possible DGP featuring dynamic selection and non-absorbing treatment that is consistent with conditional parallel trends and therefore can be addressed within a LP-DiD framework by properly conditioning on lagged outcome changes. Of course, as well documented in the literature, more general and complex dynamic selection settings can be incompatible with (conditional) parallel trends (Ghanem, Sant'Anna, and Wüthrich, 2024; Marx, Tamer, and Tang, 2024; Viviano and Bradic, 2023). In those settings, a Difference-in-Differences approach is not suitable, and researchers should use methods based on sequential ignorability assumptions (rather than parallel trends), like those proposed in Viviano and Bradic (2023).

All results presented in this Appendix are virtually unchanged if ruling out treatment reversals.

D.1 Setting

We calibrate our simulated datasets on the empirical application presented in Section 6.2 in the main text, which estimates the effect of democratization on economic growth, replicating the analysis in Acemoglu et al. (2019).

Our simulated datasets include 184 countries observed over 51 periods (years). Potential outcomes (log of GDP per capita) without treatment are given by $y_{it}(0) = \lambda_i + \gamma_t + \epsilon_{it} + \rho\epsilon_{i,t-1}$, where ϵ is a white noise random shock. We set $\rho = 0.98$ based on the corresponding estimated autoregressive coefficient in the Acemoglu et al. (2019) dataset. λ_i, γ_t and ϵ_{it} are normally distributed with mean and standard deviation estimated empirically using the Acemoglu et al. (2019) dataset.

Unit i enters treatment in the first period that satisfies that following condition:

$$\psi\Delta y_{i,t-1} + (1-\psi)u_i \leq \theta \quad \text{and} \quad 11 \leq t \leq 40, \tag{D.1}$$

where u is a white noise term with the same variance as ϵ. We set $\psi = 0.5$ 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, similar to the pre-democratization dip in GDP per capita documented in Acemoglu et al. (2019).

We allow for treatment reversals. Specifically, a unit exits treatment in the first post-treatment period that satisfies condition D.1, provided that at least 5 periods have passed since entering treatment. For simplicity, we do not allow for re-entry after a reversal.[21]

The treatment effect is positive and grows in time, stabilizing 9 years after treatment. Moreover, earlier treated cohorts experience larger treatment effects. Specificaly, define $\phi = t - p_i + 1$ and $\mu_i = \frac{p_i - 1960}{\min\{p_1,...,p_N\} - 1960}$. The treatment effect β_{it} is given by

$$\beta_{it} = \begin{cases} 0 & \text{if } t - p_i < 0 \\ \alpha\phi^2 + (1-\alpha)\left(\frac{\phi}{\mu}\right)^2 & \text{if } 0 \leq t - p_i \leq 8 \\ \beta_{i,t-1} & \text{if } t - p_i > 8\,, \end{cases}$$

where p_i is the period in which unit i enters treatment as in the previous sections. We set $\alpha = 0.5$. After a treatment reversal, β_{it} gradually falls, until it stabilizes at 0 after 9 periods.

Observed outcomes y_{it} are given by

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

D.2 Estimators

We estimate effects using (a) simple (variance-weighted) LP-DiD; (b) Re-weighted LP-DiD; (c) PMD LP-DiD and (d) Re-weighted PMD LP-DiD. Given that treatment is non-absorbing, we employ the estimator of the average effect of a treatment event developed in Section 4.3.2 in the main text, under the assumption that treatment effects stabilize after 9 periods. The PMD versions set $k = 9$ given the non-absorbing treatment setting and the assumption that effects stabilize after 9 periods. We include $\Delta y_{i,t-1}$ as a control variable in all specifications.

D.3 Results

Results are presented in Figures D.1 and D.2 and Table D.1. Figure D.1 displays the estimated effect path in comparison with the true (equally-weighted) average effect path

[21] Multiple treatment and exit events in the same country are rare in the Acemoglu et al., 2019 dataset.

and the full range of heterogeneous group-specific effects. Figure D.2 plots the full distribution of the estimates at time horizons $h = 0, 2, 5$ and -3. Table D.1 reports the root mean squared error (RMSE) at time horizons $h = -10$ to $+10$.

Overall, the LP-DiD estimator tracks well the true dynamic effect also in this second simulation. This demonstrates that LP-DiD specifications controlling for lagged outcome dynamics can address at least some simple types of dynamic selection into treatment. Although theoretically re-weighted versions of LP-DiD are preferable when there are control variables (as discussed in Section 4.1), in this simulation there are no material differences in performance between variance-weighted and equally-weighted specifications, notwithstanding significant heterogeneity in treatment effects across cohorts (Table D.1).

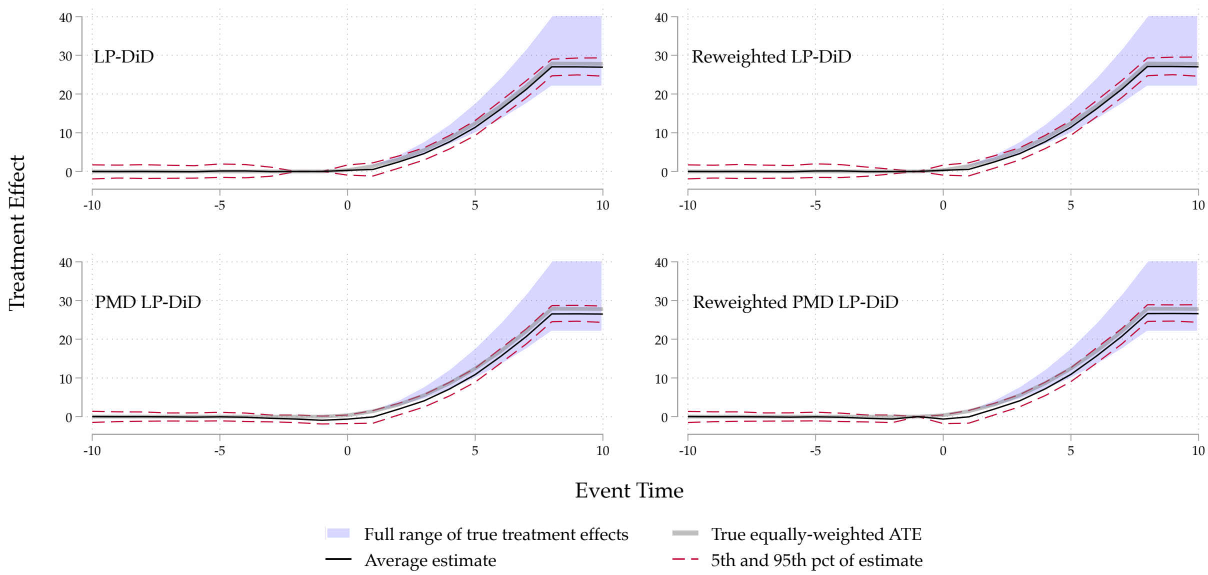
Note that this specification, including the lagged outcome variable, might suffer from Nickell (1981) bias, due to the presence of $y_{i,t-1}$ both in a regressor and in the error term. However, two simultaneous conditions must be met for this bias to be problematic. First, the autoregressive coefficient 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 this simulation the autoregressive coefficient is high, but the time dimension of the dataset is long, which is why Nickell bias is not a major concern here. In applications in which Nickell bias is a major concern, the researcher can correct for it by using a simple split-sample correction, following Chen, Chernozhukov, and Fernández-Val (2019) and Mei, Sheng, and Shi (2023).

Table D.1: Simulation with endogenous treatment timing: Root Mean Squared Error (RMSE)

Event time	LP-DiD	Rw LP-DiD	PMD LP-DiD	Rw PMD LP-DiD
-10	1.05	1.04	0.85	0.85
-9	1.06	1.06	0.75	0.75
-8	1.10	1.10	0.71	0.71
-7	1.03	1.04	0.64	0.65
-6	1.00	1.00	0.68	0.68
-5	1.01	1.01	0.66	0.66
-4	1.01	1.00	0.67	0.67
-3	0.73	0.73	0.70	0.71
-2			0.88	0.88
0	0.81	0.81	1.22	1.21
1	1.37	1.37	1.75	1.74
2	1.20	1.20	1.48	1.47
3	1.35	1.35	1.73	1.72
4	1.41	1.40	1.80	1.77
5	1.54	1.51	1.89	1.85
6	1.39	1.36	1.70	1.66
7	1.45	1.41	1.73	1.67
8	1.41	1.36	1.64	1.57
9	1.43	1.38	1.64	1.57
10	1.49	1.44	1.72	1.65

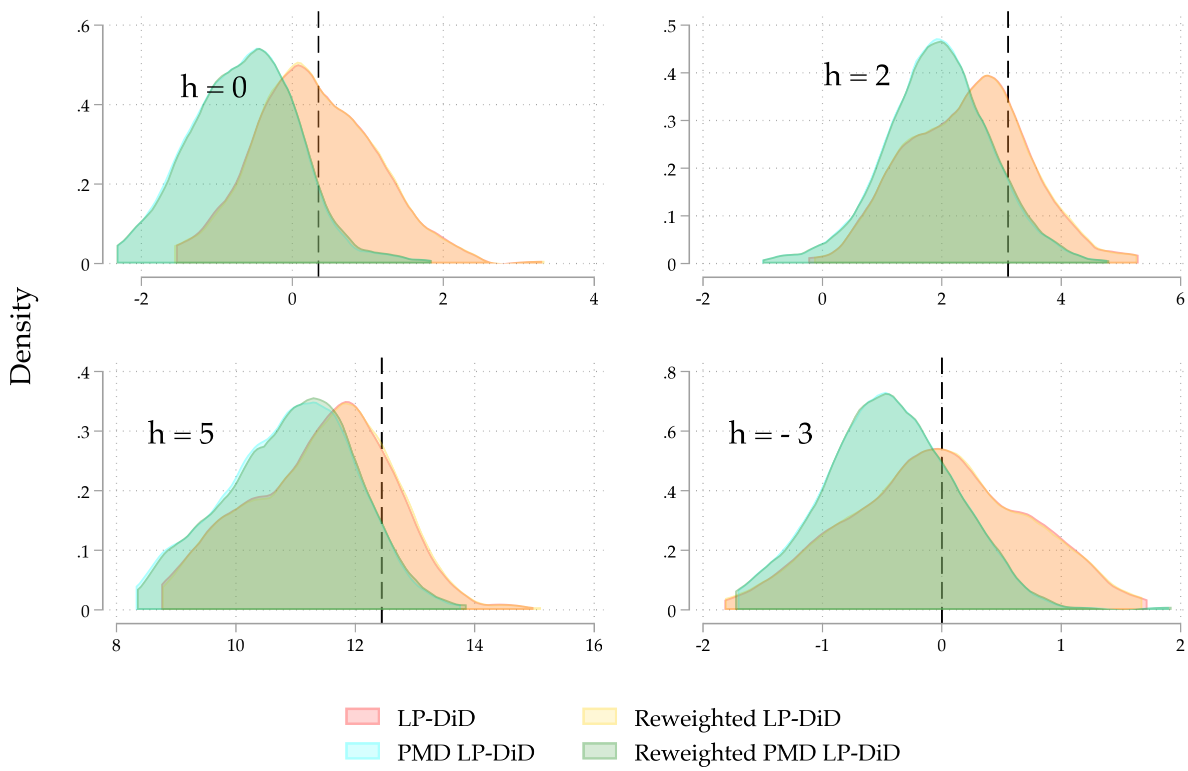
Notes: RMSE from 200 replications. Rw = reweighted; PMD = pre-mean-differenced.

Figure D.1: Simulation with endogenous treatment timing: true effect path and estimates



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

Figure D.2: Simulation with endogenous treatment timing: Distribution of estimates



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

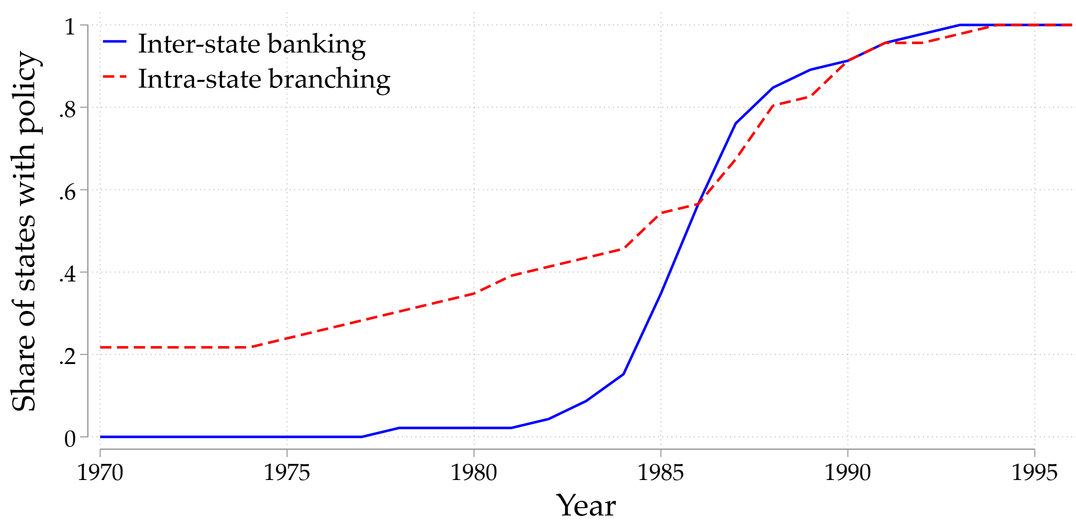
E Credit and the labor share: Additional statistics and results

This Appendix presents additional descriptive statistics and results concerning our first empirical application (Section 6.1 in the main text), which estimates the effect of banking deregulation on the labor share, replicating Leblebicioğlu and Weinberger (2020), LW hereafter.

E.1 Staggered rollout of banking deregulation reforms

Figure E.3 displays the share of US States which have adopted interstate banking and intrastate branching deregulation (similar to Figure 1 in LW).

Figure E.3: Banking deregulation in US States



Notes: Data from Leblebicioğlu and Weinberger (2020).

E.2 Results from conventional TWFE specifications

We estimate the following static TWFE specification for the effect of banking deregulation laws, which replicates LW's baseline specification:

$$LS_{st} = \beta_{Bank} Bank_{st} + \beta_{Branch} Branch_{st} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st} \,, \tag{E.1}$$

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

To assess possible pre-trends and lagged effects, LW also estimate the following dynamic TWFE specification:[22]

$$LS_{st} = \sum_{q=-9}^{9} \beta_{Bank,t+q} \Delta Bank_{s,t+q} + \sum_{q=-9}^{9} \beta_{Branch,t+q} \Delta Branch_{s,t+q} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st} . \quad \text{(E.2)}$$

Figure E.4 displays results from the static and dynamic TWFE specifications of Equations E.1 and E.2. These replicate the estimates reported in Table 2 and Figure 2 of LW. They suggest that the liberalization of inter-state banking has a sizable negative effect on the labor share, although they also show a small pre-treatment trend. Instead, the estimated effects of intra-state branching deregulation on the labor share are positive, small and very imprecise.

E.3 Goodman-Bacon (2021) decomposition diagnostic

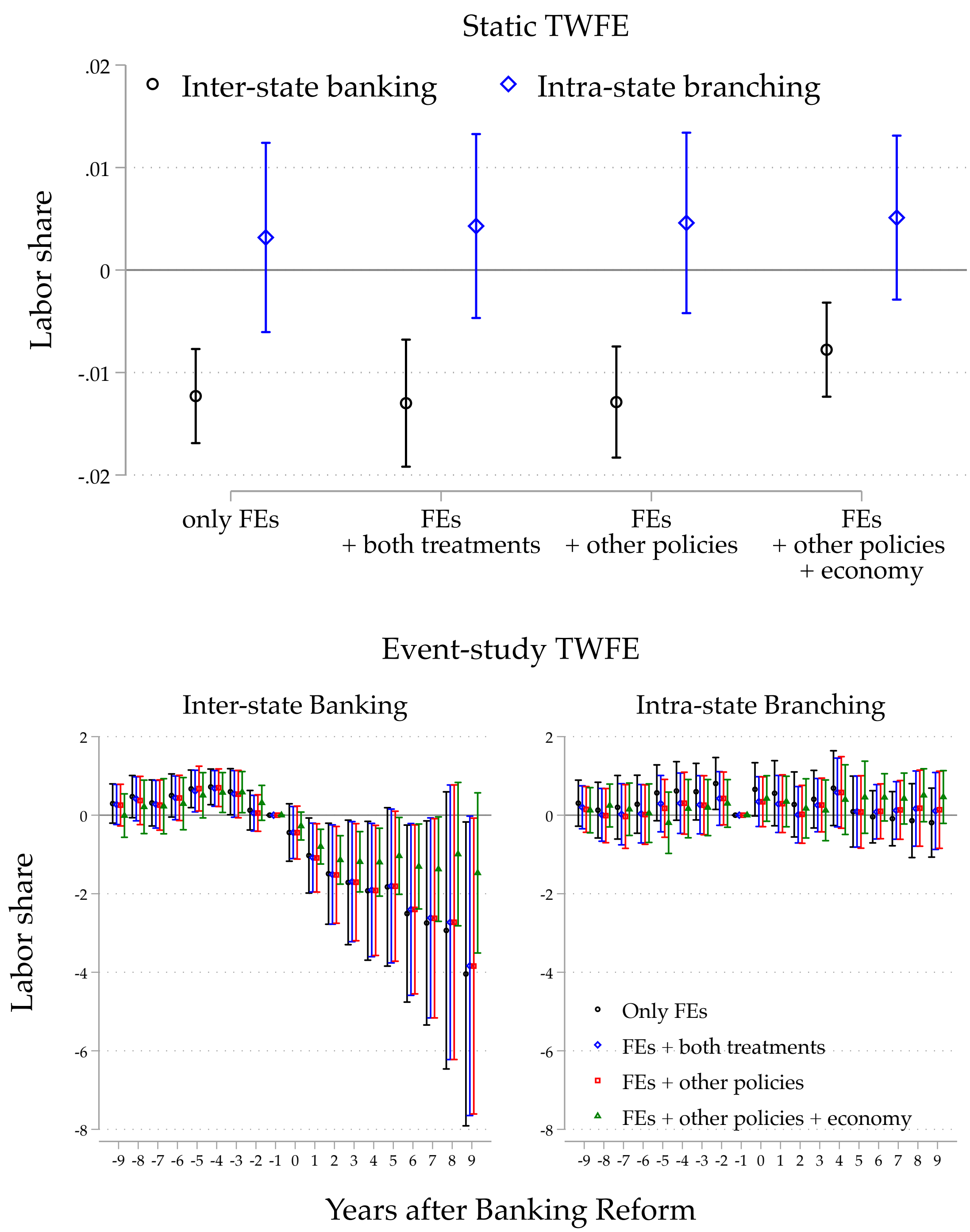
Given the staggered rollout of banking deregulation laws across US states, the TWFE specifications of Equations E.1 and E.2 might suffer from the negative weights bias highlighted by recent studies. Earlier liberalizers are implicitly used as controls for states that liberalize later on. Specifically,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

We employ the Goodman-Bacon (2021) diagnostic to decompose the static TWFE estimate into these two types of comparisons. Specifically, the Goodman-Bacon (2021) diagnostic is applied to a basic static TWFE specification that only includes two-way fixed effects and the treatment indicator for the policy under consideration. This corresponds to the first specification ('only FEs') reported in the top panel of Figure E.4.

We find that the static TWFE estimator of the effect of interstate banking deregulations

[22]The dynamic TWFE specification employed by LW 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 dynamic TWFE estimator as obtained by estimating equation 3 in the main text. This non-standard specification, however, does not influence results: applying a standard dynamic TWFE specification (as in Equation 3 in the main text) yields very similar results as those obtained by LW. This reassures us that any differences between the dynamic TWFE results of LW and our results from applying LP-DiD are due to the negative weights bias of TWFE, not to the non-standard specification of the dynamic TWFE model used by LW.

Figure E.4: Effect of banking deregulation on the labor share: TWFE estimates

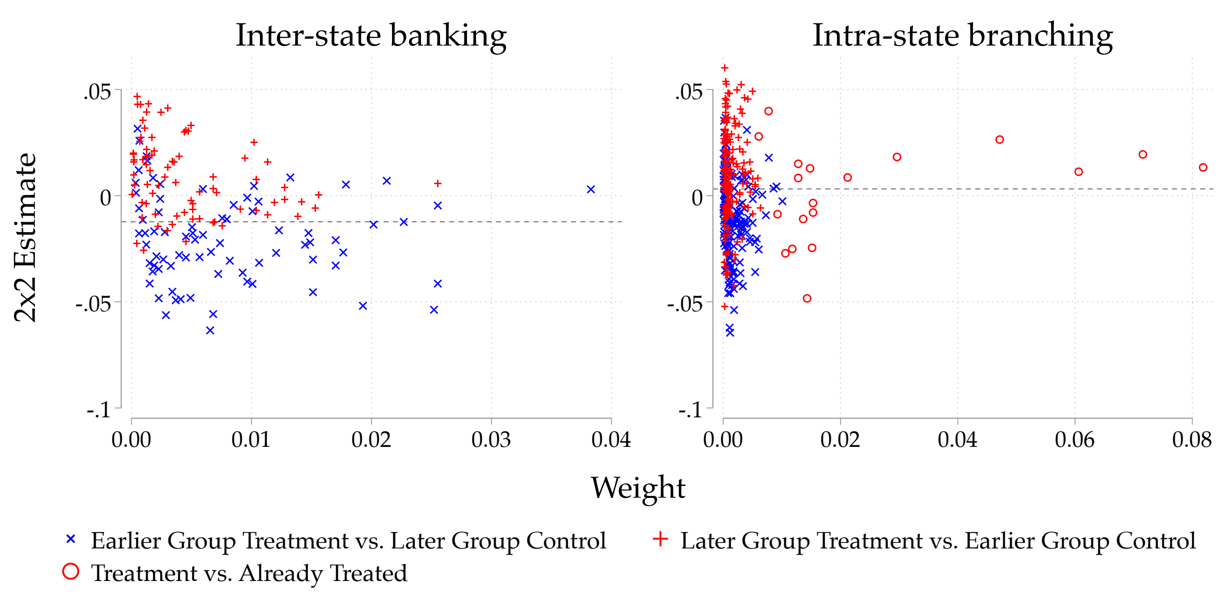


Notes: Estimates for the effect of banking deregulation on the labor share, using data from Leblebicioğlu and Weinberger, 2020. The top panel uses the static TWFE specification of equation E.1. The two bottom panels use the dynamic TWFE specification of equation E.2. These graphs replicates results from Leblebicioğlu and Weinberger, 2020. Following Leblebicioğlu and Weinberger, 2020, the additional controls are four lags of real State GDP, average corporate tax rate, and union membership rates.

assigns a 36% weight to unclean comparisons. For the estimates of the effect of intrastate branching deregulations, the problem is much more severe: unclean comparisons account for as much as 70% of the estimate.

Figure E.5 displays the results of the Goodman-Bacon, 2021 decomposition diagnostic. The figure plots each constituent 2x2 comparison that contributes to the static TWFE estimates, with its weight on the horizontal axis and its estimate on the vertical axis. The graph suggests that the estimates of the effects of branching deregulations are driven by a few 'unclean' comparisons – those involving states that deregulated before 1970 – that receive a very large weight. Notably, for both types of policies, clean comparisons produce overwhelmingly negative coefficients, while the unclean ones tend to bias the coefficients upwards.

Figure E.5: Goodman-Bacon, 2021 decomposition diagnostic for the static TWFE specification

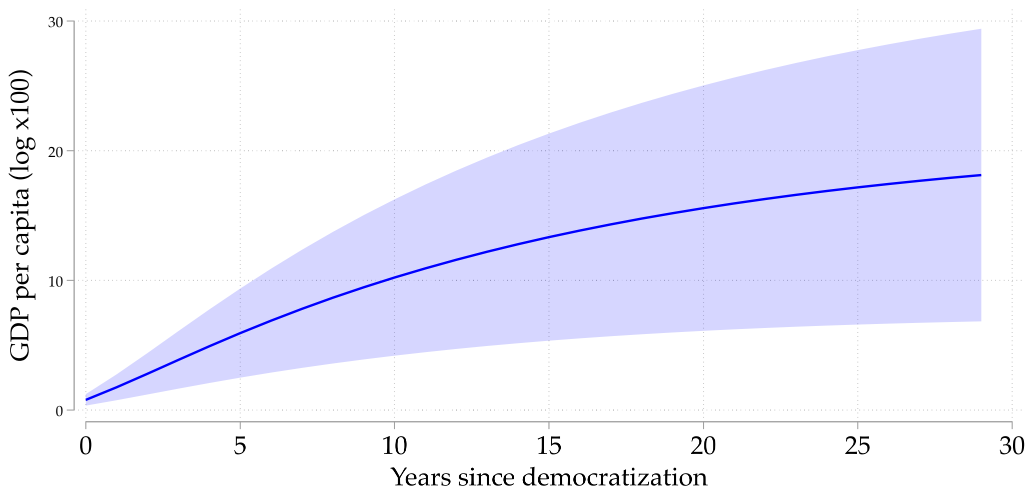


F Democracy and growth: Additional statistics and results

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

Figure F.6 displays the impulse-response function (IRF) from the estimation of the dynamic panel model of Equation 15 in the main text, reproducing the baseline results in Acemoglu et al., 2019.

Figure F.6: 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 15 in the main text. This graph replicates the baseline results from Acemoglu et al., 2019.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2019). "Democracy Does Cause Growth". In: *Journal of Political Economy* 127.1, pp. 47–100.

Álvarez, Javier and Manuel Arellano (2003). "The time series and cross-section asymptotics of dynamic panel data estimators". In: *Econometrica* 71.4, pp. 1121–1159.

Angrist, Joshua and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs". In: *Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Chaisemartin, Clément de and Xavier D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110.9, pp. 2964–96.

Chen, Shuowen, Victor Chernozhukov, and Iván Fernández-Val (2019). "Mastering Panel Metrics: Causal Impact of Democracy on Growth". In: *AEA Papers and Proceedings* 109, pp. 77–82.

Gardner, John, Neil Thakral, Linh T Tô, and Luther Yap (May 2024). "Two-stage differences in differences". https://jrgcmu.github.io/2sdd_gtty.pdf.

Ghanem, Dalia, Pedro H. C. Sant'Anna, and Kaspar Wüthrich (2024). *Selection and parallel trends*. arXiv: 2203.09001 [econ.EM]. URL: https://arxiv.org/abs/2203.09001.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Leblebicioğlu, Aslı and Ariel Weinberger (2020). "Credit and the Labour Share: Evidence from US States". In: *Economic Journal* 130.630, pp. 1782–1816.

Marx, Philip, Elie Tamer, and Xun Tang (2024). "Parallel Trends and Dynamic Choices". In: *Journal of Political Economy Microeconomics* 2.1, pp. 129–171. DOI: 10.1086/727363.

Mei, Ziwei, Liugang Sheng, and Zhentao Shi (2023). *Nickell Bias in Panel Local Projection: Financial Crises Are Worse Than You Think*. arXiv: 2302.13455 [econ.EM]. URL: https://arxiv.org/abs/2302.13455.

Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects". In: *Econometrica* 49.6, pp. 1417–1426.

Sun, Liyang and Sarah Abraham (2020). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

Viviano, Davide and Jelena Bradic (2023). *Dynamic covariate balancing: estimating treatment effects over time with potential local projections*. arXiv: 2103.01280.