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

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

March 28, 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 projection (LP) estimators of the sort used in applied macroeconometrics. Our proposed LP-DiD estimator provides an overarching toolkit with several advantages. First, the method is clear, simple, easy to compute, and transparent and flexible in its handling of treated and control units. Second, it is quite general, including its ability to control for pre-treatment values of the outcome and of other covariates, as under conditional common trends. Third, the LP-DiD can nest other estimators, providing a framework that is not only rigorous but also encompassing. 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 many. 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

In applied microeconomics, and in quantitative social science more broadly, mimicking an experimental research design via differences-in-differences (DiD) has become a widely-used statistical technique for event studies and, with appropriate identification, 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 with 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. Some of the central matters of concern here have been the appropriate implementation of DiD in an expanded set of situations where the investigator studies multiple treatments (not a single wave), where the treatment effects are heterogeneous, or where the object of interest is the (multi-period) dynamic response of the outcome to treatment as, for example, in Callaway and Sant'Anna (2020); de Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2020); Goodman-Bacon (2021). What was once a seemingly simple tool of general application increasingly appears to need bespoke adjustments to suit a specific situation in more expanded settings.

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, here we will argue that the lens which best allows us to see the key equivalence is to re-frame the expanded set of DiD problems 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 expanded settings.

Admittedly, a natural skeptical reaction we can imagine hearing at this point is: why

do we need yet another expanded DiD technique? We have three main responses as to why a local projection DiD (LP-DiD) framework makes sense and offers some useful benefits. First, the method is direct, clear, simple, easy to code and compute, and is transparent and flexible in its handling of treated and control units. Second, it is not specific, but extremely general, including notably its ability to control for pre-treatment values of the outcome and of other covariates, e.g., making it straightforward to work with only conditional common trends and address situations with endogenous treatment or selection. Third, the LP-DiD can nest other estimators, providing a framework that is not only rigorous but also encompassing.

Our proposed LP-DiD approach employs local projections (Jordà, 2005) to estimate dynamic effects and a flexible 'clean control' condition (Cengiz, Dube, Lindner, and Zipperer, 2019) to accommodate the possibility of heterogeneous effects, which could otherwise introduce bias (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020; Sun and Abraham, 2020). As we will show, under the usual DiD assumptions, the LP-DiD estimator identifies a weighted average of potentially heterogeneous cohort-specific treatment effects, with weights that 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.

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. A short Section 2 provides a brief overview of the LP-DiD approach. A long Section 3 gives a full exposition of the statistical methods. In Section 4 we use simulations to compare the performance of our LP-DiD approach with other new methods in the recent literature. In Section 5 we apply the methods to two empirical applications in the form of a replication and robustness exercise, illustrating how established results can potentially be change when we employ the LP-DiD method. Section 6 concludes.

2 A preview of the LP-DiD approach

This section provides a preview of our proposed local projections estimator for difference-in-differences studies, which we dub LP-DiD.

The LP-DiD estimator combines the local projections approach for estimating dynamic effects with a 'clean control' condition to avoid the bias that fixed-effects estimators can suffer from when treatment adoption is staggered.

Important 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.

2.1 Estimating equation

Consider a setting in which N units are observed for T time periods. The researcher is interested in estimating the effect of a binary treatment. Different groups of units enter treatment at different points in time; treatment might or might not be absorbing. Assume that the no anticipation and parallel trends assumptions hold, at least conditional on a set of observable exogenous and pre-determined covariates.[1]

[1] We will make these assumptions more precise later in Section 3.

The LP-DiD estimator can be implemented through least squares regression in two equivalent ways: by imposing a simple sample restriction, or by including a set of interaction terms.

LP-DiD implemented via sample restriction LP-DiD can be implemented through the following estimating equation, with β^k the parameter of interest at horizon k:

$$\begin{aligned} y_{i,t+k} - y_{i,t-1} = \; & \beta^k \Delta D_{it} && \} \text{ treatment indicator} \\ & + \textstyle\sum_{p=1}^{P} \gamma^k_{0,p} \Delta y_{i,t-p} && \} \text{ outcome lags} \\ & + \textstyle\sum_{m=1}^{M} \sum_{p=0}^{P} \gamma^k_{m,p} \Delta x_{m,i,t-p} && \} \text{ covariates} \\ & + \delta^k_t && \} \text{ time effects} \\ & + e^k_{it} \; ; && \text{for } k = 0, \ldots, K \,, \end{aligned} \tag{1}$$

where the sample is restricted to observations that satisfy either of two conditions:

$$\begin{cases} \text{treatment} & \Delta D_{it} = 1 \,, \\ \text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k \,, \end{cases} \tag{2}$$

where t indexes time and i indexes units; y is the outcome of interest, D is a binary treatment indicator, x_m is the m-th covariate, δ^k_t is a common time-specific effect, P is the number of lags of the outcome and of other covariates that are included, M is the number of covariates, and K is the time horizon over which dynamic effects are estimated.

The parameter H in condition 2 determines the time-window employed to define admissible ('clean') control units. Specifically, a unit is considered an admissible control for units that enter treatment in period t only if it has experienced no change in treatment status for at least H periods before t (as well as for k periods after t). The time-window H should thus be selected by the researcher based on the application. It should be such that treatment effect dynamics can be assumed to have stabilized after H periods. With absorbing treatment, it is possible to use only not-yet treated units as controls ($H = \infty$), in which case the clean control group in 2 is simply defined by the condition $D_{i,t+k} = 0$.

LP-DiD implemented via interaction terms An alternative and equivalent way to implement LP-DiD is through the following estimating equation:

$$
\begin{aligned}
y_{i,t+k} - y_{i,t-1} = \; & \beta^k \Delta D_{it} && \} \text{ treatment indicator} \\
& + \theta^k UC_{i,t} && \} \text{ UC indicator} \\
& + \textstyle\sum_{p=1}^{P} \gamma^k_{0,p}(1 + \rho^k_{0,p} UC_{i,t})\, \Delta y_{i,t-p} && \} \text{ outcome lags} \times \text{UC indicator} \\
& + \textstyle\sum_{m=1}^{M} \sum_{p=0}^{P} \gamma^k_{m,p}(1 + \rho^k_{m,p} UC_{i,t}) \Delta x_{m,i,t-p} && \} \text{ covariates} \times \text{UC indicator} \\
& + \delta^k_t (1 + \phi^k_t UC_{i,t}) && \} \text{ time effects} \times \text{UC indicator} \\
& + e^k_{it} \; ; && \quad \text{for } k = 0, \ldots, K \,,
\end{aligned}
\tag{3}
$$

where the 'unclean control' indicator $UC_{i,t}$ is a binary variable equal to 1 if unit i fails to satisfy the clean control condition 2 at time t, and 0 if it satisfies it. The time-window H plays the same role as in the sample restriction specification.

With absorbing treatment, the unclean control indicator is equal to $UC = \sum_{j=-H(j\neq 0)}^{k} \Delta D_{i,t+j}$.

2.2 Main features

Under the relevant parallel trends assumption, we will show that the LP-DiD estimator identifies a *weighted average* of potentially *heterogeneous* cohort-specific treatment effects, with *weights that are always positive* and depend on treatment variance and subsample size.[2] Intuitively, the clean control condition (equation 2) prevents previously treated units, which might still be experiencing time-varying treatment effects, to be used as controls for newly-treated units. The clean control condition thus prevents the 'negative weights' problem of the two-way fixed-effects (TWFE) estimator highlighted by de Chaisemartin and D'Haultfœuille (2020) and other recent contributions.

We can also see that in basic form the LP-DiD estimator follows the OLS regression logic of assigning weights proportional to treatment variance and subsample size and this is generally efficient. But in some settings the researcher might prefer other weighting schemes. However, this is perfectly compatible with the LP-DiD estimator and is trivially easy to implement. This can be done simply by re-weighting observations as in weighted OLS regression. For example, one can obtain an equally-weighted ATET by simply reweighing observations from each group (i.e., treatment cohort) g by $1/(\omega_g^{LP-DiD}/N_g)$, where N_g is the number of treated observations belonging to group g and ω_g is the weight

[2] This will be shown formally in Section 3 and Appendix A, where we make explicit the parallel trends assumption and characterize formally the weights assigned by the LP-DiD estimator to each group-specific effect.

assigned by LP-DiD to group g, which we characterize explicitly in Section 3.

A key feature of the LP-DiD approach is the ability to control for pre-treatment values of the outcome variable and of other covariates. This important feature is made possible by the structure of the local projection specification. That is, *unlike* a standard event-study TWFE specification or most alternative estimators proposed in the recent literature, the local projection specification implies that any lagged variable included in the estimating equation is measured before treatment. Of course, controlling for pre-treatment outcome dynamics (as well as any other exogenous or pre-determined covariate) will be appropriate in some applications with a few exceptions.[3] But what has to be said is that the LP-DiD estimator offers the virtue of flexibility in this respect: the researcher can decide whether to control for lagged outcomes and other covariates based on the application.

While the clean control condition outlined in equation 2 is generally appropriate for difference-in-differences settings in which multiple units receive a binary treatment at different points in time, the condition can be adapted flexibly based on the application. As mentioned above, if treatment is absorbing, the researcher can choose to use only not-yet treated units as controls, in which case the clean control set definition simplifies to $D_{i,t+k} = 0$. Another possible example is the continuous treatment setting. This could be dealt with by adapting the clean control condition to define clean controls as 'stayers' (or alternatively 'quasi-stayers'), in the terminology of de Chaisemartin, D'Haultfœuille, Pasquier, and Vazquez-Bare (2022). With several different treatments, one can consider units which enter one treatment but not the others as the treated units, and units who do not receive any treatment as the control group, following de Chaisemartin and D'Haultfoeuille (2022).

3 Estimation of treatment effects with LP-DiD

This section provides a formal discussion of our proposed LP-DiD approach. We consider a binary treatment with staggered adoption: multiple units can receive treatment at different points in time. We start from a simple setting with dynamic but homogeneous effects (Section 3.1), and then allow for heterogeneous effects across units (Section 3.2). We show that, under heterogeneous treatment effects, the LP-DiD estimator of equations 1 and 2 (or, equivalently, equation 3) produces a weighted average of the various group-

[3] For example, Chabé-Ferret (2017) shows that in some circumstances conditioning on pre-treatment outcomes in a difference-in-differences study can be problematic. On the issue of whether to control for pre-treatment outcomes or not, see also REFERENCES.

specific effects, with weights that are always positive and depend on subsample size and treatment variance.

3.1 Dynamic but homogeneous treatment effects

We introduce the main ideas of our approach by building from a simple setting. A sample of data on an outcome y_{it} is available for $i = 1, \ldots, N$ units observed across $t = 1, \ldots, T$ periods. All units are untreated in the first period. They may then receive a binary treatment, measured by the indicator D_{it}, equal to 1 if unit i in period t is treated or previously treated, and 0 otherwise. Treatment is an absorbing state: once a unit enters treatment, it remains treated in all subsequent periods. The treatment indicator D_{it} is thus a step function that jumps permanently from 0 to 1 when a unit receives treatment. Hence, the variable $\Delta D_{it} = D_{it} - D_{it-1}$ takes the value of 1 when the unit first receives treatment, and is zero everywhere else. It also follows that $D_{it} = \sum_{s=1}^{t} \Delta D_{is}$.

Some units may never be treated during the observed $t = 1, \ldots, T$ periods in the sample. Further, we assume that the effect of treatment is homogenous across units, an assumption that we will relax later on.

The effect of the treatment may manifest over several periods. Moreover, the effect of the treatment could potentially differ from one period to the next. That is, treatment induces a dynamic response on the outcome.[4]

We denote $y_{i,t}(j)$ for $j = 0, 1$ the potential outcomes without and with treatment. We often omit the index i when we discuss the random variable itself as opposed to the sample observation. Thus $y_t(j)$ for $j = 0, 1$ are latent, unobservable random variables whose relation to the observed y_{it} is given by the mixture $y_{it} = y_{it}(0)(1 - D_{it}) + y_{it}(1)D_{it}$.

We assume that the data-generating process (DGP) conforms to the assumptions of no anticipation and parallel trends, which are needed to justify a difference-in-differences (DiD) approach. Formally, we assume:

Assumption 1. No anticipation
Suppose treatment is administered at time s. Then,

$$E[y_t(1) - y_t(0) | D_t; D_{t-1}, \ldots, D_{t-J}] = 0 \; ; \quad \text{for } t < s \, ,$$

where the lag index runs all the way to $J = t - 2$.

[4]This dynamic response is related to the concept of an *impulse response* in macroeconometrics. Indeed, the two ideas are equivalent when treatment is a one-off occurrence. However, in macroeconomics it is more common for treatment to be administered over several periods and with varying doses, a scenario that presents its own complications, as we shall see (Section 3.2.4).

The statement of the assumption differs slightly from typical statements found in the literature in that we condition on the past history of previous treatments. The assumption essentially says that, conditional on the history of previous treatments, units do not change their behavior before treatment in anticipation of such treatment.[5]

Assumption 2. Common trends
Also known as the parallel trends assumption, it can be stated in our setting as:

$$E[y_t(0) - y_1(0)|D_t; D_{t-1}, \ldots, D_{t-J}] = E[y_t(0) - y_1(0)|D_{t-1}, \ldots, D_{t-J}]; \quad \text{for } t = 2, \ldots, T,$$

again with $J = t - 2$.

3.1.1 Conventional static fixed-effects estimator

Under these assumptions, it may seem natural to estimate the average treatment effect on the treated, that is, $\tau \equiv E(y_t(1) - y_t(0)|D_t = 1)$, from the following two-way fixed effects (TWFE) regression:

$$y_{it} = c_i + \alpha_t + \beta^{TWFE} D_{it} + u_{it}, \tag{4}$$

where $\hat{\beta}^{TWFE}$ might seem a natural estimate of τ.

However, as highlighted by several recent contributions, the TWFE regression specification of equation 4 can suffer from serious bias in the staggered treatment setting when treatment effects are dynamic.[6] The bias comes from the fact that previously treated units are effectively used as controls for newly treated units.

The intuition is as follows. Since previously treated units might still be experiencing a delayed dynamic response to treatment, these treatment effect dynamics are effectively subtracted from the 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).

[5] A common way of relaxing this assumption is to discard periods in a small neighborhood around treatment. We do not elaborate on that here.

[6] Heterogeneous effects, that we consider below in Section 3.2, would exacerbate the issue. Moreover, while as we will see dynamic but homogeneous effects can be effectively dealt with by using a event-study version of the TWFE regression with a sufficient number of lags of the treatment variable, heterogeneous effects can produce bias also in the event-study TWFE specification (Sun and Abraham, 2020).

3.1.2 Event-study two-way fixed-effects estimator

How should the fixed-effects regression specification of equation 4 be modified to address this source of bias?

It might seem natural to just control for these lagged effects and indeed Sun and Abraham (2020) show that if treatment effects are dynamic but homogeneous across units, inclusion of a sufficient number of lags of the treatment indicator would eliminate the bias. Intuitively, the lagged treatment indicators control for the lagged dynamic effects of previous treatments.[7] Therefore, under homogeneity, a event-study version of the TWFE estimator, that includes lags of the treatment indicators, is sufficient to obtain unbiased estimates.

Formally, given the usual potential outcomes expression $y_t = y_t(0) + D_t(y_t(1) - y_t(0))$, we can take expectations on both sides of the expression conditional on $D_{t-1}, \ldots, D_{t-J}$:

$$\begin{aligned} &E[y_t|D_t; D_{t-1}, \ldots, D_{t-J}] \\ &= E[y_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}] + D_t E[y_t(1) - y_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}]. \end{aligned}$$

Given that treatment is absorbing and treatment effects are homogeneous across units, the average treatment effect on the treated, conditional on treatment history, is given by:

$$\tau_t \equiv E[y_t(1) - y_t(0)|D_t = 1; D_{t-1}, \ldots, D_{t-J}] = \beta_0 + \sum_{j=1}^{J} \beta_j D_{t-j}, \tag{5}$$

where β_0 is the contemporaneous effect of treatment– that is, $\beta_0 = E[y_t(1) - y_t(0)|D_t = 1, D_{t-1} = ... = D_{t-J} = 0]$; and $\sum_{j=0}^{k} \beta_j$ is the effect of having been treated for k periods. By the Law of Iterated Expectations, $\tau \equiv E(y_t(1) - y_t(0)|D_t = 1) = E[\tau_t]$.

Based on Equation 5 and the law of iterated expectations, we have

$$E[y_t|D_t; D_{t-1}, \ldots, D_{t-J}] = E[y_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}] + D_t \tau_t \,.$$

Next, write $y_t(0) = y_1(0) + g_t(0)$ fot $t = 2, \ldots, T$. The current value of the potential outcome without treatment is the sum of the initial value $y_1(0)$ and the gain $g_t(0)$. Hence, without

[7] If treatment effects were instead heterogeneous across treatment cohorts, the coefficients on the lags of the treatment indicator could be contaminated by lagged effects from previous periods' treatments, leading to bias (Sun and Abraham, 2020). As we shall see, this is not an insurmountable problem. Indeed, as we show below, in that case it will just be necessary to extend the estimating equation with a set of appropriate interaction terms.

loss of generality, we can write:

$$E[y_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}]$$
$$= E[y_1(0)|D_t; D_{t-1}, \ldots, D_{t-J}] + E[g_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}].$$

Next, based on the common trends Assumption 2,

$$E[g_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}] = E[g_t(0)|D_{t-1}, \ldots, D_{t-J}],$$

therefore:

$$E[y_t|D_t; D_{t-1}, \ldots, D_{t-J}] = E[y_1(0)|D_t; D_{t-1}, \ldots, D_{t-J}] + E[g_t(0)|D_{t-1}, \ldots, D_{t-J}] + D_t\tau_t$$

At this stage, it is convenient to assume linearity (and to reintroduce the i subscript) and hence write:

$$E[y_{i1}(0)|D_{it}; D_{i,t-1}, \ldots, D_{i,t-J}] = c_i, \tag{6}$$
$$E[g_{it}(0)|D_{i,t-1}, \ldots, D_{i,t-J}] = \alpha_t, \tag{7}$$
$$\tau_{it} = \beta_0 + \sum_{j=1}^{J} \beta_j D_{i,t-j}. \tag{8}$$

Hence, the dynamic response to treatment can be estimated through the $\beta_j^{ES-TWFE}$ coefficients in the following event-study TWFE regression:[8]

$$y_{it} = c_i + \alpha_t + \sum_{j=0}^{J} \beta_j^{ES-TWFE} D_{i,t-j} + u_{it}, \tag{9}$$

where $\hat{\beta}(k)^{ES-TWFE} = \sum_{j=0}^{k} \hat{\beta}_j^{ES-TWFE}$ could then be constructed to provide an unbiased estimate of the average effect of having been treated for k periods.

Summarizing, under parallel trends and dynamic but homogeneous treatment effects, an event-study specification with lags of the treatment indicator is needed (and sufficient) to account for the lagged dynamic effect of previous treatments.

[8] Equations 6-8 imply $E[y_{it}] = c_i + \alpha_t + D_{it}[\beta_0 + \sum_{j=1}^{J} \beta_j D_{i,t-j}]$. Absorbing treatment implies that $D_{it}[\beta_0 + \sum_{j=1}^{J} \beta_j D_{i,t-j}] = \sum_{j=0}^{J} \beta_j D_{i,t-j}$.

3.1.3 A local projections estimator

Consider the following k-periods forward treatment effect:

$$\tau_t(k) \equiv E[y_{t+k}(1) - y_{t+k}(0) | \Delta D_t = 1]\,. \tag{10}$$

Given the assumption of homogeneous effects, this object depends only on k and not on t, so we can just write $\tau_t(k) = \tau(k)$. Moreover, for each given unit, homogeneity implies $y_{i,t+k}(1) - y_{i,t+k}(0) = \tau(k)$

Observed outcomes are thus given by

$$y_{t+k} = y_{t+k}(0) + \Delta D_t \tau(k) + \sum_{j=1}^{J} \Delta D_{t-j} \tau(k+j) + \sum_{j=1}^{k} \Delta D_{t+j} \tau(k-j)$$

Taking expectations conditional on D_t and $D_{t+k}, \ldots, D_{t+1}; D_{t-1}, \ldots, D_{t-J}$,

$$\begin{aligned} &E[y_{t+k} | D_t; D_{t+k}, \ldots, D_{t+1}; D_{t-1}, \ldots, D_{t-J}] = \\ &E[y_{t+k}(0) | D_t; D_{t+k}, \ldots, D_{t+1}; D_{t-1}, \ldots, D_{t-J}] + \\ &\Delta D_t \tau_t(k) + \textstyle\sum_{j=1}^{J} \Delta D_{t-j} \tau(k+j) + \sum_{j=1}^{k} \Delta D_{t+j} \tau(k-j)\,, \end{aligned} \tag{11}$$

Note that we have expanded the conditioning set to include values of the treatment indicator referring to future periods. This is to soak up variation in the outcome not related to the current treatment, but to a possible future treatment that might occur between $t+1$ and $t+k$, just as including lagged values is meant to soak up variation in the outcome due to previous treatments.

The parallel trends assumption implies that:

$$\begin{aligned} &E[y_{i,t+k}(0) | D_{it}; D_{i,t+k}, \ldots, D_{i,t+1}; D_{i,t-1}, \ldots, D_{t-J}] = \\ &E[y_{i,t+k}(0) | D_{i,t+k}, \ldots, D_{i,t+1}; D_{i,t-1}, \ldots, D_{t-J}] = c_i + \alpha_{t+k}. \end{aligned}$$

The natural estimating equation is then easily seen to be:

$$y_{i,t+k} = c_i + \alpha_{t+k} + \beta^{k\ LP\text{–}DiD} \Delta D_{i,t} + \sum_{j=0}^{J+k} \gamma_j^k \Delta D_{i,t+k-j} \times \mathbf{1}\{j \neq k\} + u_{i,t+k}; \qquad \text{for } k = 0, 1, \ldots, K\,, \tag{12}$$

where $\mathbf{1}\{j \neq k\} = 0$ if $j = k$. This additional notation plays no role but is convenient because it allows us to single out β^k, which is an estimator of $\tau_t(k)$, as the parameter of

interest.

This regression therefore controls not only for previously treated units, but also for units that are treated anytime between $t+1$ to $t+k$ and which could pollute estimation of $\tau(k)$. Based on this regression, $\beta^{k\ LP\text{–}DiD}$ provides a natural estimate of $\tau(k)$ under linearity. It is important to highlight that here we are assuming that treatment of one unit does not affect outcomes or treatment likelihood in other units.

As is well known from the literature on local projections, starting from Jordà (2005), the error term $u_{i,t+k}$ may very well be serially correlated (and possibly heterogeneous across units). However, depending on the setting, two-way clustering, or the Driscoll-Kraay heteroscedasticity and autocorrelation robust estimator will adjust standard errors for quite general variance structures.

In applications of the local projections estimator, it is common to employ the long difference $\Delta_k y_{it} \equiv y_{i,t+k} - y_{i,t-1}$ on the left side of the estimating equation, especially when y_{it} is expressed in logs. A reason is that $100 \times \beta^k$ can then be interpreted as an approximate percentage change in the outcome at time $t+k$ due to treatment at time t, facilitating interpretation of effect sizes. This transformation also has the advantage of mechanically removing unit-specific fixed effects.

From Equation 12 then, we the arrive at the 'long difference' expression for the LP-DiD estimator:

$$\Delta_k y_{it} = \delta_t^k + \beta^{k\ LP\text{–}DiD} \Delta D_{it} + \sum_{j=0}^{J+k} \theta_j^k \Delta D_{i,t+k-j} \times \mathbf{1}\{j \neq k\} + v_{it}^k \,; \qquad \text{for } k = 0, 1, \ldots, K, \quad (13)$$

where $\delta_t^k = \alpha_{t+k} - \alpha_{t-1}$ and $v_{it}^k = u_{it+k} - u_{it-1}$, while unit-fixed effects c_i cancel out as mentioned earlier. Given the homogeneous treatment effects assumption, the coefficients θ_j^k are sufficient to soak up the effect of any "unclean" controls on the long difference.

3.1.4 Adding covariates

Equation 12 and Equation 13 can be easily extended to include exogenous or predetermined covariates. In fact, in applications of the local projections estimator, it is common to include lags of the outcome variable since they soak up serially correlated variation due to unobservables.

To discuss the role of covariates, we start by stating conditional versions of the no anticipation and common trends assumptions, where $\boldsymbol{x}_t$ is a vector of predetermined and exogenous variables, possibly including lags of the outcome variable.

Assumption 3. Conditional no anticipation

As before, suppose treatment is administered at time s. Then:

$$E[y_t(1) - y_t(0)|D_t; D_{t-1}, \ldots, D_{t-J}, \boldsymbol{x}_t] = 0, \quad t < s.$$

Assumption 3 states that treatment at time t is unexpected, given the information available today, as it does not lead to behavior modification by units before it is administered.

Assumption 4. Conditional common trends

$$\begin{aligned} &E[y_t(0) - y_1(0)|D_t; D_{t-1}, \ldots, D_{t-J}, \boldsymbol{x}_t] = \\ &E[y_t(0) - y_1(0)|D_{t-1}, \ldots, D_{t-J}, \boldsymbol{x}_t]; \quad t = 2, \ldots, T. \end{aligned}$$

With the introduction of covariates, the LP-DiD estimating equation becomes:

$$\begin{aligned} &\Delta_k y_{it} = \delta_t^k + \beta^{k\ LP\text{–}DiD} \Delta D_{it} + \\ &\textstyle\sum_{j=0}^{J+k} \theta_j^k \Delta D_{i,t+k-j} \times \mathbf{1}\{j \neq k\} + \boldsymbol{\rho}_k \Delta \boldsymbol{x}_{it} + v_{it}^k; \quad \text{for } k = 0, 1, \ldots, K, \end{aligned} \tag{14}$$

where once again $\tau_t(k) = \tau(k) = \beta^k$ under linearity. Note that here the assumption of homogeneous effects allows us to omit interaction terms between the covariates and the treatment indicator.

In applications, researchers have to choose carefully which variables to include in $\boldsymbol{x}_t$, as the inclusion of 'bad controls' can introduce bias (eg Angrist and Pischke 2009, pp.64–68). It is generally a good idea to include lagged values of the covariates, i.e., dated $t-1$ or earlier, which are predetermined relative to treatment. Any control variable dated at time t must be fully exogenous. In general, this will seldom be the case in economic applications, when variables are jointly determined. Variables dated $t+1$ to $t+k$ are usually not recommended for the same reason. Their values could respond to treatment and future values of the outcome, which could easily bias estimation. For this reason, we stick to the notation $\boldsymbol{x}_t$.

When $\boldsymbol{x}_t$ includes the lag of the outcome, Nickell (1981) bias can arise from the presence of $y_{i,t-1}$ both as a regressor and in the error term, which is equal to $u_{i,t+k} - u_{i,t-1}$. 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 Alvarez and Arellano (2003) show. In applications in which 'Nickell bias' is a concern, the researcher can nevertheless correct for it by using a simple split-sample correction, following Chen et al. (2019).

3.2 Heterogeneous treatment effects

Up to this point we have assumed homogeneous treatment effects, i.e., that the treatment effect is the same across different treatment cohorts. In this section we extend the analysis to allow for heterogeneous treatment effects which may vary from one cohort to another. 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).

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. Group $g = 0$ is the never treated group. The group-specific k-periods forward average treatment effect for group g is denoted as $\tau_g(k)$. We denote the time period in which group g enters treatment as p_g.

3.2.1 Staggered absorbing treatment with dynamic heterogeneous treatment effects

In this setting units enter treatment at most once and treatment is permanent, but treatment effects can vary from one group to another and treatment can affect the outcome over several periods.

For simplicity, we start by abstracting from covariates, imposing assumptions 1 and 2. Under heterogeneous group-specific effects, equation 11 no longer holds, and instead we have:

$$\begin{aligned} E[y_{t+k}|D_t; D_{t+k}, \ldots, D_{t+1}; D_{t-1}, \ldots, D_{t-J}] &= E[y_{t+k}(0)|D_t; D_{t+k}, \ldots, D_{t+1}; D_{t-1}, \ldots, D_{t-J}] \\ &+ \textstyle\sum_{g=1}^{G} \Delta D_t \tau_g(k) \times \mathbf{1}\{t = p_g\} \\ &+ \textstyle\sum_{g=1}^{G} \sum_{j=1}^{J} \Delta D_{t-j} \tau_g(k+j) \times \mathbf{1}\{t = p_g + j\} \\ &+ \textstyle\sum_{g=1}^{G} \sum_{j=1}^{k} \Delta D_{t+j} \tau_g(k-j) \times \mathbf{1}\{t = p_g - j\}\,. \end{aligned} \tag{15}$$

where $\mathbf{1}\{t = p_g\}$ is an indicator equal to 1 if group g enters treatment at time t. Again, parallel trends implies that $E[y_{t+k}(0)]$ is independent of D and equal to $c_i + \alpha_{t+1}$.

The main insight from Equation 15 is that interaction terms between time indicators and the leads and lags of the treatment indicator are necessary and sufficient to 'clean' the estimated counterfactual from the bias coming from the influence of previously treated units.

Hence, an unbiased estimate of each group-specific effect can be obtained from an

estimate of β_g^k for $g = 1, ..., G$ in the following regression:[9]

$$\begin{aligned}\Delta_k y_{igt} &= \beta_g^k \Delta D_{gt} \\ &+ \delta_t^k \\ &+ \textstyle\sum_{j=0}^{J+k} \theta_j^k \Delta D_{g,t+k-j} \times \mathbf{1}\{j \neq k\} \\ &+ \textstyle\sum_{j=0}^{J+k} \delta_t^{k,j} \Delta D_{g,t+k-j} \times \mathbf{1}\{j \neq k\} \\ &+ v_{i,g,t}^k \; ; \\ &\qquad\qquad \text{for } k = 0, 1, \ldots, K \, .\end{aligned} \tag{17}$$

If one is not interested in each single group-specific effect, however, it is possible to estimate an average effect by estimating β_k^{LP-DiD} from the following regression

$$\begin{aligned}\Delta_k y_{it} &= \beta^{k\ LP-DiD} \Delta D_{it} \\ &+ \delta_t^k \\ &+ \textstyle\sum_{j=0}^{J+k} \theta_j^k \Delta D_{it+k-j} \times \mathbf{1}\{j \neq k\} \\ &+ \textstyle\sum_{j=0}^{J+k} \delta_t^{k,j} \Delta D_{it} \times \mathbf{1}\{j \neq k\} \\ &+ v_{it}^k \; ; \\ &\qquad\qquad \text{for } k = 0, 1, \ldots, K \, .\end{aligned} \tag{18}$$

For a given unit (or, equivalently, group), the leads and lags of ΔD are either all zero (if the unit either enters treatment at t or is not yet treated at $t + k$) or they are all zero except one (if the unit has entered treatment before period t or between $t + 1$ and $t + k$). Therefore, what the estimating equation 18 does is to estimate $\beta^{k\ LP-DiD}$ using only observations that are either entering treatment ($\Delta D_{it} = 1$) or 'clean controls' ($D_{i,t+k} = 0$). Observations that have been treated before t or between $t + 1$ and $t + k$ do not contribute to the no-treatment counterfactual change δ_t^k and do not affect the estimate of $\beta^{k\ LP-DiD}$.

The exact same $\beta^{k\ LP-DiD}$ coefficient can thus be estimated from the following simpler

[9] Sun and Abraham (2020) and Gardner (2021) offer two alternative regression based-solutions that modify the conventional TWFE specification to address heterogeneity and dynamics. Sun and Abraham (2020) extend the TWFE event-study specification of equation 9 with a set of interactions terms between the lags of the treatment indicator and group-specific indicators:

$$y_{it} = c_i + \alpha_t + \sum_{g=1}^{G} \sum_{j=0}^{J} \gamma_{g,j} \left(\mathbf{1}\{g_i = g\} \times D_{it-j} \right) + u_{it}. \tag{16}$$

Gardner (2021) proposes a two-stages approach: first one estimates the unit and period effects in the sample from observations with $D_{it} = 0$. The estimated unit and period effects from the first stage are then subtracted from the outcome, and this 'adjusted' outcome is regressed on D_{it}

estimating equation:

$$\begin{aligned} y_{i,t+k} - y_{i,t-1} = \quad & \beta^{k\ LP\text{–}DiD} \Delta D_{it} && \} \text{ treatment indicator} \\ & + \theta^k UC_{i,t} && \} \text{ UC indicator} \\ & + \delta_t^k (1 + \phi_t^k UC_{i,t}) && \} \text{ time effects} \times \text{UC indicator} \\ & + e_{it}^k \ ; && \text{for } k = 0, \ldots, K \,, \end{aligned} \tag{19}$$

where UC_{it} is a binary variable equal to one if a unit has either been previously treated ($D_{i,t-1} = 1$) or will be treated between $t+1$ and $t+k$ ($D_{it} = 0, D_{i,t+k} = 1$), and equal to zero if the unit is either untreated ($D_{i,t+k} = 0$) or newly treated ($\Delta D_{i,t} = 1$). In other words, UC_{it} is a binary variable that identifies 'unclean controls', that is, units with $\Delta D_{it} = 0$ but that have previously entered treatment or will enter it between $t+1$ and $t+k$.

Finally, the same estimate of $\beta_k^{LP\text{–}DiD}$ can also be obtained from the following equivalent regression:

$$\begin{aligned} y_{i,t+k} - y_{i,t-1} = \quad & \beta^{k\ LP\text{–}DiD} \Delta D_{it} && \} \text{ treatment indicator} \\ & + \delta_t^k && \} \text{ time effects} \\ & + e_{it}^k \ ; && \text{for } k = 0, \ldots, K \,, \end{aligned} \tag{20}$$

restricting the sample to observations that are either:

$$\begin{cases} \text{treatment} & \Delta D_{it} = 1 \,, \\ \text{clean control} & D_{i,t+k} = 0 \end{cases} \tag{21}$$

3.2.2 What does LP-DiD identify?

Under the common trends assumption, each β_g^k from equation 17 provides an unbiased estimate of the treatment effect for group g at horizon k. Furthermore, $\beta^{k\ LP\text{–}DiD}$ from equations 20-21 (or, equivalently, equations 18 or 19) provides some weighted average of these group-specific effects.

Indeed, the Frisch-Waugh-Lovell theorem implies that the LP-DiD estimator $\beta^{k\ LP\text{–}DiD}$ is 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 detailed formal derivation is in Appendix A.

First, we need to introduce some further definitions. Recall that we denoted the time period in which group g enters treatment as p_g. For each treatment group $g > 0$, define the clean control sample (CCS) for group g at time horizon k (denoted as $CCS_{g,k}$) as the set of observations for time $t = p_g$ that satisfy condition 21. Therefore $CCS_{g,k}$ includes the

observations at time p_g for all units that either enter treatment at p_g or are still untreated at $p_g + k$. In other words, $CCS_{g,k}$ includes observations at p_g for group g and its *clean controls*.

Under the assumption of parallel trends, the LP-DiD estimator $\beta^{k\ LP-DiD}$ identifies the following weighted-average effect:

$$E(\hat{\beta}^{k\ LP-DiD}) = \sum_{g \neq 0} \omega_{g,k}^{LP-DiD} \tau_g(k) \tag{22}$$

where $\tau_g(k)$ is the ATET for treatment-cohort g at the the k-periods horizon.

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

$$\omega_{g,k}^{LP-DiD} = \frac{N_{CCS_{g,k}}[n_{gk}(n_{c,g,k})]}{\sum_{g \neq 0} N_{CCS_{g,k}}[n_{g,k}(n_{c,g,k})]}, \tag{23}$$

where $N_{CCS_{g,k}}$ is the number of observations in the clean control sample for group g at time-horizon k; $n_{g,k} = N_g / N_{CCS_{g,k}}$ is the share of treated units in the $CCS_{g,k}$ subsample; and $n_{c,g,k} = N_{c,g,k} / N_{CCS_{g,k}}$ is the share of control units in the $CCS_{g,k}$ subsample. See Appendix A for the formal derivation of these weights.

3.2.3 Introducing covariates under heterogeneous effects

The introduction of covariates, under the conditional version of the no-anticipation and parallel trends assumptions (Assumptions 3 and 4) means that Equation 19 has to be expanded once more. Using the same logic of the Kitagawa-Oaxaca-Blinder decomposition, treatment can affect how observables influence the outcome, so that the baseline regression now becomes:[10]

$$\begin{aligned} y_{i,t+k} - y_{i,t-1} = \quad & \beta_k^{LP-DiD} \Delta D_{it} && \} \text{ treatment indicator} \\ & + \theta^k UC_{i,t} && \} \text{ UC indicator} \\ & + \boldsymbol{\rho}_k \Delta \boldsymbol{x}_{it} && \} \text{ covariates} \\ & + \boldsymbol{\gamma}_k (\Delta \boldsymbol{x}_{it} UC_{it}) && \} \text{ covariates} \times \text{UC indicator} \\ & + \delta_t^k (1 + \phi_t^k UC_{i,t}) && \} \text{ time effects} \times \text{UC indicator} \\ & + e_{it}^k \ ; && \text{for } k = 0, \ldots, K. \end{aligned} \tag{24}$$

[10] We do not directly investigate the Kitagawa-Oaxaca-Blinder decomposition based on Equation 14 since it has been described in great detail in the review article by Fortin, Lemieux, and Firpo (2011) and for local projections in Cloyne, Jorda, and Taylor (2020).

Or, in the (equivalent) version with the 'clean controls' sample restriction,

$$\begin{aligned} y_{i,t+k} - y_{i,t-1} = \; & \beta_k^{LP-DiD} \Delta D_{it} && \} \text{ treatment indicator} \\ & + \boldsymbol{\rho}_k \Delta \boldsymbol{x}_{it} && \} \text{ covariates} \\ & + \delta_t^k && \} \text{ time effects} \\ & + e_{it}^k; && \text{for } k = 0, \ldots, K, \end{aligned} \tag{25}$$

restricting the sample to observations that respect condition 21.

Of course, the same remarks made in Section 3.1.4 about the choice of control variables, their timing, and the possibility of (and possible remedies to) Nickell bias when outcome lags are included, apply here too.

Appendix A shows that, under linearity assumptions, the inclusion of covariates does not change the weights of the LP-DiD estimator, which remain the ones of equation 23. In more general settings, the weights are proportional to the residuals of a regression of the treatment indicator on time effects and the covariates, as detailed in Appendix A.

3.2.4 Repeated treatment

In many settings, treatment is not absorbing: units can enter and exit treatment multiple times. Although we will not investigate this case in much detail here, it is useful to briefly discuss how this situation complicates the analysis.

Our key point here is 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 in settings in which treatment can turn on and off. By appropriately modifying the 'clean control' condition (equation 21), the researcher can implement the approach that is considered most appropriate to the specific application.

First, consider a simple setting, in which the parallel trends assumption holds unconditionally (Assumption 2) and treatment assignment is not serially correlated. Specifically, suppose that $p_{it} \equiv E(D_{it}) \perp D_{it-1}, \ldots, D_{it-J}$, where p_{it} is the propensity score. Receiving treatment at time t does not alter the probability of receiving treatment at times $t+1, ..., T$.

In this setting one can recover, for example, the effect of entering treatment for the first time and staying treated, relative to a counterfactual of remaining untreated, by using the LP-DiD specification of Equation 20 and defining the clean control condition as

$$\begin{cases} \text{treatment} & D_{it} = D_{i,t+j} = ... = D_{i,t+k} = 1; \quad D_{i,t-1} = ... = D_{i,1} = 0, \\ \text{clean control} & D_{i,t+k} = ... = D_{it} = D_{i,t-1} = ... = D_{i,1} = 0 \end{cases} \tag{26}$$

In some settings there might not be enough units that have never received treatment

in each period t, making it impossible to implement condition 26. However, it might be possible to assume that treatment effect dynamics stabilize after H periods (formally, $\tau_{it}(H) = \tau_{it}(H+j)$ for $j \geq 0$) and employ the following clean control condition:

$$\begin{cases} \text{treatment} & \Delta D_{it} = 1, \quad \Delta D_{i,t+k} = ... = \Delta D_{t+1} = \Delta D_{i,t-1} = ... = \Delta D_{i,t-H} = 0 \\ \text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k \end{cases} \tag{27}$$

Now consider a more complicated setting, in which treatment assignment is serially correlated. Specifically, suppose that the propensity score is such that $p_t(\Omega_D, \boldsymbol{x}) \equiv P(D_{it}|D_{it-1}, \ldots, D_{it-J}; \boldsymbol{x}_t) \neq P(D_{it}|\boldsymbol{x}_t) = p_t(\boldsymbol{x})$.

This situation presents the researcher with two choices, neither one more correct than the other, as Alloza, Gonzalo, and Sanz (2019) highlight. If interest is in isolating the direct effect of a single treatment, without taking into account any 'indirect' effects through the probability of future treatments, then one must condition on such future treatments. If interest is in characterizing the overall effect of receiving treatment at t, inclusive of possible effects through inducing (or discouraging) future treatments, no such conditioning is necessary. The latter would be a more accurate description of what is likely to happen in practice, the former is a more accurate description of the treatment effect as generally conceived in the policy evaluation literature.

Formally, the treatment effect on the treated conditional on subsequent treatments is defined as:

$$\tau_{it}(k) \equiv E\left[y_{i,t+k}(1) - y_{i,t+k}(0)|\Delta D_{it} = 1; D_{i,t+1}, \ldots, D_{i,t+k}\right]; \text{ for } k = 0, \ldots, K, \tag{28}$$

This can be recovered by using a clean control condition like 26 or 27.

Instead, the treatment effect on the treated inclusive of indirect effects through the probability of future treatment is:

$$\mathcal{T}_{it}(k) \equiv E[y_{i,t+k}(1) - y_{i,t+k}(0)|\Delta D_{it} = 1] \qquad \text{for } k = 0, \ldots, K. \tag{29}$$

This can be recovered by using a clean control condition that does not condition on subsequent treatments, for example:

$$\begin{cases} \text{treatment} & \Delta D_{it} = 1, \quad \Delta D_{i,t-1} = ... = \Delta D_{i,t-H} = 0 \\ \text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, 0 \end{cases} \tag{30}$$

4 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 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-Abraham estimator; and (c) the Callaway-Sant'Anna 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-Abraham and Callaway-Sant'Anna 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_{oit} that a unit would experience if not treated is given by

$$Y_{oit} = \rho Y_{o,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}\,, \tag{31}$$

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:

$$\beta_{it} = \begin{cases} 0 & \text{if } t - \tau_i < 0 \\ \alpha_0(t-\tau_i) + \alpha_1(t-\tau_i)^2 + \alpha_2 \frac{(t-\tau_i)^2}{(\tau_i/\tau_1)^2} & \text{if } 0 \leq t - \tau_i < 20 \\ \alpha_0 20 + \alpha_1 20^2 + \alpha_2 \frac{20^2}{(\tau_i/\tau_1)^2} & \text{if } t - \tau_i \geq 20\,, \end{cases} \tag{32}$$

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

$$Y_{it} = Y_{oit} + \beta_{it} \tag{33}$$

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.
- The Sun-Abraham estimator.
- The Callaway-Sant'Anna 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.[11] Our 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 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.

5 Empirical Applications

To illustrate the use of the LP-DiD estimator in practice, we present two empirical applications. In the first, we use the LP-DiD estimator to estimate the effect of banking

[11] 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.

deregulation laws on the labor share in US States, replicating Leblebicioğlu and Weinberger (2020). In the second, we replicate the Acemoglu, Naidu, Restrepo, and Robinson (2019) country-panel study of the effect of democracy on economic growth.

5.1 Credit and the labor share

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.

5.1.1 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:

$$LS_{st} = \beta_{Bank} Bank_{st} + \beta_{Branch} Branch_{st} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st}\,, \tag{34}$$

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:

$$LS_{st} = \sum_{q=-9}^{9} \beta_{Bank,t+q} \Delta Bank_{s,t+q} + \sum_{q=-9}^{9} \beta_{Branch,t+q} \Delta Branch_{s,t+q} + \eta X_{st} + \alpha_s + \alpha_t + \epsilon_{st}\,. \tag{35}$$

5.1.2 Forbidden comparisons in the TWFE specifications

Given the staggered rollout of banking deregulation laws across US states, the TWFE specifications of equations 34 and 35 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 34 and 35 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 34 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 34, 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.

5.1.3 LP-DiD specification

In order to avoid the biases of the conventional TWFE specifications, and to allow for matching based on pre-treatment outcome dynamics, we re-estimate the effect of banking

deregulation laws using the following LP-DiD specification:[12]

$$LS_{s,t+k} - LS_{s,t-1} = \alpha_t + \beta_k^{LP-DiD} \Delta Bank_{s,t} + \sum_{m=1}^{M} \gamma_m^k \Delta LS_{s,t-m} + \sum_{m=1}^{M} \eta_m^k X_{s,t-m} + e_{s,t+k} \tag{36}$$

restricting the sample to observation that are either:

$$\begin{cases} \text{treatment} & \Delta Bank_{s,t} = 1 \\ \text{control} & Bank_{s,t+k} = 0 \end{cases} \tag{37}$$

5.1.4 Results

Figure 5 displays results from the static TWFE specification of equation 34, while Figure 6 displays results from the event-study TWFE specification of equation 35. 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.

5.2 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

[12] 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.

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.

5.2.1 Dynamic panel specifications

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

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{c,t-j} + \alpha_c + \delta_t + \epsilon_{ct}\,, \tag{38}$$

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. 38 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.

5.2.2 LP-DiD specifications

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

$$y_{c,t+k} - y_{c,t-1} = \beta^k \Delta D_{ct} + \delta_t^k + \sum_{j=1}^{p} \gamma_j^k y_{c,t-j} + \epsilon_{ct}\,. \tag{39}$$

restricting the sample to:

$$\begin{cases} \text{democratizations} & D_{it} = 1, D_{i,t-1} = 0 \\ \text{clean controls} & D_{i,t+h} = 0 \text{ for } -H \leq h \leq k\,. \end{cases} \tag{40}$$

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

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.[13]

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 39 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 40), and treatment status between $t+1$ and $t+k$ 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 $t+k$ 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

[13] A different possible choice would have been to define the control group as $\Delta D_{i,t+h} = 0$ for $-H \leq h \leq k$. This would have meant allowing countries that have been continually a democracy from $t-H$ to $t+k$ 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.

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 $t+k$ from the control group.

5.2.3 Results

Figure 8 displays the impulse response function from the estimation of the dynamic panel model of equation 38. This reproduces the baseline results of Acemoglu, Naidu, Restrepo, 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 39). 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 ($H = 1$) and not constraining treatment status between t and $t+k$ in the control group; the other three apply the clean-control condition in equation 40, respectively with time-windows $H = 1, 20$ and 40.

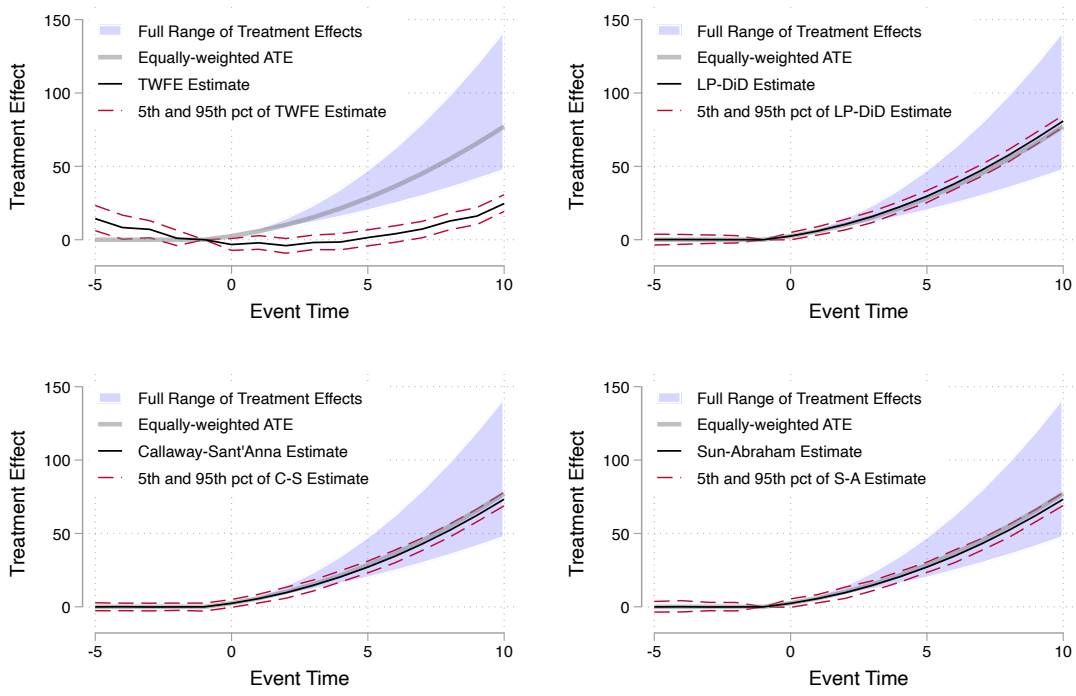
Broadly speaking, the result of a positive effect of democracy on GDP per capita appears robust to stricter definitions of the control group. However, at longer time horizons (25 to 30 years after democratization), the effect declines more and is much more uncertain in the specifications with a stricter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

6 Conclusion

We propose a simple, transparent, easy and fast technique for difference-in-differences estimation with dynamic heterogeneous treatment effects. Our LP-DiD estimator has several advantages and nest many existing estimators. 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

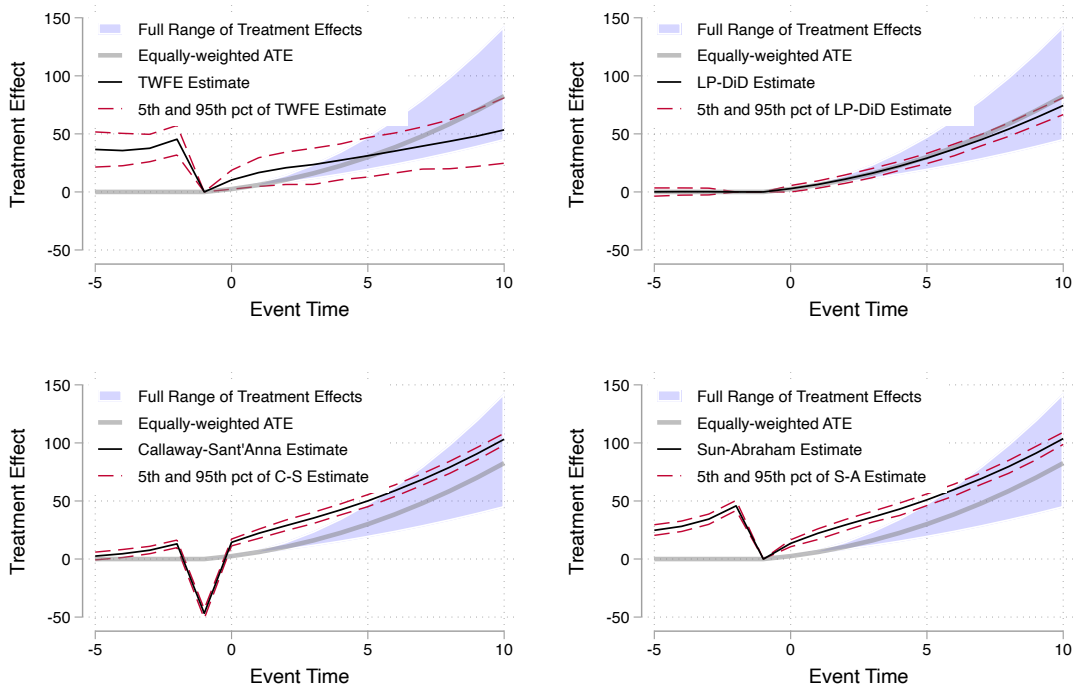
estimator and empirical exercises illustrate its use.

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

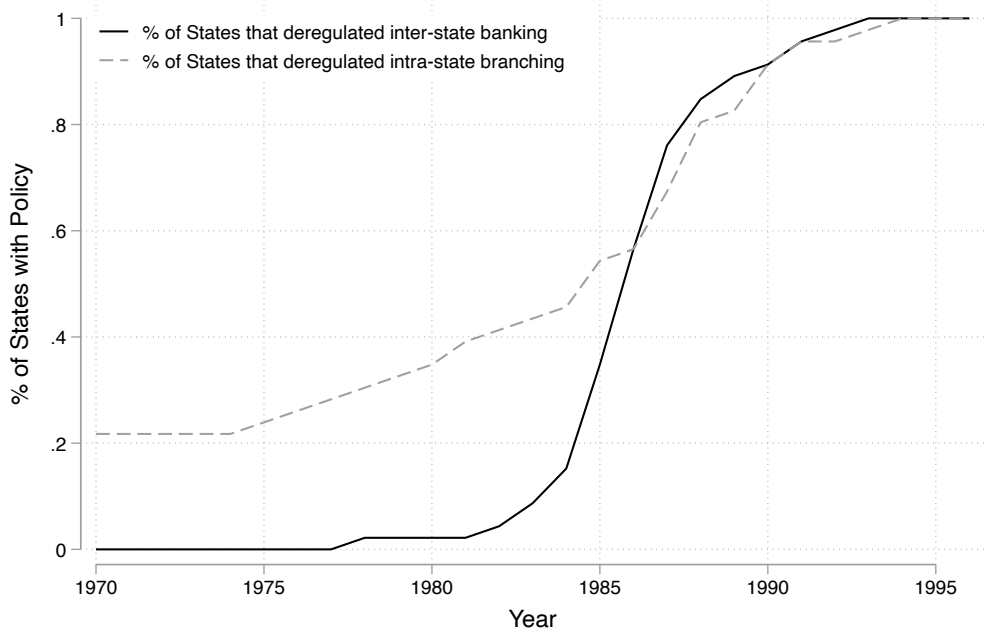
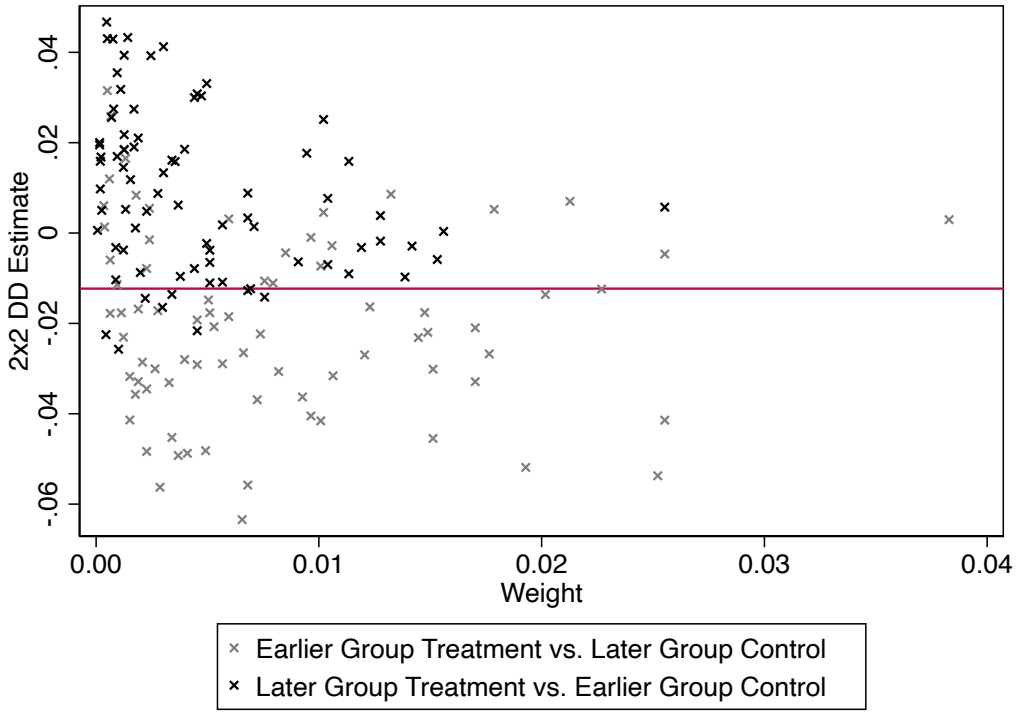


Figure 4: Goodman-Bacon (2021) decomposition diagnostic for the static TWFE specification of equation 34

(a) treatment: interstate banking deregulation



(b) treatment: intrastate branching deregulation

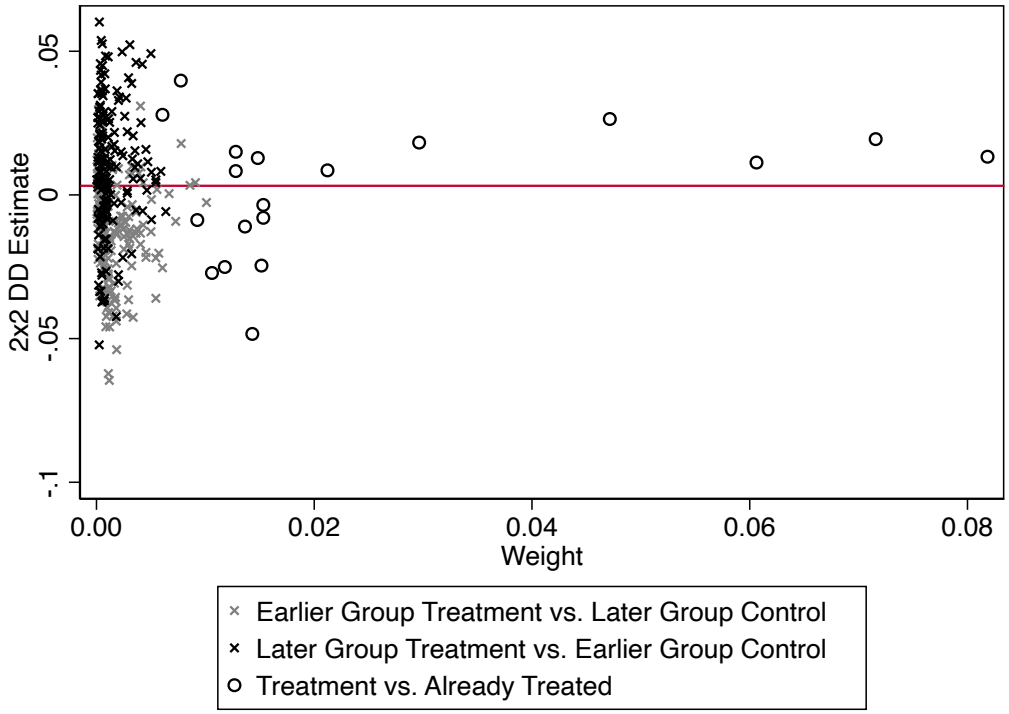


Figure 5: Effect of banking deregulation on the labor share: static TWFE estimates

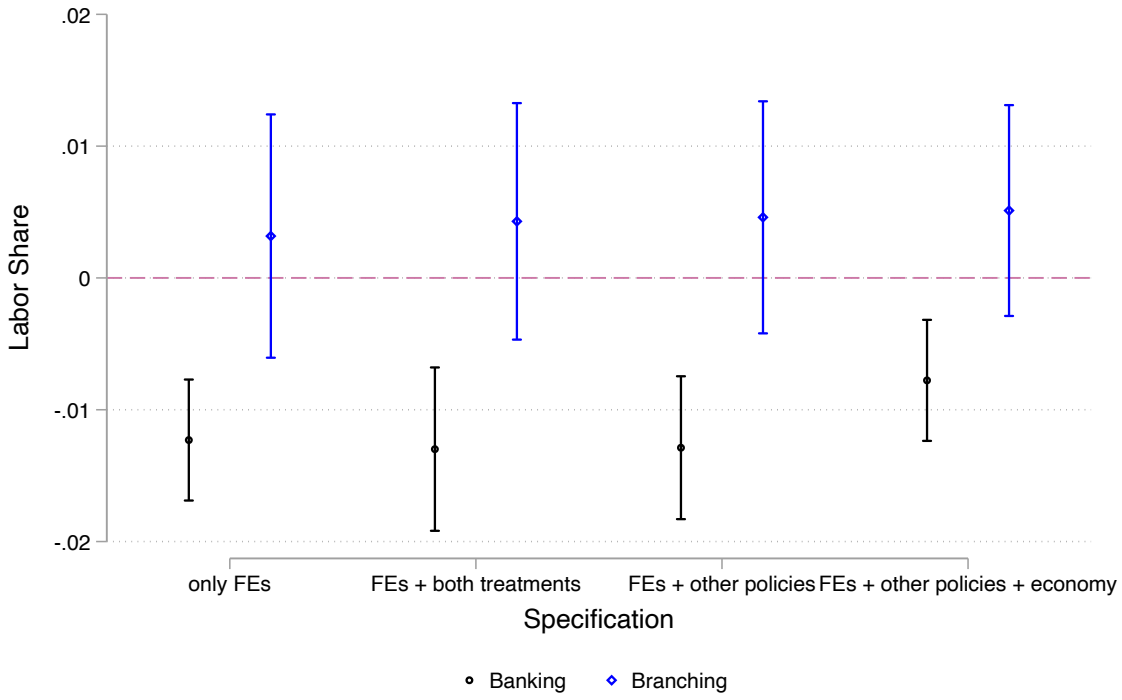
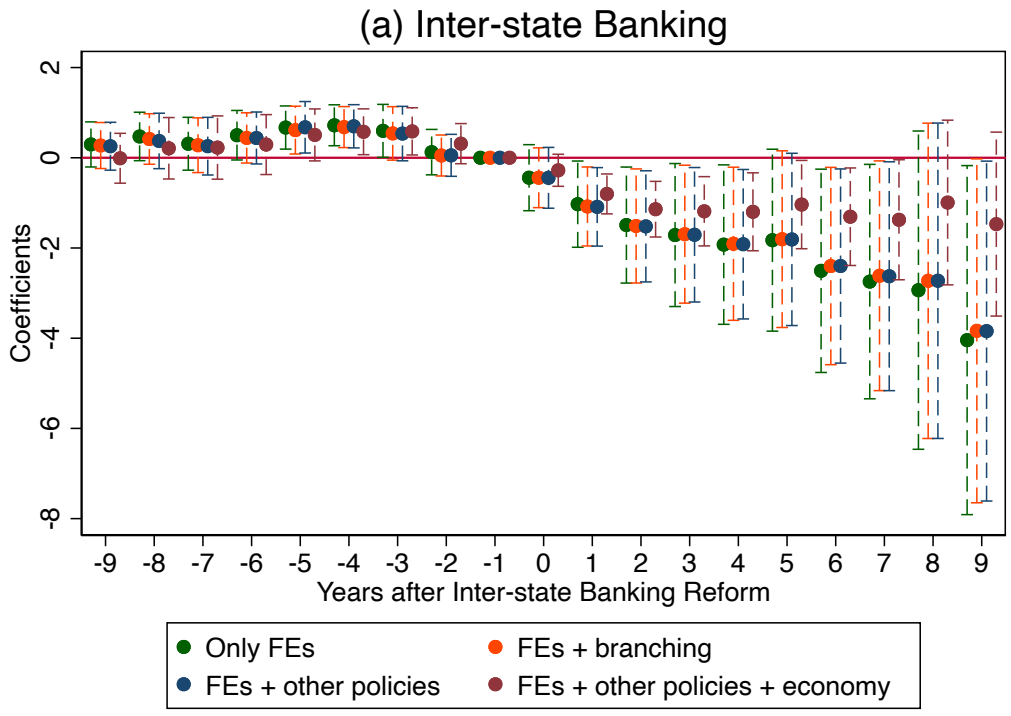


Figure 6: Effect of banking deregulation on the labor share: event-study TWFE estimates



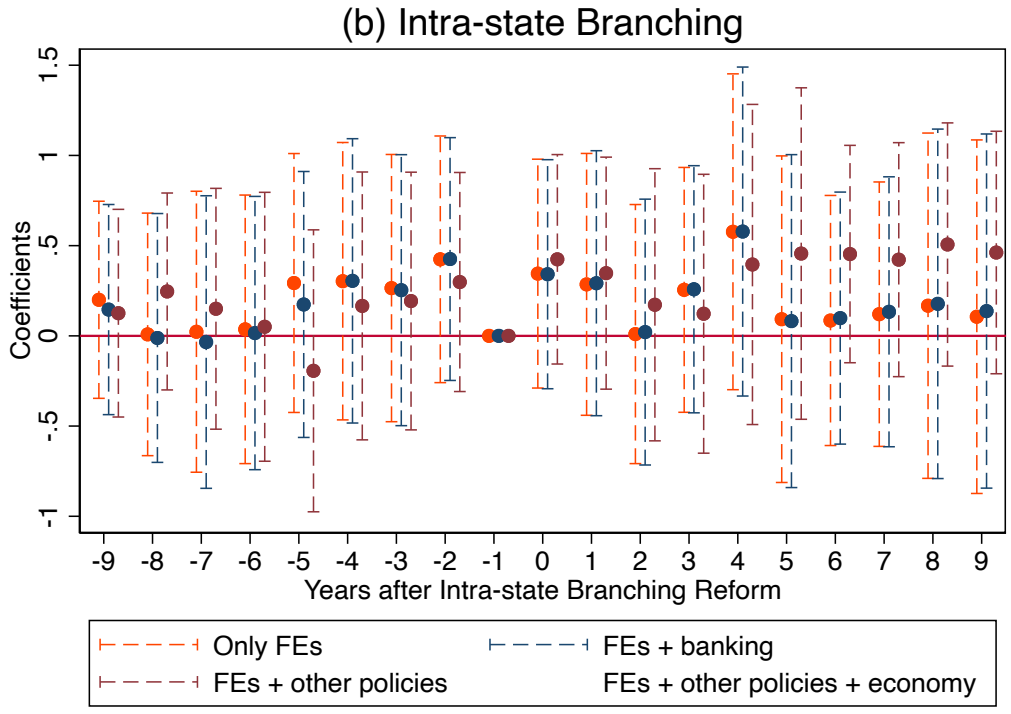
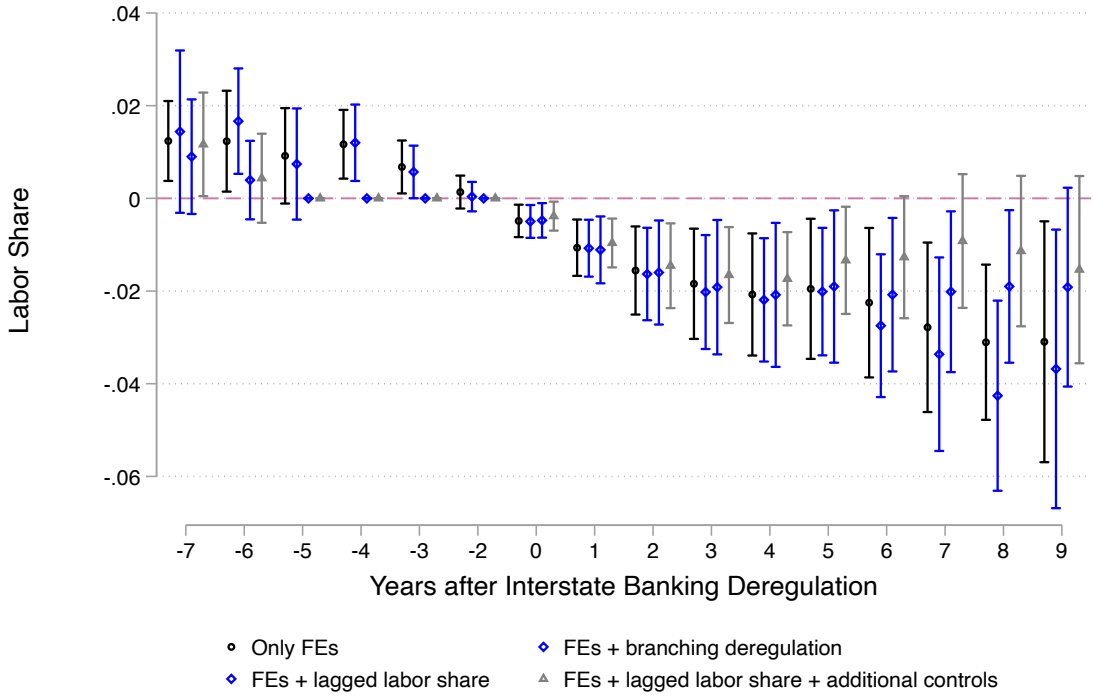


Figure 7: Effect of banking deregulation on the labor share: LP-DiD Estimates with clean controls

(a) effect of inter-state banking deregulation



(b) effect of intra-state branching deregulation

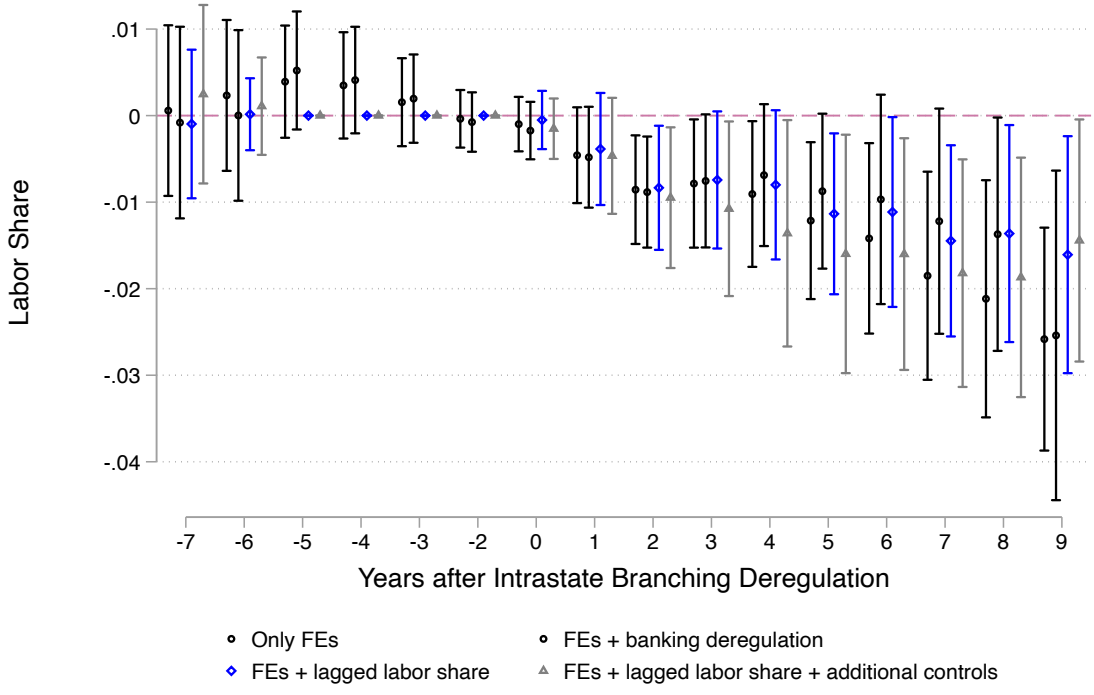


Figure 8: Effect of democracy on growth - dynamic panel estimates

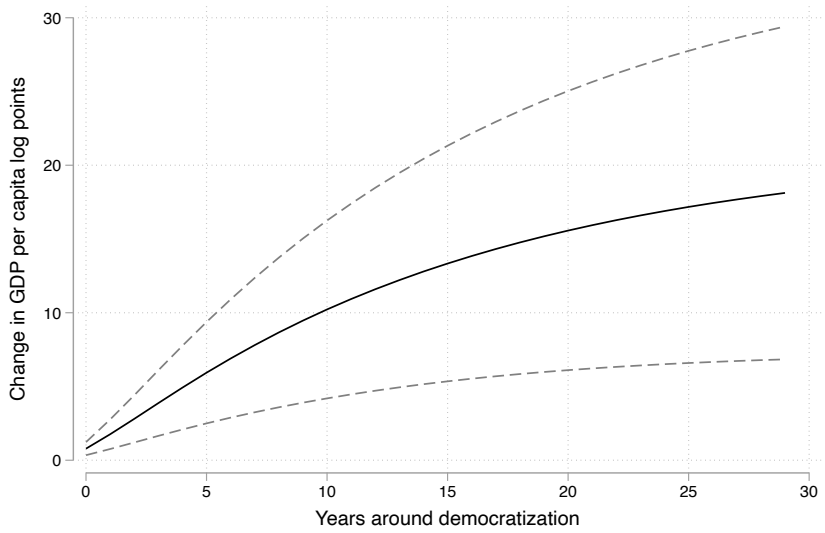


Figure 9: Effect of democracy on growth - LP-DiD estimates

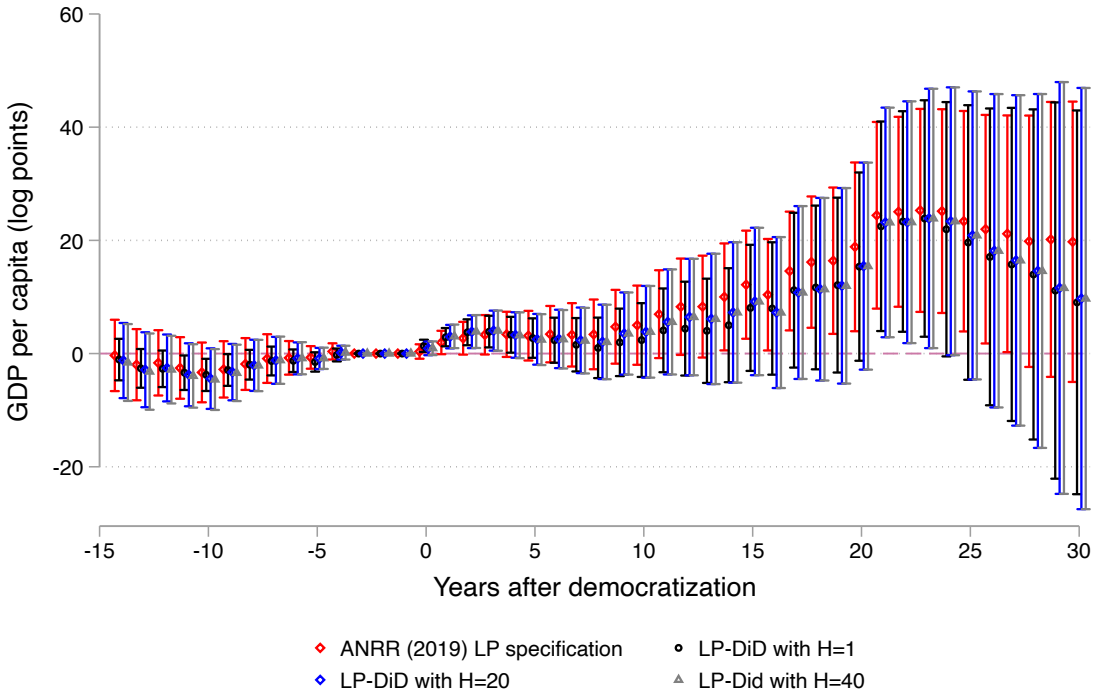


Table 1: Root Mean Squared Error (RMSE) – Exogenous Treatment Scenario

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.

M. Alloza, J. Gonzalo, and C. Sanz. Dynamic effects of persistent shocks. 2019.

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

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

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.

S. Chabé-Ferret. "should we combine difference in differences with conditioning on pre-treatment outcomes? Toulouse School of Economics Working Paper n.17-824, 2017.

S. Chen, V. Chernozhukov, and I. Fernández-Val. Mastering panel metrics: Causal impact of democracy on growth. *AEA Papers and Proceedings*, 109:77–82, May 2019. doi: 10.1257/pandp.20191071. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20191071.

J. Cloyne, O. Jorda, and A. Taylor. Decomposing the fiscal multiplier. NBER Working Papers 26939, National Bureau of Economic Research, Inc, 2020. URL https://EconPapers.repec.org/RePEc:nbr:nberwo:26939.

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 and X. D'Haultfoeuille. Two-way fixed effects and differences-in-differences estimators with several treatments. Working Paper 29734, National Bureau of Economic Research, February 2022. URL http://www.nber.org/papers/w29734.

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.

N. Fortin, T. Lemieux, and S. Firpo. Decomposition Methods in Economics. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4, chapter 1, pages 1–102. Elsevier, 2011. URL https://ideas.repec.org/h/eee/labchp/4-01.html.

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.

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

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.

S. Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1911408.

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 group-specific ATE by the LP-DiD estimator, first in a baseline version without control variables (equations 22 and 23 in the main text) and then in more general specifications with control variables.

A.1 Baseline version without control variables

Assumptions about the DGP

Assume that parallel trends and no anticipation hold unconditionally (Assumptions 1 and 2 in the main text). Specifically, assume that the DGP satisfies

$$E(y_{it}|D_{it}=0) = \alpha_i + \alpha_t\,, \tag{1}$$

where i indexes units, t indexes time, D is a binary treatment indicator, α_i is a unit-specific fixed effect, and α_t is a common time-varying effect.

Therefore actual outcomes are equal to

$$y_{it} = \alpha_i + \alpha_t + \tau_{it}D_{it} + u_{it}\,, \tag{2}$$

where $\tau_{it} = E(y_{it}(1) - y_{it}(0))$ is the treatment effect for unit i at time t and $E(u_i|i,t,D) = 0$. Treatment is binary, staggered and absorbing. All units are untreated in the first period.

This DGP implies that the observed long-difference $\Delta y_{i,t+k} = y_{i,t+k} - y_{i,t-1}$ is equal to

$$\Delta y_{i,t+k} = \delta_t^k + \tau_{i,t+k}D_{i,t+k} - \tau_{i,t-1}D_{i,t-1} + v_{i,t}^k\,, \tag{3}$$

where $\delta_t^k = \alpha_{t+k} - \alpha_t$ and $v_{it}^k = u_{i,t+k} - u_{i,t-1}$.

LP-DiD specification

Consider the following LP-DiD specification with clean controls:

$$\Delta y_{i,t+k} = \delta_t^k + \beta^{k\ LP\text{–}DiD}\Delta D_{it} + \epsilon_{it}^k\,, \tag{4}$$

restricting the sample to observations that are either:

$$\begin{cases} \text{treated} & \Delta D_{it} = 1\,, \\ \text{clean control} & D_{i,t+k} = 0\,. \end{cases} \tag{5}$$

Here, $\beta^{k\ LP\text{–}DiD}$ is the LP-DiD estimate of the average k–periods treatment effect.

Derivation of the weights

As in the main text, we define groups (or treatment cohorts) $g \in \{0,1,2,...,G\}$ as exhaustive and mutually exclusive sets of units. Groups are defined so that all units within a group enter

treatment in the same period; two units belonging to different groups enter treatment in two different periods; group $g = 0$ is the never-treated group. Denote the time period in which group g enters treatment as p_g.

For each treatment group $g > 0$, define the clean control sample (CCS) for group g at time horizon k (denoted as $CCS_{g,k}$) as the set of observations for time $t = p_g$ that satisfy condition 5. Therefore $CCS_{g,k}$ includes the observations at time p_g for all units that either enter treatment at p_g or are still untreated at $p_g + k$. In other words, $CCS_{g,k}$ includes observations at p_g for group g and its *clean controls*.

By definition of groups and CCSs, each observation that satisfies condition 5 enters into one and only one CCS. Therefore, the unbalanced panel dataset defined by the clean control condition in 5 can always be reordered as a 'stacked' dataset, in which observations are grouped into consecutive and non-overlapping CCSs.

Moreover, for any observation $\{i, t\} \in CCS_{g,k}$, we have $\Delta D_{i,t} = \Delta D_{i,p_g} = D_{i,p_g}$. This follows from the fact that for any $\{i, t\} \in CCS_{g,k}$, 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 $\{i, t\} \in CCS_{g,k}$ 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,

$$E\left(\beta^{k\ LP-DiD}\right) = \frac{\sum_{j=1}^{G} \sum_{i \in CCS_{j,k}} \left[\widetilde{\Delta D}_{i,p_j} E\left(\Delta y_{i,p_j+k}\right)\right]}{\sum_{j=1}^{G} \sum_{i \in CCS_{j,k}} \widetilde{\Delta D}_{i,p_j}^2}, \tag{6}$$

where $\widetilde{\Delta D}_{i,p_g}$ is the residual from a regression of ΔD on time indicators in the sample defined by condition 5.

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

$$\widetilde{\Delta D}_{i,p_g} = \Delta D_{i,p_g} - \frac{\sum_{i \in CCS_{g,k}} \Delta D_{i,p_g}}{N_{CCS_{g,k}}} = D_{i,p_g} - \frac{\sum_{i \in CCS_{g,k}} D_{i,p_g}}{N_{CCS_{g,k}}} = D_{i,p_g} - \frac{N_g}{N_{CCS_{g,k}}}, \tag{7}$$

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

The first equality in 7 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 of equations 1 and 2, we have

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

where the weights are given by

$$
\omega^{LP\text{–}DiD}_{g,k} = \frac{N_g\widetilde{\Delta D}_{g,p_g}}{\sum_{g\neq 0}N_g\widetilde{\Delta D}^2_{g,p_g}} = \frac{N_g\left(1-\frac{N_g}{N_{CCS_{g,k}}}\right)}{\sum_{g\neq 0}N_g\left(1-\frac{N_g}{N_{CCS_{g,k}}}\right)} = \frac{N_{CCS_{g,k}}[n_{gk}(n_{c,g,k})]}{\sum_{g\neq 0}N_{CCS_{g,k}}[n_{g,k}(n_{c,g,k})]}\,, \tag{8}
$$

where $n_{g,k} = \frac{N_g}{N_{CCS_{g,k}}}$ is the share of treated units in the $CCS_{g,k}$ subsample; and $n_{c,g,k} = \frac{N_{c,g,k}}{N_{CCS_{g,k}}}$ is the share of control units in the $CCS_{g,k}$ subsample. Recall that $\tau_g(k)$ was defined in the main text as the group-specific k-periods forward average treatment effect for group g.

A.2 Weights with control variables

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

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

A.2.1 Covariates with linear and homogeneous effects

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

$$\Delta y_{i,t+k} = \delta_t^k + \boldsymbol{\rho}_k \Delta \boldsymbol{x}_{it} + \tau_{i,t+k} D_{i,t+k} - \tau_{i,t-1} D_{i,t-1} + v_{i,t}^k \,, \tag{9}$$

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

$$\begin{array}{rll} y_{i,t+k} - y_{i,t-1} = & \beta_k^{LP-DiD} \Delta D_{it} & \} \text{ treatment indicator} \\ & + \boldsymbol{\rho}_k \Delta \boldsymbol{x}_{it} & \} \text{ covariates} \\ & + \delta_t^k & \} \text{ time effects} \\ & + e_{it}^k; & \text{for } k = 0, \ldots, K \,, \end{array} \tag{10}$$

restricting the sample to observations that respect condition 5.

Weights derivation All the definitions of groups and 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 10 can be rewritten as

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

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

$$E\left(\hat{\beta}_k^{LP-DiD}\right) = \frac{\sum_{j=1}^{G} \sum_{i \in CCS_{j,k}} \left[\widetilde{\Delta D}_{i,p_j} E\left(\Delta y_{i,t+k} - \hat{\rho}_k \Delta \boldsymbol{x}_{it}\right)\right]}{\sum_{j=1}^{G} \sum_{i \in CCS_{j,k}} \widetilde{\Delta D}_{i,p_j}^2}, \tag{11}$$

where $\widetilde{\Delta D}_{i,p_g}$ is the residual from a regression of ΔD on time indicators in the sample defined by condition 5.

The equivalence of eq. 7 above still holds; therefore, for all observations belonging to the same group $g > 0$, we have $\widetilde{\Delta D}_{i,p_g} = \widetilde{\Delta D}_{g,p_g} = 1 - \frac{N_g}{N_{CCS_{g,k}}}$

Given the assumptions about the DGP, we have

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

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:

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

where the weights are given by equation 8 above.

A.2.2 More general setting

Now consider a more general setting, in which Assumptions 3 and 4 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

$$E\left(\beta^{k\ LP-DiD}\right) = \frac{\sum_{j=1}^{G}\sum_{i\in CCS_{j,k}}\left[\widetilde{\Delta D}^{c}_{i,p_j} E\left(\Delta y_{i,p_j+k}\right)\right]}{\sum_{j=1}^{G}\sum_{i\in CCS_{j,k}}\left(\widetilde{\Delta D}^{c}_{i,p_j}\right)^2}, \tag{12}$$

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 5.

The weights are thus given by

$$\omega^{c\ LP-DiD}_{g,k} = \frac{N_g \widetilde{\Delta D}^{c}_{g,p_g}}{\sum_{g\neq 0} N_g \left(\widetilde{\Delta D}^{c}_{g,p_g}\right)^2}, \tag{13}$$