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

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

This draft: 24th December 2022

Abstract

Recent applied microeconometrics research proposes various difference-in-differences (DiD) estimators for the problem of dynamic heterogeneous treatment effects. We show that the problem can be resolved by the local projections (LP) estimators of the sort used in applied macroeconometrics, combined with a flexible 'clean control' condition to appropriately define treated and control units. Our proposed LP-DiD estimator provides an overarching toolkit with several advantages. The method is clear, simple, easy to compute, and transparent and flexible in its handling of treated and control units. Moreover, it is quite general, including its ability to control for pre-treatment values of the outcome and of other time-varying covariates. The LP-DiD estimator does not suffer from the negative weighting problem, and indeed can be implemented with any weighting scheme the investigator desires. Simulations demonstrate the good performance of the LP-DiD estimator in common settings. Two empirical applications illustrate how LP-DiD addresses the bias of conventional fixed effects estimators, leading to potentially different results.

*We thank Colin Cameron, Sophie Hatte, Thomas Høgholm Jørgensen, and seminar partecipants at UC Davis, University of Milan La Statale, 2022 NBER Summer Institute, Asia School of Business (ASB), UMass Amherst and University of Pisa for helpful comments and exchanges. The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

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

[‡]Department of Economics, University of Massachusetts, Amherst (dgirardi@umass.edu).

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

[¶]Department of Economics and Graduate School of Management, University of California, Davis; NBER; and CEPR (amtaylor@ucdavis.edu).

1 Introduction

Difference-in-differences (DiD) is a widely used method for estimating causal impacts with observational data. In its canonical form, with only two time periods, only two groups of which one is treated, and under suitable assumptions (e.g., no anticipation and parallel trends), the DiD estimator can identify the average treatment effect on the treated.

Yet, as the scale and scope of DiD applications have widened over time and expanded into multi-period settings, its underpinnings have been stretched and doubts about the generality of its underlying assumptions have proliferated, as highlighted in many notable recent studies. The central matter of concern has been the appropriate implementation of DiD in an expanded set of situations where different units enter treatment at different dates (*staggered* treatment adoption) and treatment effects can occur gradually over time and be heterogeneous (Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2020; Goodman-Bacon, 2021; Borusyak et al., 2021). What was once a seemingly simple tool of general application increasingly appears to need bespoke adjustments to suit each specific situation.

In this paper we take a different angle on this problem, drawing out a potentially important link to a broader, flexible, encompassing family of alternative statistical techniques close at hand. Put simply, we bring to the fore an essential congruity between the concerns of applied microeconomists who encounter the challenge of estimating dynamic, heterogeneous, staggered treatment effects, and the concerns of applied macroeconomists who have long faced the task of estimating dynamic impulse-responses in time-series or panel data. Once understood this way, the scope for fertile interaction between these two strands of empirical work might seem obvious, despite its failure to happen quite yet. To prompt such a conversation, in this paper we re-frame the expanded set of DiD settings from the perspective of estimation via *local projections*, or LP, where the latter is the statistical technique introduced in a time-series context in Jordà (2005). We take these LP techniques from macroeconometrics in conjunction with the potential outcomes approach of microeconometrics to derive results for a wide range of DiD settings, seeking to develop a more general toolkit for implementing the DiD method.

Our proposed LP-DiD approach employs local projections to estimate dynamic effects and a flexible 'clean control' condition in the spirit of Cengiz, Dube, Lindner, and Zipperer (2019) to avoid the bias that can plague fixed-effects estimators when treatment adoption is staggered (Borusyak, Jaravel, and Spiess, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020). Intuitively, the bias of fixed-effects estimators arises because previously treated units, which might still be experiencing lagged time-varying and heterogeneous treatment effects, are implicitly used as controls for newly treated units. The clean control condition of LP-DiD avoids this bias by restricting the estimation sample so that 'unclean' observations, which outcome dynamics are still potentially influenced by a previous change in treatment status, are not part of the control group.

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

A skeptical reaction we can imagine hearing at this point is: why do we need yet another expanded DiD technique? Indeed, several alternative DiD estimators have recently been proposed to address the different settings that can arise in empirical applications without incurring in the 'negative weights' bias of two-way fixed-effects regression (Sun and Abraham, 2020; Callaway and Sant'Anna, 2020; Borusyak, Jaravel, and Spiess, 2021; de Chaisemartin and D'Haultfœuille, 2020; de Chaisemartin, D'Haultfœuille, Pasquier, and Vazquez-Bare, 2022; Gardner, 2021). Important distinguishing features of our proposed LP-DiD approach are the simplicity of its implementation, its ability to control for pre-treatment values of the outcome and of other covariates, and the flexibility it offers in the definition of the appropriate sets of treated and control units. The LP-DiD estimator is not specific to a particular setting, but can be applied in a variety of situations, providing an encompassing framework. In addition to its flexibility, 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.

Evidence from two Monte Carlo simulations suggests that the LP-DiD estimator performs well in staggered difference-in-differences settings, also in comparison with other estimators that have recently been proposed. Our simulations consider a binary staggered treatment with dynamic and heterogeneous effects. In the first simulation treatment timing is exogenous. Under this scenario, LP-DiD performs as well as the Sun and Abraham (2020) and Callaway and Sant'Anna (2020) estimators, while being computationally simpler and faster. In our second simulation, the probability of entering treatment depends on lagged outcome dynamics. In this second scenario, the ability of LP-DiD to match on pre-treatment outcomes allows it to outperform other estimators. The purpose of these simulations is not mainly that of performing a horse race between LP-

DiD and other estimators, but to show that LP-DiD performs well in plausible scenarios and that there is a class of settings – those in which matching on pre-treatment outcome dynamics or other pre-determined covariates is appropriate and important – in which LP-DiD could become the 'go-to' approach. We also note that in the exogenous treatment timing case, the LP-DiD estimate is identical to the estimate from a stacked regression approach as implemented in Cengiz, Dube, Lindner, and Zipperer (2019). However, the LP-DiD implementation is simpler (as it does not require stacking the data by events), and can be more easily generalized (e.g., conditioning on past outcomes).

Our two empirical applications employ LP-DiD to estimate the impact of banking deregulation on the labor share (replicating Leblebicioğlu and Weinberger 2020) and the effect of democratization on economic growth (replicating Acemoglu, Naidu, Restrepo, and Robinson 2019). These are two examples of important empirical settings in which conventional dynamic panel estimates are potentially subject to bias because of previously treated units being effectively used as controls, and matching on pre-treatment outcomes and other covariates is likely to be important.

The rest of this paper will be structured as follows. Section 2 draws a connection between the DiD method and the LP estimator, and presents our proposed LP-DiD specification. In Section 3 we use simulations to assess the performance of our LP-DiD approach, also in comparison with other new methods in the recent literature. In Section 4 we apply the LP-DiD estimator in two empirical applications. Section 5 concludes.

2 The local projections implementation of difference-in-differences

The aim of this Section is to clarify the connection between the difference-in-differences (DiD) method and the local projections (LP) estimator. We show how LP can be used to implement DiD in several different settings and with high flexibility.

While we start from simpler settings for the sake of clarity (sections 2.2 to 2.4), the core of this Section is the discussion of the case of binary staggered treatment with dynamic and heterogeneous treatment effects (section 2.5). In the staggered setting, the conventional two-way fixed-effects (TWFE) implementation of DiD, both in the static and event-study version, can suffer from 'negative weights' bias, as uncovered by recent important studies (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak et al., 2021; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020). We show that a LP approach can successfully address this problem.

Our main result is that in the staggered setting a properly specified LP regression, with clearly defined treated and control units following the 'clean control approach' of Cengiz et al. (2019), is able to identify a convex weighted average treatment effect, without incurring in the negative weights problem. We characterize explicitly the weights assigned to each cohort-specific treatment effect, and show that they are non-negative and proportional to group size and treatment variance. We also discuss how a simple re-weighted LP regression can recover an equally-weighted average treatment effect on the treated. We conclude the Section by briefly discussing how LP-DiD can be adapted flexibly to accommodate more general settings, including non-absorbing treatment, continuous treatment and inclusion of covariates (section 2.6).

2.1 General setup and notation

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

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 (ie, the set of units with $p_i = \infty$). We denote the time period in which group g enters treatment as p_g.

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

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

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

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

$$\tau_g(h) = E\left[y_{i,p+h}(p) - y_{i,p+h}(0) | p_i = p\right] \quad (1)$$

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

In other words, $\tau_g(h)$ represents the average dynamic effect, h periods after entering treatment, for all units belonging to a group g that enters treatment at time p.[2]

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

Assumption 1. No anticipation

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

Assumption 2. Parallel trends

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

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

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

where α_i is a unit-specific fixed effect, and δ_t is a time-specific effect common to all units.

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

Static two-way fixed-effects regression (static TWFE)

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

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

$$y_{it} = \alpha_i + \delta_t + \sum_{h=-Q}^{H} \beta_h^{TWFE} D_{i,t-h} + e_{it} \tag{4}$$

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

Local Projections regression (LP)

$$\Delta_h y_{it} = \delta_t^h + \beta^{h\ LP} \Delta D_{it} + e_{it}^h, \text{ for } h = 0, 1, ..., H. \quad (5)$$

where α_i are unit-specific intercepts; δ_t are common time-specific effects; $\Delta_h y_{it}$ is a h-periods forward long difference, defined as $\Delta_h y_{it} = y_{t+h} - y_{t-1}$; and e is an error term. The β terms are population regression coefficients, while the OLS estimates of these coefficients will be denoted by $\hat{\beta}$.

2.2 Basic DiD setting with two groups and two time periods

The connection between LP and DiD is easiest to see in the basic two-groups/two-periods setting. In this 2x2 setting, a LP regression at time horizon $h = 0$ is completely equivalent to a first-difference regression and to the static TWFE regression, which are well-known ways to implement the DiD method.

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

We are interested 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), the ATT in this setting can be rewritten as follows:

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

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

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

$$\Delta_0 y_{it} = y_{i2} - y_{i1} = \delta + \beta^{0\ LP} D_{i2} + e_i$$

Therefore we have

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

In other words, in the basic 2x2 setting the LP regression at time 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 well known, in this setting also the estimand β^{TWFE} from the static TWFE regression of equation 3 is equivalent to the coefficient from a first-difference regression and corresponds to β^{2x2} (Angrist and Pischke, 2009, pp.233-236). We thus have $\beta^{h=0\ LP} = \beta^{TWFE} = \beta^{2x2} = ATT$.

2.3 Two groups and multiple time periods

Now consider a slightly extended setting, with two groups (treated and control) observed over multiple time periods $T > 2$, and where all the treated units enter treatment in the same time period. Also in this setting, a 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 t^*, with $1 < t^* < T$, and remain treated thereafter, while control units are never treated over the sample period. Therefore, in pre-treatment periods $t < t^*$ no unit is treated. In post-treatment periods $t \geq t^*$, units in the treatment group are treated, while units in the control group are not.

In terms of our general setup and notation, we are setting $g \in \{0, 1\}$, where group 0 is the control group and group 1 the treatment group. For all units in the treatment group, $p_i = p_1 = t^*$. For all units in the control group, $p_i = p_0 = \infty$. With only one treated cohort, the dynamic ATT (equation 1) does not need the treatment group indicator, and becomes simply $\tau(h) = E[y_{i,t^*+h}(t^*) - y_{i,t^*+h}(0)|p_i = t^*]$.

Using the no-anticipation and parallel trends assumptions (Assumptions 1 and 2), we can write

$$\begin{aligned}
\tau(h) &\equiv E[y_{i,t^*+h}(t^*) - y_{i,t^*+h}(0)|p_i = t^*] = \\
&= E[(y_{i,t^*+h}(t^*) - y_{i,t^*-1}(0)) - (y_{i,t^*+h}(0) - y_{i,t^*-1}(0))|p_i = t^*] = \\
&= E[y_{i,t^*+h}(t^*) - y_{i,t^*-1}(0)|p_i = t^*] - E[y_{i,t^*+h}(0) - y_{i,t^*-1}(0)|p_i = \infty] = \\
&= E[\Delta_h y_{i,t^*}|p_i = t^*] - E[\Delta_h y_{i,t^*}|p_i = \infty] \equiv \beta(h)^{DiD}
\end{aligned}$$

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

The population coefficient $\beta^{h\ LP}$ from a LP regression (equation 5) corresponds exactly to this estimand. To see this, note that in this setting the LP regression of equation 5 is equivalent to the following cross-sectional regression, estimated on a subsample including all units but only the time period $t = t^*$:[3]

$$y_{i,t^*+h} - y_{i,t^*-1} = \delta^h + \beta^{h\ LP} \Delta D_{i,t^*} + e^h_{i,t^*}$$

Therefore we have

$$\beta^{h\ LP} = E[\Delta_h y_{i,t^*} | \Delta D_{i,t^*} = 1] - E[\Delta_h y_{i,t^*} | \Delta D_{i,t^*} = 0] = \beta(h)^{DiD} = \tau(h)$$

It follows from results in the recent literature on DiD (for example de Chaisemartin and D'Haultfœuille 2020; Gardner 2021; Sun and Abraham 2020; Goodman-Bacon 2021) that in this setting with only one treated cohort, and under Assumptions 1 and 2, also the coefficients in the event-study TWFE regression (equation 4) correspond to the $\tau(h)$ estimands.[4] Moreover, the β^{TWFE} estimand from the static TWFE regression (equation 3) equals the ATT, defined as $E(\tau_{it} | D_{it} = 1)$.[5]

2.4 Staggered treatment adoption with dynamic but homogeneous treatment effects

Let us now allow for the presence of multiple treated groups which enter treatment at different points in time (treatment is *staggered*). In this Section we provisionally assume that the average treatment effect path does not differ across treatment cohorts (treatment effects are *homogeneous*). In terms of our general setup and notation, we have $G > 1$ and $\tau_g(h) = \tau(h)$ for all $g > 0$.

In this setting with staggered treatment and dynamic but homogeneous treatment effects, a LP regression (equation 5) augmented with an adequate number of lags and

[3]This equivalence holds because at time periods different from t^* there is no variation in the regressor ΔD_{it}. For this reason, observations for any time period $t \neq t^*$ do not contribute to the estimated coefficient $\beta^{h\ LP}$. In other words, in this setting the coefficient $\beta^{h\ LP}$ is only identified from the observations for time $t = t^*$.

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

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

leads of the treatment indicator is able to recover the average treatment effect path under parallel trends and no-anticipation.

To derive this conclusion, consider that, under Assumptions 1 and 2 and homogeneous treatment effects, mean observed outcomes at time $t+h$ are given by

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

Subtracting $E[y_{i,t-1}]$ from both sides and defining $\delta_t^h = \delta_{t+h} - \delta_{t-1}$, we obtain[6]

$$E[\Delta_h y_{i,t}] = \delta_t^h + \tau(h)\Delta D_{i,t} + \sum_{j=1}^{h} \tau(h-j)\Delta D_{i,t+j} + \sum_{j=1}^{\infty}[\tau(h+j) - \tau(j-1)]\Delta D_{i,t-j}$$

Therefore the dynamic ATT $\tau(h)$ corresponds to the $\beta^{h\ LP}$ population coefficient in the following LP regression:

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

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

What do the static and event-study TWFE specifications of equations 3 and 4 identify in this setting with staggered treatment and dynamic but homogeneous effects?

Results from the recent DiD literature imply that a static TWFE regression (equation 3) can suffer from bias if treatment effects are dynamic (in the sense that $\tau(h) \neq \tau(h+1)$ for some h), even under parallel trends, no-anticipation and homogeneity across treatment cohorts.[7] 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 TWFE estimate with a negative weight (de Chaisemartin and D'Haultfœuille, 2020).

Under the assumption of homogeneous treatment effects, however, event-study TWFE

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

[7]Heterogeneous effects, that we consider below in Section 2.5, would exacerbate the issue.

regression (equation 4) does not suffer from this bias and, like the LP regression with lags and leads of treatment discussed above, is able to recover the average treatment effect path under parallel trends and no anticipation, as long as a sufficient number of lags of the treatment indicator is included (see Sun and Abraham (2020), in particular Proposition 4 and Equation 19). Intuitively, the lagged treatment indicators control for the lagged dynamic effects of previous treatments, which in this setting are the same (in expectation) for all units.

2.5 Staggered treatment adoption with dynamic and heterogeneous treatment effects

Let us now abandon the assumption of homogeneity of the treatment effect path, and allow for heterogeneous treatment effects across different cohorts. Formally, we have $\tau_g(h) \neq \tau_{g'}(h)$ for at least some time-horizon h and some pair of groups $g' \neq g$. This case has been the main focus of a growing recent literature (e.g., de Chaisemartin and D'Haultfœuille 2020; Sun and Abraham 2020; Callaway and Sant'Anna 2020; Goodman-Bacon 2021; Borusyak et al. 2021).

With heterogeneous treatment effects, the static TWFE estimator of equation 3 is biased both because of dynamic lagged effects and heterogeneity. Define a cell as a given treatment group g in a given period t. de Chaisemartin and D'Haultfœuille (2020) show that β^{TWFE} in equation 3 provides a weighted average of all cell-specific ATTs, but with weights that can be negative. Negative weights introduce bias: for example, positive cell-specific effects can enter the formula for the TWFE cofficient with a negative sign. Another way to see this problem is through the Goodman-Bacon (2021) decomposition theorem, which shows that the static TWFE estimator in equation 3 is an average of all potential 2x2 comparisons in the data, with weights based on subsample shares and treatment variances. The problem is that some of these 2x2 comparisons are 'unclean' comparisons in which previously treated units are used as controls for newly-treated units. These 'unclean comparisons' are the source of the 'negative weights' bias of static TWFE.[8]

With heterogeneous treatment effects across cohorts, also the event-study TWFE specification of equation 4 is generally biased (Sun and Abraham, 2020). Sun and

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

Abraham (2020) show that the relative-period coefficients (ie, the coefficients on leads and lags of treatment in equation 4) can be contaminated by effects from other periods.

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

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

Subtracting $E[y_{i,t-1}]$ from both sides, we obtain[9]

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

Without appropriate adjustment to take into account the last two sums on the right side of equation 8, the simplest LP regression function of equation 5 would be misspecified in this setting. Indeed, the population regression coefficient $\beta^{h\ LP}$ in equation 5 would correspond to the following expression

$$
\begin{aligned}
E[\beta^{h\ LP}] &= E(\Delta_h y_{it}|t, \Delta D_{it} = 1) - E(\Delta_h y_{it}|t, \Delta D_{it} = 0) &= \\
&= E(\textstyle\sum_{g}^{G}[\tau_g(h) \times \mathbf{1}\{t = p_g\}]) \\
&\quad - E(\textstyle\sum_{g=1}^{G}[\sum_{j=1}^{\infty}((\tau_g(h+j) - \tau_g(j-1)) \times \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\})]) \\
&\quad - E(\textstyle\sum_{g=1}^{G}[\sum_{j=1}^{h}(\tau_g(h-j) \times \Delta D_{i,t+j} \times \mathbf{1}\{t = p_g - j\})])
\end{aligned}
\tag{9}
$$

Equation 9 shows that, without appropriate adjustment, the population regression coefficient $\beta^{h\ LP}$ from the simplest LP regression of equation 5 corresponds to a weighted average dynamic ATT, plus two bias terms.

The first source of bias is the presence of previously treated units in the control group, ie 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

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

effects. Note that this bias exists as long as $\tau_g(h+j) \neq \tau_g(j-1)$: treatment effects evolve gradually over time. As a result, the dynamic changes in treatment effects that these previously treated units might be experiencing enter equation 9 with a *negative* sign. This is a manifestation of the 'negative weights' bias discussed by the recent literature on DiD (Goodman-Bacon (2021); de Chaisemartin and D'Haultfœuille (2020); Callaway and Sant'Anna (2020); Sun and Abraham (2020); Borusyak et al. (2021))

Moreover, in the LP setting, a second potential source of bias is the presence in the control group of units that are treated between $t+1$ and $t+h$, ie observations such that $\Delta D_{it} = 0$ but $\Delta D_{i,t+j} \neq 0$ for some j in $1 \leq j \leq h$.[10]

We are now ready to present our main contribution: A properly specified LP regression (which we call LP-DiD) solves these problems and identifies a convex combination of cohort-specific effects. LP-DiD consists in estimating the LP specification of equation 5 in a restricted sample that only includes newly treated observations ($\Delta D_{it} = 1$) and not-yet treated ones ($\Delta D_{i,t-j} = 0$ for $-h < j < \infty$). Under the assumption of absorbing binary treatment, the restriction imposed on the control group ($\Delta D_{i,t-j} = 0$ for $-h < j < \infty$) simplifies to $D_{i,t+h} = 0$. Intuitively, as recent literature has made clear and as equations 8 and 9 illustrate, 'negative weights' bias comes from unclean comparisons in which previously treated units are used as controls for newly-treated units. Excluding these 'unclean' observations from the control group eliminates the bias.

Formally, consider the following LP regression, which we dub LP-DiD:

LP-DiD regression

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

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

restricting the sample to observations that are either:

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

By removing previously treated observations and observations treated between $t+1$ and $t+h$ from the control group, $\beta^{h\ LP-DiD}$ from regression 10 provides a convex combination of all group-specific effects $\tau_g(h)$. Indeed we have

$$\begin{aligned} E[\beta^{h\ LP-DiD}] &= E(\Delta_h y_{it}|t, \Delta D_{it} = 1) - E(\Delta_h y_{it}|t, \Delta D_{it} = 0, D_{i,t+h} = 0) = \\ &= E(\textstyle\sum_g^G [\tau_g(h) \times \mathbf{1}\{t = p_g\}]) \end{aligned}$$

2.5.1 Weights of the LP-DiD estimator

Let us now characterize explicitly the weights assigned to each cohort-specific effect $\tau_g(h)$ when the LP-DiD specification (equation 10) is estimated through OLS. The key result is that, under parallel trends and no-anticipation (Assumptions 1 and 2), the LP-DiD estimator identifies a weighted average of all cohort-specific treatment effects, with weights that are always positive and depend on treatment variance and subsample size. Here we present this result; a simple formal derivation based on the Frisch-Waugh-Lovell theorem is in Appendix A.

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

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

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

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

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

where $N_{CCS_{g,h}}$ is the number of observations in the clean control sample for group g at time-horizon h; $n_{g,h} = N_g / N_{CCS_{g,h}}$ is the share of treated units in the $CCS_{g,h}$ subsample; and $n_{c,g,h} = N_{c,g,h} / N_{CCS_{g,h}}$ is the share of control units in the $CCS_{g,h}$ subsample.[11]

In short, the LP-DiD estimator $\beta^{h\ LP-DiD}$ identifies a variance-weighted ATT (*VWATT* in the terminology of Goodman-Bacon (2021)).

If a researcher is instead interested in a equally-weighted ATT, they can employ a re-weighted LP-DiD regression. Indeed, equations 11 and 12 imply that estimation of a LP-DiD regression (equation 10) through weighted least squares, assigning to an observation belonging to $CCS_{g,h}$ a weight equal to $1/(\omega_{g,h}^{LP-DiD}/N_g)$, identifies the equally-weighted ATT.

In practical applications, the quantity $(\omega_{g,h}^{LP-DiD}/N_g)$ can be obtained by computing subsamples sizes and shares of treated and control units in the sample and using equation 12, or through an auxiliary regression. Specifically, consider an auxiliary regression of ΔD on time indicators in the sample defined by condition 10. Define $\widetilde{\Delta D}_{g,p_g}$ as the residual at time p_g for a unit belonging to group g.[12] The Frisch-Waugh-Lovell theorem implies that $(\omega_{g,h}^{LP-DiD}/N_g) = \frac{\widetilde{\Delta D}_{g,p_g}}{\sum_{g \neq 0} N_g \widetilde{\Delta D}^2_{g,p_g}}$ (see also Appendix A).

In this setting without inclusion of covariates, the LP-DiD estimator $\beta^{h\ LP-DiD}$ is equivalent to the estimate from a stacked regression approach as implemented in Cengiz et al. (2019).[13] However, the LP-DiD implementation is simpler, faster and less prone to errors in practical applications, given that it does not require the reshaping of the dataset in stacked format. Moreover, as we discuss below, the LP-DiD specification is easier to generalize, for example by conditioning on pre-treatment values of the outcome or of other covariates.

2.6 Extensions: Covariates, non-absorbing treatment and non-binary treatment

2.6.1 Inclusion of covariates

The LP-DiD specification of equation 10 can easily be augmented to include both time-invariant and time-varying covariates. Including covariates might be necessary for identification, if the parallel trends assumption only holds conditional on some variables, or to increase precision.

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

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

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

In particular, a distinctive feature of the LP-DiD approach is that it allows to control for *pre-treatment* values of time-varying covariates, including lagged outcome dynamics. This is made possible by the structure of the LP specification. Unlike a standard event-study TWFE specification or most alternative estimators proposed in the recent literature, the LP specification implies that any lagged variable included in the estimating equation is measured before treatment.[14]

With binary absorbing treatment, a LP-DiD specification which controls for P lags of the outcome dynamics and contemporaneous and lagged values of a vector of M covariates can be written as follows.

LP-DiD regression with exogenous covariates and lagged outcome dynamics

$$\begin{aligned} y_{i,t+h} - y_{i,t-1} = \;& \beta^{h\ LP-DiD} \Delta D_{it} && \}\text{ treatment indicator} \\ & + \textstyle\sum_{p=1}^{P} \gamma_p^h \Delta y_{i,t-p} && \}\text{ outcome lags} \\ & + \textstyle\sum_{m=1}^{M} \sum_{p=0}^{P} \gamma_{m,p}^h \Delta x_{m,i,t-p} && \}\text{ covariates} \\ & + \delta_t^h && \}\text{ time effects} \\ & + e_{it}^h\ ; && \quad \text{for } h = 0, \ldots, H\,, \end{aligned} \tag{13}$$

restricting the sample to observations that are either

$$\begin{cases} \text{newly treated} & \Delta D_{it} = 1\,, \\ \text{or clean control} & D_{i,t+h} = 0 \end{cases} \tag{14}$$

Appendix A.2 discusses the weights assigned to each group-specific effect in the specification that includes control variables. The main result is that the weights are guaranteed to remain the same as in equation 12 if covariates have linear and homogeneous effects. In more general settings, the presence of covariates will alter the weighting scheme in ways that are difficult to characterize analytically, but for the purpose of obtaining a equally-weighted ATT the weights can still be recovered through the auxiliary regression described above.

If a researcher wants to preserve the variance-weighting scheme of the baseline specification without covariates, or to avoid other possible drawbacks from the inclusion

[14]Of course, controlling for pre-treatment outcome dynamics (as well as any other exogenous or predetermined covariate) will be appropriate in some applications but not in others. A discussion of the conditions under which it is appropriate or necessary to control for lagged outcome dynamics and other covariates in the DiD setting is outside the scope of this paper. What matters here is that the LP-DiD estimator offers flexibility in this respect: the researcher can decide whether to control for lagged outcomes and other covariates based on the application.

of covariates in linear regression in the DiD setting (discussed by Sant'Anna and Zhao (2020)), it is easy to control for covariates semiparametrically using propensity-score based methods in the spirit of Sant'Anna and Zhao (2020). Jorda and Taylor (2016) discuss the implementation of propensity-score based methods in the LP setting and apply them to estimate the effects of fiscal consolidation.

2.6.2 Non-absorbing treatment

In many applications, treatment is not absorbing: units can enter and exit treatment multiple times. The LP-DiD framework offers flexibility to accommodate the different definitions of the causal effect of interest and the different identification assumptions that might be appropriate under non-absorbing treatment.

Appropriate modification of the 'clean control' sample restriction of equation 10 will generally be necessary to implement LP-DiD in the non-absorbing treatment setting.

One can recover, for example, the effect of entering treatment for the first time and staying treated, relative to a counterfactual of remaining untreated, by using the LP-DiD specification of equation 10 but modifying the 'clean control' sample restriction as follows:

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

In some settings the 'clean control' condition in equation 15 might not be feasible or appropriate. Consider, for example. the problem of estimating the effect of minimum wage increases in a panel of regions. For later time periods t, there will be very few regions that have never experienced a minimum wage increase until period $t+h$.

This case can be dealt with in a simple way in the LP-DiD framework, under the additional assumption that dynamic treatment effects stabilize after a finite number of periods. Formally, we introduce the following assumption.

Assumption 3. Dynamic effects stabilize after K^* periods:

$$\tau_g(K^*) = \tau_g(K^* + k) \text{ for } k \geq 0$$

Define $K = K^* + 1$. Assumption 3 implies that, for any time horizon h and for any $j \geq K$, we have $\tau_g(h+j) = \tau_g(j-1)$. Therefore, under Assumptions 1, 2 and 3, equation 9 becomes

$$
\begin{aligned}
E[\beta^{h\ LP}] \ = \ & E(\textstyle\sum_g^G[\tau_g(h) \times \mathbf{1}\{t = p_g\}]) \\
& - E(\textstyle\sum_{g=1}^G[\sum_{j=1}^K((\tau_g(h+j) - \tau_g(j-1)) \times \Delta D_{i,t-j} \times \mathbf{1}\{t = p_g + j\})]) \\
& - E(\textstyle\sum_{g=1}^G[\sum_{j=1}^h(\tau_g(h-j) \times \Delta D_{i,t+j} \times \mathbf{1}\{t = p_g - j\})])
\end{aligned}
\tag{16}
$$

Equation 16 implies that bias only comes from observations that experience a change in treatment status between time $t-K$ and $t-1$ or between $t+1$ and $t+h$.

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

$$
\begin{cases}
\text{treatment} & (D_{i,t+j} = 1 \text{ for } 0 \leq j \leq h) \text{ and } (D_{i,t-j} = 0 \text{ for } 1 \leq j \leq K) \\
\text{or clean control} & \Delta D_{i,t-j} = 0 \text{ for } -h \leq j \leq K
\end{cases}
\tag{17}
$$

In Section 4.2 we will illustrate the use of LP-DiD under non-absorbing treatment through an empirical application which estimates the effect of democracy on economic growth.

2.6.3 Continuous treatment

A detailed formal discussion of the issues that can arise under continuous treatment is outside the scope of this paper (see for example de Chaisemartin et al. 2022). However, we do argue that the LP-DiD framework offers flexibility to accommodate the different definitions of the causal effect of interest and the different identification assumptions that might be appropriate with continuous treatment. For example, the clean control condition can be adapted to define clean controls as 'stayers' (or alternatively 'quasi-stayers'), in the terminology of de Chaisemartin, D'Haultfœuille, Pasquier, and Vazquez-Bare (2022).[15]

3 Simulations

We conduct two Monte Carlo simulations to illustrate the performance of the LP-DiD estimator. We consider a binary staggered treatment, with dynamic and heterogeneous treatment effects. In the first simulation, treatment is exogenous; the parallel trends

[15] For example, in a recent working paper Hoyos (2022) employs our proposed LP-DiD approach, with an appropriately specified clean control sample restriction, to estimate the effect of (continuous) changes in average tariff rates on GDP per capita, in a panel of industrialized and developing economies. In this application, treatment is both non-absorbing and continuous. Hoyos (2022)'s application of the LP-DiD clean control condition consists in excluding from the estimation sample observations that experience a large change in average tariff rates between year $t-10$ and $t-1$ or between $t+1$ and $t+h$.

assumption holds and the conventional TWFE model only fails because of heterogeneous dynamic effects, which lead to the 'negative weighting' problem. In the second simulation, treatment is endogenous; specifically, the probability of receiving treatment depends on previous outcome dynamics.

We compare the performance of our LP-DiD estimator with (a) a conventional event-study TWFE specification; (b) the Sun and Abraham (2020) estimator; and (c) the Callaway and Sant'Anna (2020) estimator. Results suggest that, unlike the conventional TWFE specification, LP-DiD tracks well the true effect path even in the presence of heterogeneity. With exogenous treatment, LP-DiD performs as well as the Sun and Abraham (2020) and Callaway and Sant'Anna (2020) estimators. When the probability of treatment depends on lagged outcome dynamics, the ability of LP-DiD to match on pre-treatment outcomes makes it outperform other estimators.

Setting

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

$$Y_{0it} = \rho Y_{0,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}\,, \tag{18}$$

with $-1 < \rho < 1$, and with $\lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25)$.

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

$$\beta_{it} = \begin{cases} 0 & \text{if } t - \tau_i < 0 \\ \alpha_0(t-\tau_i) + \alpha_1(t-\tau_i)^2 + \alpha_2\frac{(t-\tau_i)^2}{(\tau_i/\tau_1)^2} & \text{if } 0 \leq t-\tau_i < 20 \\ \alpha_0 20 + \alpha_1 20^2 + \alpha_2\frac{20^2}{(\tau_i/\tau_1)^2} & \text{if } t-\tau_i \geq 20\,, \end{cases} \tag{19}$$

where τ_i is the period in which unit i enters treatment (with $\tau_i > T$ if unit i is never treated during the sample period) and τ_1 is the treatment period for the 'earliest adopter' in the sample. We set $\alpha_0 = 2$; $\alpha_1 = 0.05$; $\alpha_2 = 0.95$.

Observed outcomes Y_{it} are therefore given by

$$Y_{it} = Y_{0it} + \beta_{it} \tag{20}$$

Simulation 1: Exogenous treatment timing In simulation 1, we assume that treatment is exogenous. Specifically, units are randomly assigned to 10 groups, each of size $N/10$. One group never receives treatment; the other nine groups receive treatment respectively at time $\tau = 11, 13, 15 \ldots, 27$.

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

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

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

Results

We perform 200 replications of each of our two simulations. We apply four estimators to our synthetic data:

- A conventional two-way-fixed-effects model, using an event-study specification with leads and lags of a treatment indicator.
- Our LP-DiD estimator.[16]
- The Sun and Abraham (2020) estimator.
- The Callaway and Sant'Anna (2020) estimator.

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

Results from the simulation with exogenous treatment timing (Simulation 1) are presented in Figure 1 and Table 1. The conventional event-study TWFE specification does an extremely poor job in our setting, due to the heterogeneity of treatment effects.[17] Our

[16]We do not apply any re-weighting, thus using the variance-weighting scheme described in Section 2.5.1 and Appendix A.

[17]The fact that in our simulated DGP the size of the effect is a function of the date of treatment makes the 'negative weighting' problem particularly severe, and therefore the performance of the TWFE specification particularly poor. We choose this DGP in order to test the performance of our estimator in a setting in which the flaws of the conventional estimator are particularly severe.

LP-DiD estimator, instead, tracks quite well the average true effect (Figure 1). Table 1, which reports the Root Mean Squared Error of each estimator at different time horizons, shows that in this setting the LP-DiD estimator does at least as well as the Sun-Abraham and Callaway-Sant'Anna estimators.

Results from the simulation with endogenous treatment timing (Simulation 2) are reported in Figure 2 and Table 2. In applying our LP-DiD estimator in this setting, we include one lag of the change in the outcome variable as a control. While the Sun and Abraham (2020) and Callaway and Sant'Anna (2020) estimators do allow for the inclusion of time-invariant control variables, there is no straightforward way to control for pre-treatment lags of the outcome in their specification, as these estimators are not designed to condition on pre-determined time-varying covariates.

The ability of the LP-DiD estimator to match on pre-treatment outcome dynamics in a straightforward way, allows it to outperform other estimators in the presence of this particular failure of the parallel-trends assumption. The LP-DiD estimator tracks quite well the true dynamic effect also in this setting (Figure 2) and it also has the lowest RMSE (Table 2).

Computational speed

We also employ our simulated dataset to assess quantitatively the computational advantage of LP-DiD relative to other recently proposed estimators. We record the computation time required for estimating the treatment effect path in a single simulation of our synthetic dataset with exogenous treatment timing. We use the STATA software on a laptop with 2.80 GHz Quad-core Intel i7 Processor and 16 GB of Ram. The LP-DiD estimator runs in 1.2 seconds, similar to the (biased) event-study TWFE estimator (1.04 seconds) and more than 100 times faster than the Callaway and Sant'Anna (2020) and Sun and Abraham (2020) estimators, that in our setting require respectively 144.6 and 198.5 seconds.

4 Empirical Applications

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

4.1 Credit and the labor share

We replicate the Leblebicioğlu and Weinberger (2020) analysis of the effect of banking deregulation on the labor share in US states.

Starting in the late 1970s, US states began removing restrictions on the ability of out-of-state banks to operate in the state (interstate banking deregulation) and on the ability of in-state banks to open new branches (intra-state branching deregulation). Leblebicioğlu and Weinberger (2020) estimate the effect of both inter-state and intra-state banking deregulation laws on the labor share of value added. They conclude that inter-state banking deregulation has a sizable negative effect on the labor share, while they find no effect of intra-state branching deregulation.

The dataset covers the 1970–1996 period. (In 1997, inter-state banking deregulation was imposed in all states by federal law.) Figure 3, which reproduces Figure 1 in Leblebicioğlu and Weinberger (2020), displays the share of US states with a liberalized banking sector.

4.1.1 Conventional TWFE specifications

We first consider the following static TWFE specification for the effect of banking deregulation laws, which replicates Leblebicioğlu and Weinberger (2020)'s baseline specification:

$$LS_{st} = \beta_{Bank} Bank_{st} + \beta_{Branch} Branch_{st} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st} \,, \tag{21}$$

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

To assess possible pre-trends and lagged effects, Leblebicioğlu and Weinberger (2020) also estimate the following event-study TWFE specification:

$$LS_{st} = \sum_{q=-9}^{9} \beta_{Bank,t+q} \Delta Bank_{s,t+q} + \sum_{q=-9}^{9} \beta_{Branch,t+q} \Delta Branch_{s,t+q} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st} \,. \tag{22}$$

4.1.2 Forbidden comparisons in the TWFE specifications

Given the staggered rollout of banking deregulation laws across US states, the TWFE specifications of equations 21 and 22 suffer from the issues highlighted by recent studies (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). Earlier liberalizers are used as controls for states that liberalize later on. Specifically, the specifications in equations 21 and 22 produce a weighted average of two types of comparisons: (1) newly treated states vs. not-yet treated states and (2) newly treated states vs. earlier treated states (Goodman-Bacon, 2021).

We employ the Goodman-Bacon (2021) diagnostic to decompose the TWFE estimate from equation 21 into these two types of comparisons. While 'unclean' 2x2 comparisons with earlier treated units as controls contribute to (and potentially bias) the TWFE estimates of both the policies studied, the estimates of the effect of intrastate branching deregulation are affected most severely. The static TWFE estimator of the effect of interstate banking deregulations assigns a overall weight of 63% to 'clean' comparisons of earlier treated versus not-yet treated states, and 36% to 'unclean' comparisons that use earlier treated units as controls. For the estimates of the effect of intrastate branching deregulations, the problem is much more severe: 'clean' comparisons receive a weight of only 30%. The remaining 70% is accounted for by two types of unclean comparisons: later treated units versus earlier treated units (23%) and treated units versus units that are already treated in the first period of the panel (47%).

Figure 4 displays the results of the Goodman-Bacon (2021) decomposition diagnostic. The figure plots each constituent 2x2 comparison that contributes to the static TWFE estimates of equation 21, 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.

4.1.3 LP-DiD specification

In order to avoid the biases of the conventional TWFE specifications, and to allow for matching based on pre-treatment outcome dynamics, we re-estimate the effect of banking deregulation laws using the following LP-DiD specification:[18]

$$LS_{s,t+h} - LS_{s,t-1} = \alpha_t^h + \beta_h^{LP-DiD} \Delta Bank_{s,t} + \sum_{m=1}^{M} \gamma_m^h \Delta LS_{s,t-m} + \sum_{m=1}^{M} \eta_m^h X_{s,t-m} + e_{s,t}^h \qquad (23)$$

[18]Given that treatment is absorbing in this data, and there is a sufficient number of not-yet treated States at all points in time, we employ the version of the clean control condition which uses only untreated units as controls.

restricting the sample to observation that are either:

$$\begin{cases} \text{treatment} & \Delta Bank_{s,t} = 1 \\ \text{control} & Bank_{s,t+h} = 0 \end{cases} \tag{24}$$

4.1.4 Results

Figure 5 displays results from the static TWFE specification of equation 21, while Figure 6 displays results from the event-study TWFE specification of equation 22. These results replicate the estimates reported in Table 2 and Figure 2 of Leblebicioğlu and Weinberger (2020). They suggest that the liberalization of inter-state banking has a sizable negative effect on the labor share, although they also show some (relatively small) pre-treatment trend. Instead, the estimated effects of intra-state branching deregulation on the labor share are positive, small and very imprecise.

Figure 7 displays results from the LP-DiD estimator with clean controls. The negative effect of inter-state banking deregulation on the labor share is confirmed, including when controlling for pre-treatment outcome dynamics. Estimates of the effect of intra-state branching deregulation, instead, change dramatically. 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.

4.2 Democracy and economic growth

Our second empirical application estimates the effect of democracy on economic growth, replicating the analysis in Acemoglu, Naidu, Restrepo, and Robinson (2019).

The dataset covers 175 countries from 1960 to 2010. The treatment indicator is a binary measure of democracy, which Acemoglu, Naidu, Restrepo, and Robinson (2019) build from several datasets to mitigate measurement error. The main outcome variable of interest is the log of GDP per capita, obtained from the World Bank Development Indicators.

Three features of this application make it particularly meaningful and interesting. First, there is potential for negative weighting: fixed effects regression would use older democracies as controls for new democracies. Second, treatment is non-absorbing: democracies can slide back into autocracy, and there are indeed multiple instances of reversals in the data. Third, controlling for pre-treatment outcome dynamics is crucial, since there is evidence of selection: Acemoglu, Naidu, Restrepo, and Robinson (2019) show that democratisation tends to be preceded by a dip in GDP per capita.

4.2.1 Dynamic panel specifications

The baseline results in Acemoglu, Naidu, Restrepo, and Robinson (2019) are obtained from the following dynamic fixed effects specification:

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct}\,, \tag{25}$$

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 fact that democratizations tend to be preceded by a decline in GDP per capita.

Estimated coefficients from eq. 25 are then used to build a impulse response function for the dynamic effect of GDP. These estimates also allow to derive the cumulative long-run effect of a permanent transition to democracy, given by $\frac{\hat{\beta}}{1-\sum_{j=1}^{p}\hat{\gamma}_j}$.

This dynamic fixed effects specification, however, might suffer from bias if treatment effects are dynamic and heterogeneous, as highlighted in the recent literature.

4.2.2 LP-DiD specifications

Consider the following LP-DiD specification for estimating the effect of democracy on growth:

$$y_{c,t+h} - y_{c,t-1} = \beta_h^{LP\ DiD} \Delta D_{ct} + \delta_t^h + \sum_{j=1}^{p} \gamma_j^h y_{c,t-j} + \epsilon_{ct}^h\,. \tag{26}$$

restricting the sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1, D_{i,t-1} = 0 \\ \text{clean controls} & D_{i,t+k} = 0 \text{ for } -K \leq k \leq h\,. \end{cases} \tag{27}$$

In words, for all years t and for each time-horizon h, treated units are countries that democratize at t, and control units are countries that have been non-democracies continually from $t-K$ to $t+h$.

This is an example of how the LP-DiD framework can be easily adapted to a setting in which treatment is not absorbing, and treatment reversals (in this case, democracies sliding back into autocracy) are possible.[19]

In a section of their analysis, Acemoglu, Naidu, Restrepo, and Robinson (2019) employ a semiparametric local projections specification that can be seen as a special version of the LP-DiD estimator above. Specifically, they estimate equation 26 with the following condition for the control group: $D_{it} = D_{i,t-1} = 0$. Their specification can thus be seen as an LP-DiD specification, in which the time-window for defining admissible ('clean') control units is only one period ($H = 1$ in equation 27), and treatment status between $t + 1$ and $t + h$ is not constrained.

Seeing the Acemoglu, Naidu, Restrepo, and Robinson (2019) semiparametric specification as a version of LP-DiD provides a useful novel perspective on their analysis and what possible deviations from their specification should be considered. Acemoglu, Naidu, Restrepo, and Robinson (2019) exclude from the control group continuing democracies and countries that transition out of democracy at time t. Countries that experience a transition to autocracy at time $t - 1$ or earlier are still used as controls. Moreover, also countries that democratize between time $t + 1$ and $t + h$ are included in the control group.

This perspective suggests testing robustness to stricter definitions of the control group. For example, consider Argentina, which democratized in 1973 and became a dictatorship again in 1976. The Acemoglu, Naidu, Restrepo, and Robinson (2019) approach means that Argentina contributes to the counterfactual for measuring the effect of (among others) the 1978 democratization of Spain. It seems natural to consider an alternative specification that excludes Argentina from the counterfactual for countries that (like Spain) democratize shortly after 1973–1976, reflecting the concern that the country might have experienced prolonged dynamic effects from the 1973–1976 transitions in and out of democracy. Moreover, in measuring the effect of the 1978 democratization of Spain on GDP growth in the subsequent 10 years, Acemoglu, Naidu, Restrepo, and Robinson (2019) allow Ecuador, which was a nondemocracy in 1977 and 1978 but democratized in 1979, to be part of the control group. It appears useful to test robustness to exclusion of countries that democratize between $t + 1$ and $t + h$ from the control group.

4.2.3 Results

Figure 8 displays the impulse response function from the estimation of the dynamic panel model of equation 25. This reproduces the baseline results of Acemoglu, Naidu, Restrepo,

[19] A different possible choice would have been to define the control group as $\Delta D_{i,t+k} = 0$ for $-K \leq k \leq h$. This would have meant allowing countries that have been continually a democracy from $t - K$ to $t + h$ in the control group, under the assumption that dynamic effects of democratization have stabilized for those countries, and therefore their democracy status affects the level but not the dynamics of output.

and Robinson (2019). The implied long-run effect of democracy on growth is 21 percent with a standard error of 7 percent.

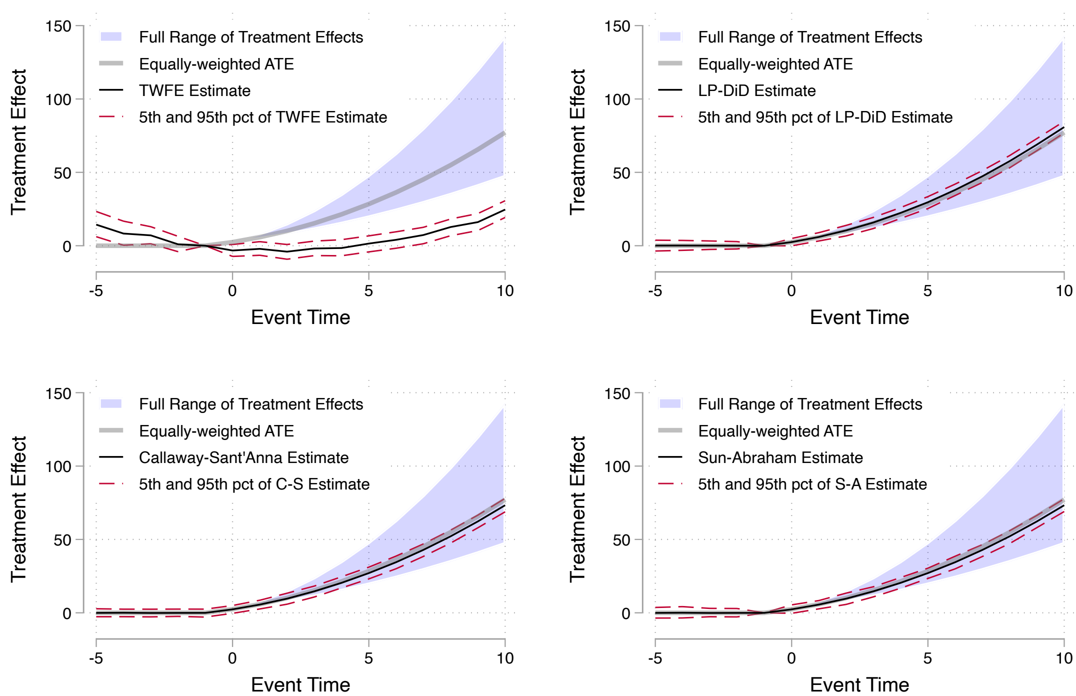
Figure 9 displays results from LP-DiD specifications (equation 26). We present four LP-DiD specifications: The first follows Acemoglu, Naidu, Restrepo, and Robinson (2019) in setting a time-window of just one period for defining clean controls ($K = 1$) and not constraining treatment status between t and $t + h$ in the control group; the other three apply the clean-control condition in equation 27, respectively with time-windows $K = 1, 20$ and 40.

Broadly speaking, the result of a positive effect of democracy on GDP per capita appears robust to stricter definitions of the control group. However, at longer time horizons (25 to 30 years after democratization), the effect declines more and is much more uncertain in the specifications with a stricter definition of the control group. Interestingly, the time-window K makes little difference in this application, while what makes some difference (at least at longer time horizons) is excluding from the control group countries that democratize between t and $t + h$. This difference emerges at long time horizons because with large h the number of countries that democratize between $t + 1$ and $t + h$ can become substantial, making the trade-off between a cleaner control group and statistical power more important.

5 Conclusion

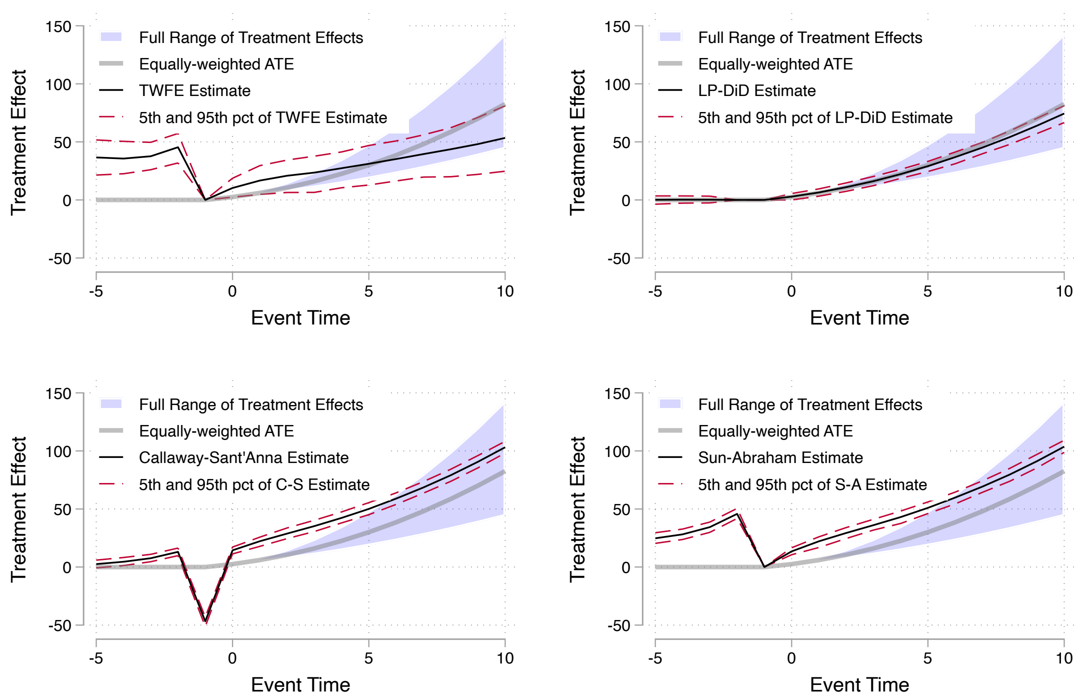
We propose a simple, transparent, easy and fast technique for difference-in-differences estimation with dynamic heterogeneous treatment effects. Our LP-DiD estimator has several advantages and provides an encompassing framework, which can be flexibly adapted to address a variety of settings. It does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

Figure 1: Results from Montecarlo simulation – Exogenous Treatment Scenario



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

Figure 2: Results from Montecarlo simulation – Endogenous Treatment Scenario



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

Figure 3: Banking deregulation in US States

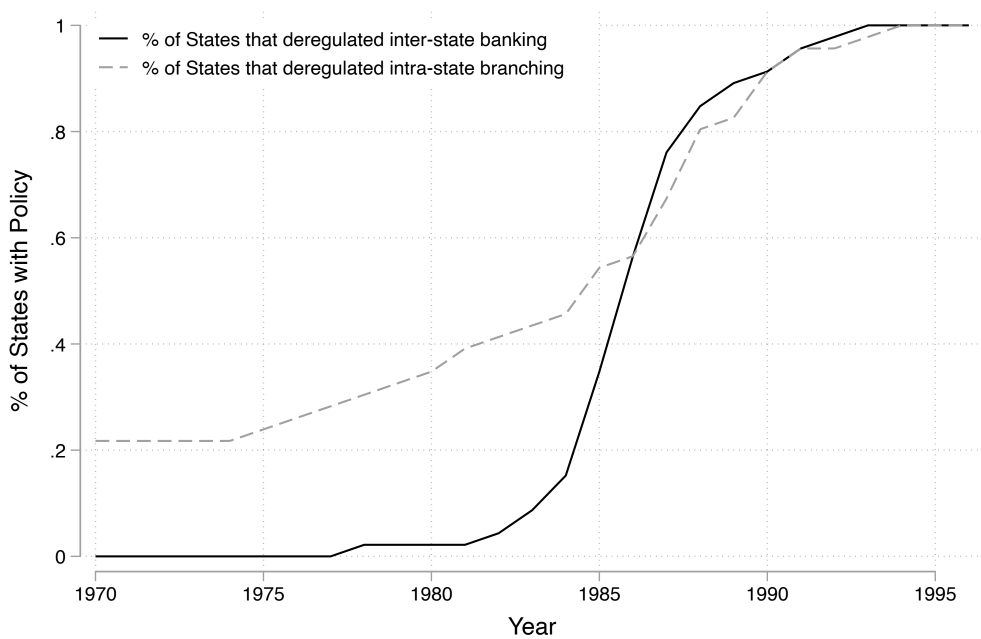
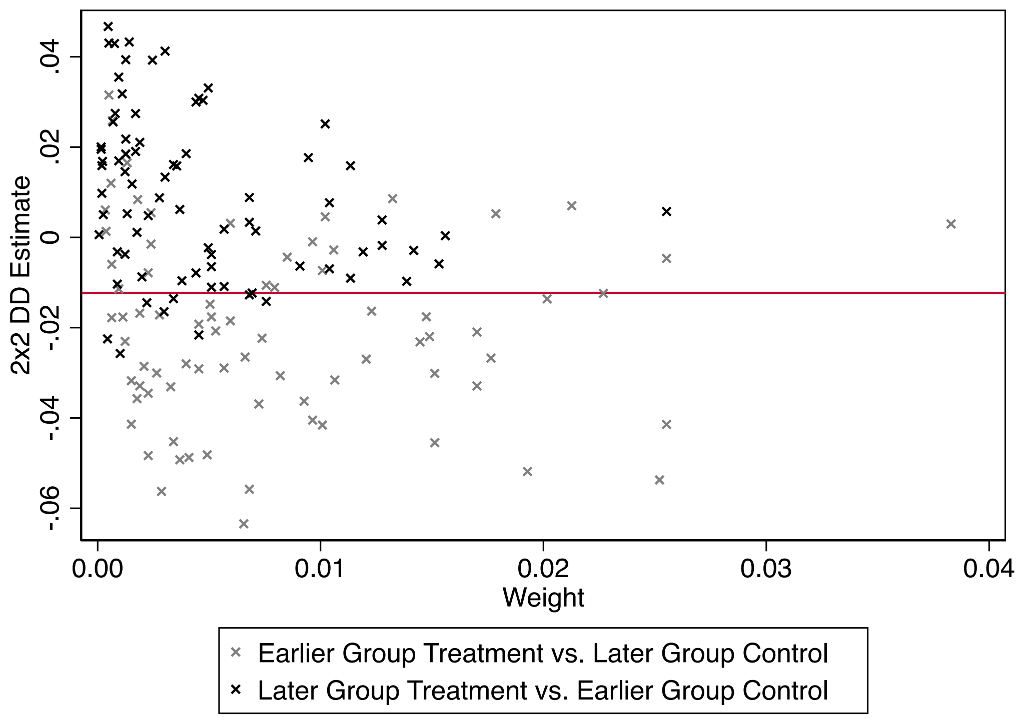


Figure 4: Goodman-Bacon (2021) decomposition diagnostic for the static TWFE specification of equation 21

(a) treatment: interstate banking deregulation



(b) treatment: intrastate branching deregulation

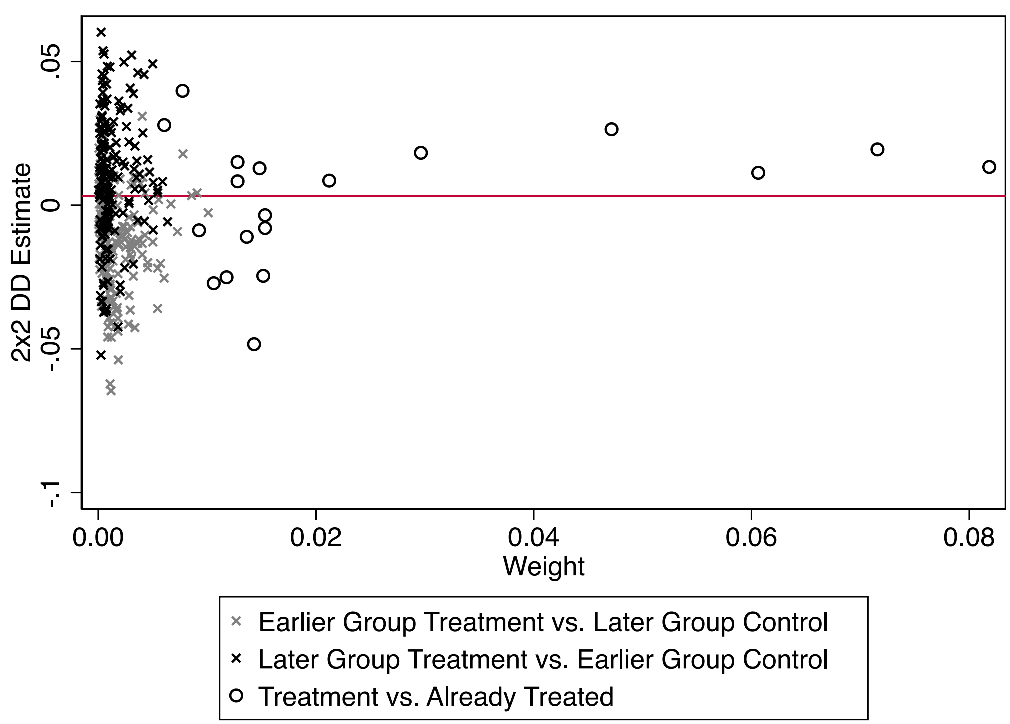


Figure 5: Effect of banking deregulation on the labor share: static TWFE estimates

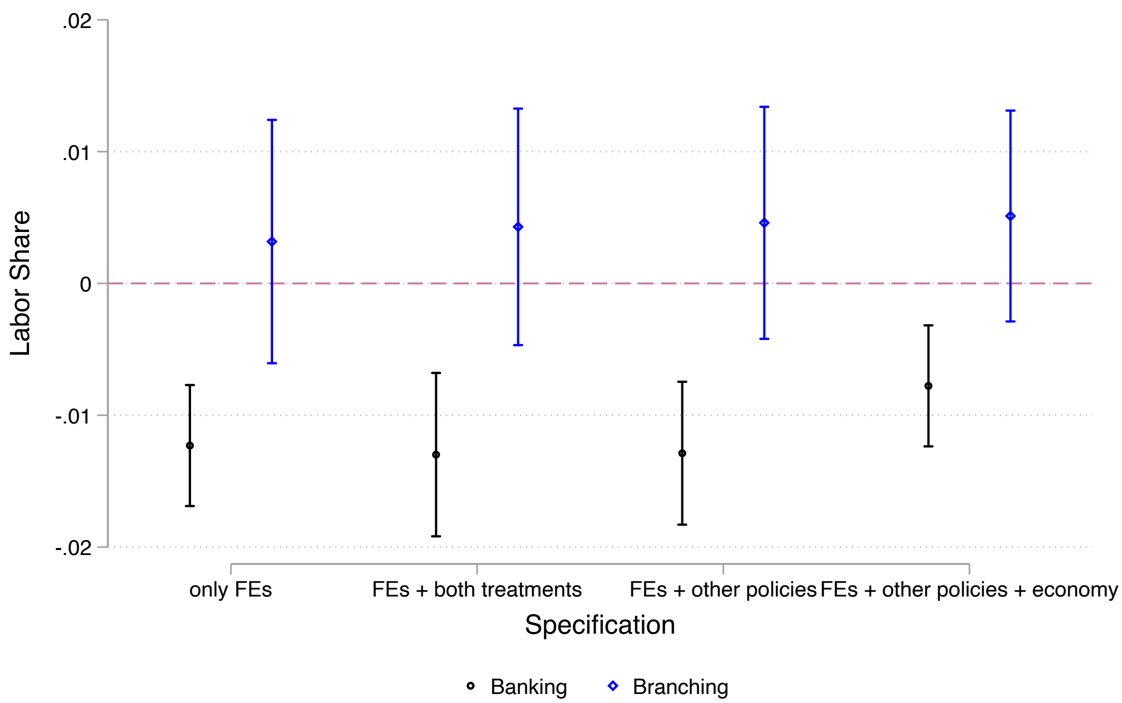


Figure 6: Effect of banking deregulation on the labor share: event-study TWFE estimates

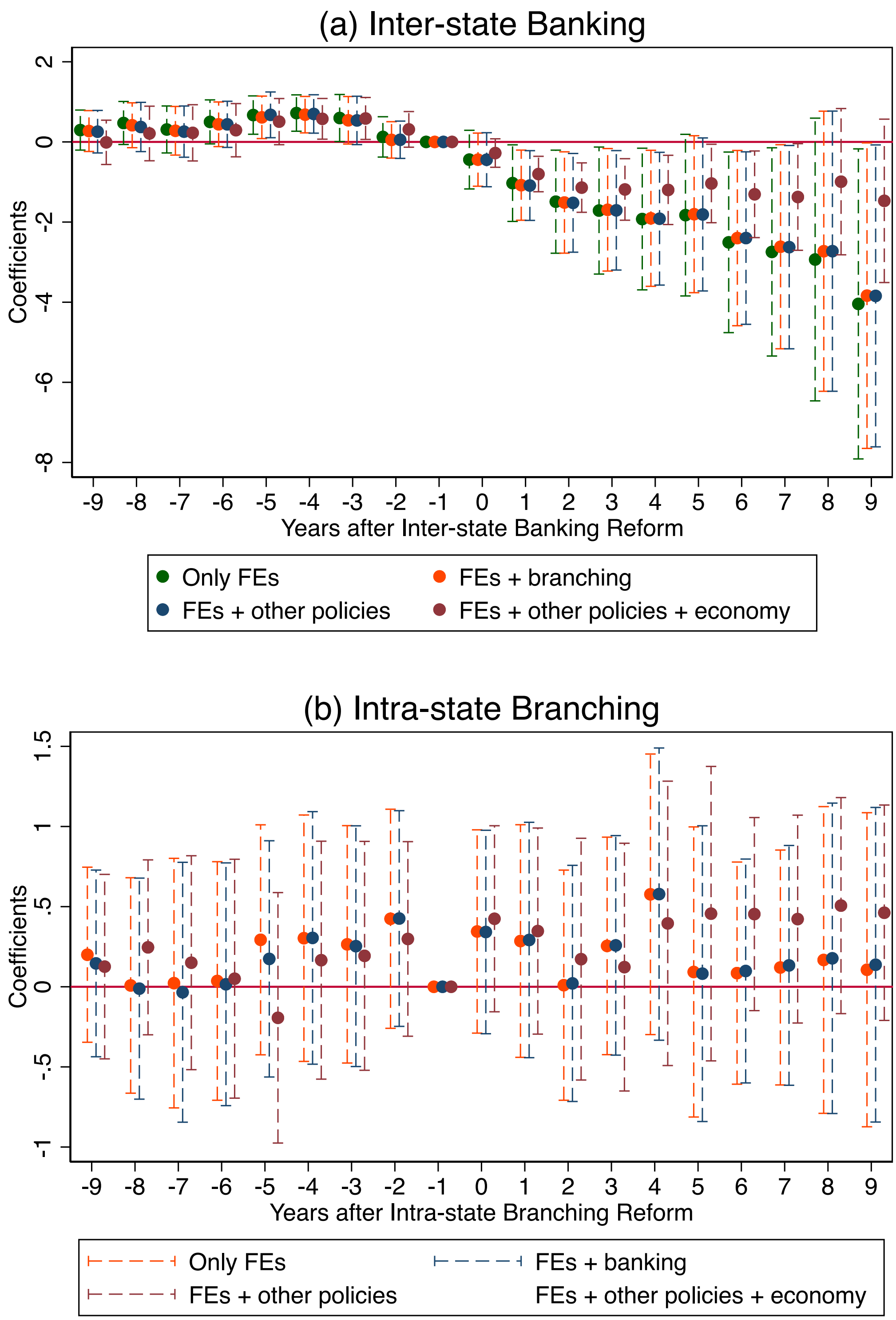
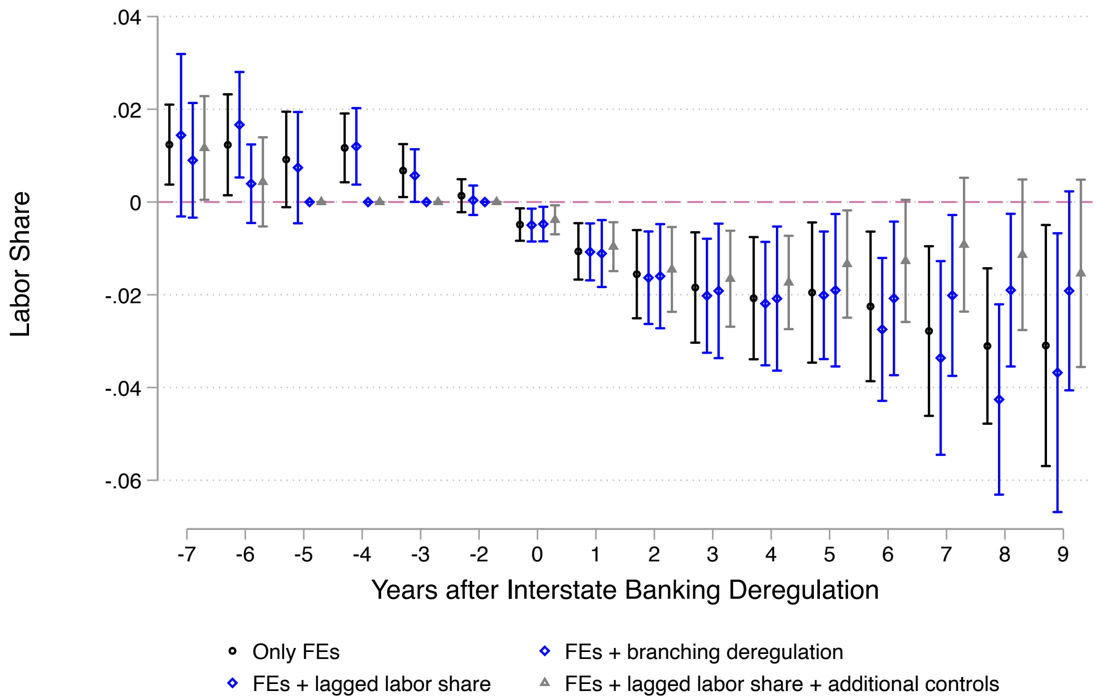


Figure 7: Effect of banking deregulation on the labor share: LP-DiD Estimates with clean controls

(a) effect of inter-state banking deregulation



(b) effect of intra-state branching deregulation

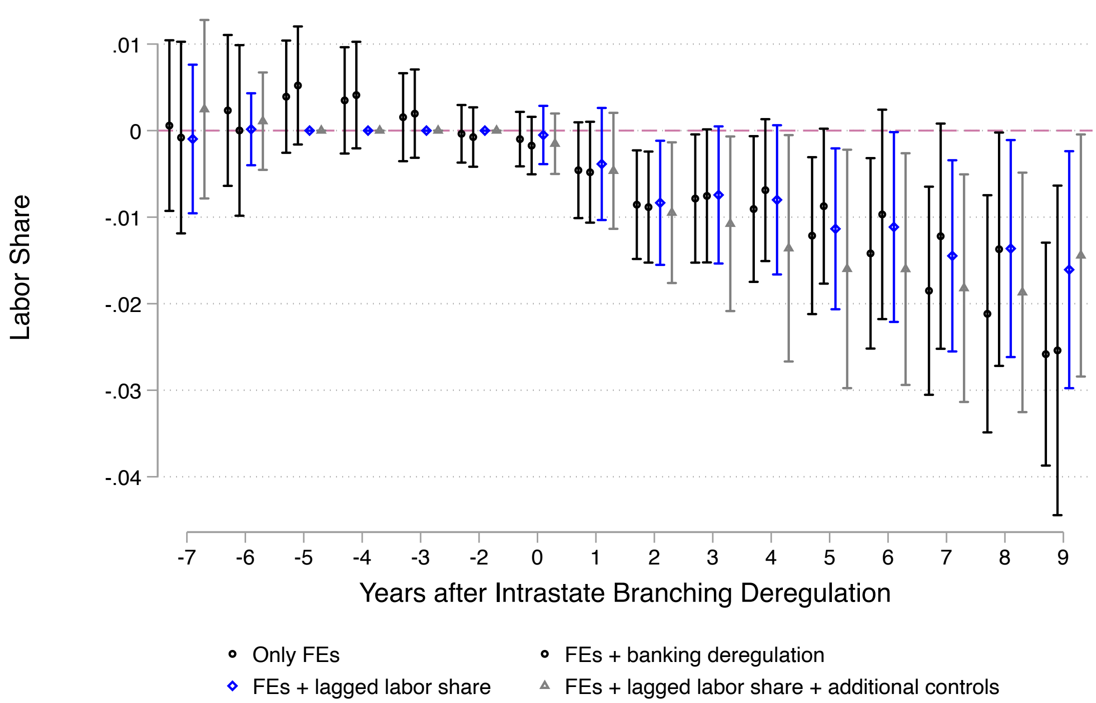


Figure 8: Effect of democracy on growth - dynamic panel estimates

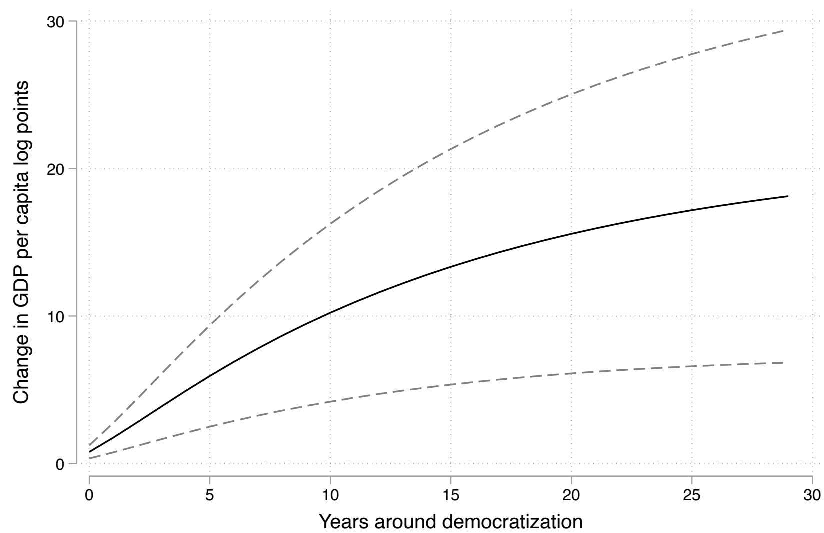


Figure 9: Effect of democracy on growth - LP-DiD estimates

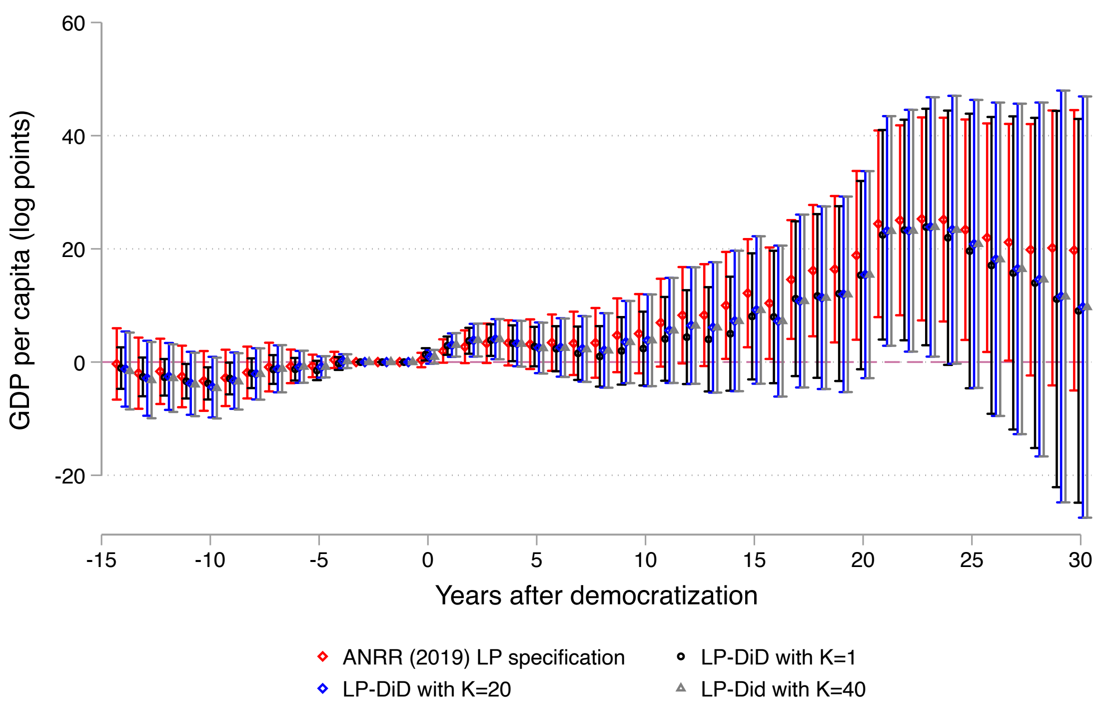


Table 1: Root Mean Squared Error (RMSE) – Exogenous Treatment Scenario

Event time	Event-Study TWFE	LP-DiD	Sun-Abraham	Callaway-Sant'Anna'
-5	15.34	2.21	2.24	1.54
-4	9.67	2.08	2.35	1.56
-3	7.85	1.76	1.7	1.52
-2	3.32	1.55	1.78	1.49
0	6.2	1.58	1.82	1.61
1	8.43	1.88	1.8	1.92
2	14.46	2.14	2.37	2.29
3	17.44	2.3	2.21	2.41
4	23.18	2.56	2.6	2.62
5	27.14	2.72	2.61	2.87
6	32.49	2.92	3.16	3.17
7	38.03	3.19	3.35	3.45
8	42.33	3.67	3.94	3.91
9	49.61	4.15	4.2	4.23
10	52.65	4.39	4.7	4.7

Notes: RMSE from 200 replications.

Table 2: Root Mean Squared Error (RMSE) – Endogenous Treatment Scenario

Event time	Event-Study TWFE	LP-DiD	Sun-Abraham	Callaway-Sant'Anna'
-5	40.29	2.15	24.94	3.28
-4	36.6	1.99	28.28	4.99
-3	38.27	1.73	34.47	7.86
-2	46.07	0	45.87	13.21
0	9.73	1.81	11.15	12.05
1	13.17	2.01	16.36	16.39
2	13.38	2.32	19.05	18.5
3	12.18	2.41	20.26	19.48
4	10.63	2.4	20.74	20.01
5	10.31	2.82	21.18	20.28
6	11.3	3.24	21.53	20.61
7	14.59	4.29	21.45	20.68
8	19.73	5.61	21.37	20.69
9	26.82	7.37	21.37	20.81
10	34.36	9.37	21.37	20.8

Notes: RMSE from 200 replications.

References

D. Acemoglu, S. Naidu, P. Restrepo, and J. A. Robinson. Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100, 2019. doi: 10.1086/700936.

J. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press, 2009.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation, 2021. URL https://arxiv.org/abs/2108.12419.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 2020.

D. Cengiz, A. Dube, A. Lindner, and B. Zipperer. The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics*, 134(3):1405–1454, 05 2019. doi: 10.1093/qje/qjz014.

C. de Chaisemartin and X. D'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, September 2020. doi: 10.1257/aer.20181169.

C. de Chaisemartin, X. D'Haultfœuille, F. Pasquier, and G. Vazquez-Bare. Difference-in-differences estimators for treatments continuously distributed at every period. *Available at SSRN*, 2022.

J. Gardner. Two-stage differences in differences. April 2021. URL https://jrgcmu.github.io/2sdd_current.pdf.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2021.03.014.

M. Hoyos. Tariffs and growth: Heterogeneous effects by economic structure. Unpublished job market paper, 2022.

Ò. Jordà. Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182, March 2005.

O. Jorda and A. M. Taylor. The time for austerity: Estimating the average treatment effect of fiscal policy. *The Economic Journal*, 126(590):219–255, 2016. doi: https://doi.org/10.1111/ecoj.12332.

A. Leblebicioğlu and A. Weinberger. Credit and the Labour Share: Evidence from US States. *The Economic Journal*, 130(630):1782–1816, 02 2020. doi: 10.1093/ej/ueaa025.

D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

P. H. Sant'Anna and J. Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: https://doi.org/10.1016/j.jeconom.2020.06.003.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.

Appendix

A Weights of the LP-DiD estimator

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

A.1 Baseline version without control variables

Assumptions about the DGP

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

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

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

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

LP-DiD specification

Consider the following LP-DiD specification with clean controls:

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

restricting the sample to observations that are either:

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

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

Derivation of the weights

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

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

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,

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

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

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

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

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

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

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

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

$$= \sum_{j=1}^{G} \sum_{i \in CCS_{j,h}} \frac{\widetilde{\Delta D}_{i,p_j}}{\sum_{j=1}^{G} \sum_{i \in CCS_{j,h}} \widetilde{\Delta D}^2_{i,p_j}} E\left(\tau_{i,p_j+h} D_{i,p_j}\right)$$

$$= \sum_{j=1}^{G} \sum_{i \in j} \frac{\widetilde{\Delta D}_{i,p_j}}{\sum_{j=1}^{G} \sum_{i \in j} \widetilde{\Delta D}^2_{i,p_j}} \tau_{i,p_j+h}$$

$$= \sum_{g \neq 0} \frac{N_g \widetilde{\Delta D}_{g,p_g}}{\sum_{g \neq 0} N_g \widetilde{\Delta D}^2_{g,p_g}} \tau_{g,p_g+h}$$

$$= \sum_{g \neq 0} \omega^{LP-DiD}_{g,h} \tau_g(h) \,,$$

where the weights are given by

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

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

A.2 Weights with control variables

What are the weights of the LP-DiD estimator in a more general specification that includes exogenous and pre-determined control variables? If covariates have a linear and homogenous effect on the outcome, and parallel trends holds conditional on covariates, it is possible to show that the weights assigned to each group-specific effect by the LP-DiD estimator are unchanged by the inclusion of exogenous or pre-determined covariates. In more general settings, the weights are proportional to the residuals of a regression of the treatment indicator on time effects and the covariates.

To explore the role of covariates, we now assume that no anticipation and parallel trends hold after conditioning on a set of observable exogenous or pre-determined covariates (Assumptions **??** and **??** in the main text).

A.2.1 Covariates with linear and homogeneous effects

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

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

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

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

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

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

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

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

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

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

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

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

Given the assumptions about the DGP, we have

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

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

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

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

A.2.2 More general setting

Now consider a more general setting, in which Assumptions **??** and **??** from 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

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

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

The weights are thus given by

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

As noted in the main text (Section 2.6.1), it is always possible to preserve the weights of the baseline LP-DiD estimator by employing semi-parametric propensity-score methods.