Difference-in-Differences Estimators of Intertemporal Treatment Effects*

Clément de Chaisemartin† Xavier D'Haultfœuille‡

First version: July 8, 2020
This version: March 18, 2022

Abstract

We study treatment-effect estimation, with a panel where groups may experience multiple changes of their treatment dose. We make parallel trends assumptions, but do not restrict treatment effect heterogeneity, unlike the linear regressions that have been used in such designs. We extend the event-study approach for binary-and-staggered treatments, by redefining the event as the first time a group's treatment changes. This yields an event-study graph, with reduced-form estimates of the effect of having been exposed to a weakly higher amount of treatment for ℓ periods. We show that the reduced-form estimates can be combined into an economically interpretable cost-benefit ratio.

Keywords: differences-in-differences, dynamic treatment effects, heterogeneous treatment effects, event-study graph, parallel trends, panel data, policy evaluation, cost-benefit analysis.

JEL Codes: C21, C23

*We are very grateful to Timothy Armstrong, Michal Kolesár, Isabelle Méjean, Ulrich Müller, Aureo de Paula, Ashesh Rambachan, Jonathan Roth, Pedro Sant'Anna, Jesse Shapiro and seminar participants at Berkeley, CREST, the Harris School of Public Policy, Mannheim, McGill, Oxford, Princeton, Paris School of Economics, Santa Clara, Sciences Po, UC Dublin, Universidad Javeriana, Université Paris Dauphine, and Yale for their helpful comments. Xavier D'Haultfœuille gratefully acknowledges financial support from the research grants Otelo (ANR-17-CE26-0015-041). Part of this research was conducted while Xavier D'Haultfœuille was at PSE, which he thanks for its hospitality.

†Economics Department, Sciences Po, clement.dechaisemartin@sciencespo.fr

‡CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr.

1 Introduction

We study treatment-effect estimation, using a panel of groups, indexed by g, that are exposed to different doses of the treatment at different time periods, indexed by t. $D_{g,t}$, group g's period-t treatment, may have an effect on $Y_{g,t}$, group g's period-t outcome, the so-called instantaneous effect. But $D_{g,t}$ may also have an effect on g's future outcomes, the so-called dynamic effects.

When the treatment is binary and staggered, meaning that the treatment can only increase and can do so at most once, a common practice to estimate instantaneous and dynamic effects is to run a two-way fixed effects (TWFE) event-study regression of the outcome on group fixed effects, period fixed effects, and indicators for whether group g started receiving the treatment ℓ periods ago. Sun and Abraham (2021) have shown that the coefficients on the indicators for having started receiving the treatment ℓ periods ago may be biased for the average effect of having been treated for $\ell + 1$ periods, if treatment effects vary across groups and over time. A recent literature, discussed in more details below, has proposed alternative estimators robust to heterogeneous treatment effects (see e.g. Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Borusyak et al., 2021) in binary-and-staggered designs.

In this paper, we consider general designs, where the treatment may neither be binary nor staggered. Such situations are frequent in practice. In a survey of the 100 most-cited papers published by the American Economic Review (AER) from 2015 to 2019, we find that 26 papers estimate a TWFE regression, but only four have a binary-and-staggered treatment. Papers that do not have a binary-and-staggered treatment and estimate dynamic effects do so using three different methods. In some papers, there is no variation in treatment timing: all treated groups start getting treated at the same date, with group-specific treatment intensities. Then, researchers have estimated TWFE regressions of the outcome on the treatment intensity interacted with period fixed effects. In more complicated designs where there may be variation in treatment timing and a group's treatment may increase or decrease multiple times, some researchers have estimated TWFE regressions of the outcome on the treatment and its first K lags, the so-called distributed-lag regression. In those more complicated designs, other researchers have estimated a panel-data version of the local-projection method proposed by Jordà (2005) for time-series data. We start by showing that under a parallel trends assumption, those three regressions may produce biased estimates of the treatment's instantaneous and dynamic effects, if effects are heterogeneous across groups and over time. In particular, they do not satisfy the so-called no-sign reversal property: one could have that the treatment's instantaneous and dynamic effects are positive in every (g, t) cell, but the expectations of those regression coefficients are negative. We also show that the panel-data version of the local-projection method may yield biased estimates even if effects are homogeneous.

Instead, we propose a different strategy to estimate the treatment's instantaneous and dynamic

effects in general designs. Our paper's main idea is to propose a generalization of the event-study approach to such designs, by defining the event as the period where a group's treatment changes for the first time. With a binary-and-staggered treatment, the event per our definition is the period where a group gets treated, so our definition extends the standard one to general designs.

We start by showing that for any group g whose treatment changed for the first time at period F_g, the instantaneous and dynamic effects of that change can be unbiasedly estimated. Let $Y_{g,t}(d_1, ..., d_t)$ denote the potential outcome of group g at period t, if her treatments from period 1 to t are equal to $(d_1, ..., d_t)$. Let $\delta_{g,\ell} = E(Y_{g,F_g+\ell} - Y_{g,F_g+\ell}(D_{g,1}, ..., D_{g,1}))$ be the expected difference between group g's actual outcome at $F_g + \ell$ and the counterfactual "status quo" outcome it would have obtained if its treatment had remained equal to its period-one value from period one to $F_g + \ell$. To estimate $\delta_{g,\ell}$, we propose a difference-in-differences (DID) estimator $\text{DID}_{g,\ell}$ comparing the $F_g - 1$-to-$F_g + \ell$ outcome evolution between group g, and groups whose treatment has not changed yet at $F_g + \ell$ and with the same treatment as g at period one. The $\text{DID}_{g,\ell}$ estimators are unbiased under parallel trends assumptions. To test those parallel trends assumptions, we propose placebo estimators comparing the outcome trends of switchers and non-switchers before the switchers switch.

We then aggregate those estimators into an estimator of the effect of having experienced a weakly higher amount of treatment for ℓ periods. For any real number x and $t \in \{1, ..., T\}$, let $\boldsymbol{x}_t$ denote a $1 \times t$ vector with coordinates equal to x. When the treatment is binary, for groups untreated at period one, $D_{g,1} = 0$, so

$$\delta_{g,\ell} = E(Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, 1, D_{g,F_g+1}, ..., D_{g,F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})).$$

For groups treated at period one, $D_{g,1} = 1$, so

$$-\delta_{g,\ell} = E(Y_{g,F_g+\ell}(\mathbf{1}_{F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{1}_{F_g-1}, 0, D_{g,F_g+1}, ..., D_{g,F_g+\ell})).$$

The right-hand side of the two equations above are effects of having experienced a weakly higher amount of treatment for $\ell + 1$ periods. Accordingly, we aggregate the $\text{DID}_{g,\ell}$ estimators into a DID_ℓ estimator, multiplying by minus one the $\text{DID}_{g,\ell}$ of groups treated at period one. With a non-binary treatment, we also propose a method to aggregate the $\text{DID}_{g,\ell}$ estimators, to estimate the effect of having experienced a weakly higher amount of treatment for $\ell+1$ periods. Unlike the TWFE regression coefficients described above, our DID_ℓ estimators satisfy the no-sign reversal property.

Ultimately, our approach leads to an event-study graph, with the distance to the first treatment change on the x-axis, the DID_ℓ estimators on the y-axis to the right of zero, and placebo estimators on the y-axis to the left of zero. We believe this event-study graph is useful to test the parallel trends assumption, and to provide reduced-form evidence of whether increasing the

treatment for $\ell+1$ periods increases or decreases the outcome on average. However, interpreting the magnitude of the DID_ℓ estimators might be complicated. For instance, with three periods and three groups such that $(D_{1,1}=0, D_{1,2}=4, D_{1,3}=1)$, $(D_{2,1}=0, D_{2,2}=2, D_{2,3}=3)$, and $(D_{2,1}=0, D_{2,2}=0, D_{2,3}=0)$, DID_1 estimates the average of $E(Y_{1,3}(0,4,1)-Y_{1,3}(0,0,0))$ and $E(Y_{2,3}(0,2,3)-Y_{2,3}(0,0,0))$. Accordingly, DID_1 does not estimate by how much the outcome increases on average when the treatment increases by a given amount for a given number of periods.

To circumvent this important limitation, we propose two strategies. First, we recommend complementing the reduced-form event-study graph described above with a first-stage event-study graph, where the outcome is replaced by the treatment. The estimators on the first-stage graph show the average value of $|D_{g,F_g+\ell}-D_{g,1}|$ across all groups entering in DID_ℓ. In the example above, the first two estimates on the first-stage graph are equal to $1/2(D_{1,2}-D_{1,1}+D_{2,2}-D_{2,1})=3$ and $1/2(D_{1,3}-D_{1,1}+D_{2,3}-D_{2,1})=2$. This reflects the fact that in this example, DID_1 is an effect produced by increasing the previous and current treatment by 3 and 2 units on average. Second, a weighted average across ℓ of our reduced-form estimators divided by a weighted average across ℓ of our first-stage estimators is unbiased for a parameter with a clear economic interpretation. That parameter may be used to conduct a cost-benefit analysis comparing groups' actual treatments to the status quo scenario where they would have kept all along the same treatment as in period one. In other words, that parameter can be used to determine if the policy changes that took place over the duration of the panel led to a better situation than the one that would have prevailed if no policy change had been undertaken, a natural policy question. Importantly, that parameter can also be interpreted as an average total effect per unit of treatment, where "total effect" refers to the sum of the instantaneous and dynamic effects of a treatment.

Like the regression-based estimators used by the AER papers in our survey, the estimators we propose rely on parallel trends assumptions. But unlike them, our estimators are robust to heterogeneous treatment effects. To our knowledge, our estimators are the first DID estimators of instantaneous and dynamic treatment effects that are robust to heterogeneous effects, and that can be used in general designs. Our estimators can be used with non-staggered binary or discrete treatments, and with staggered continuous treatments.[1] Accordingly, our estimators are widely applicable. All our estimators are computed by the `did_multiplegt` Stata package, available from the SSC repository.

Finally, we use our estimators to revisit Favara and Imbs (2015). These authors use the differential timing and intensity of banking deregulation episodes in US states during the 90s and early 2000s to identify the effect of deregulations on credit supply. Using our method, we show that banking deregulations initially have a small effect on credit supply, but that these effects increase over time. These findings differ sharply from those obtained by Favara and Imbs (2015).

[1] de Chaisemartin et al. (2022) extend our estimators to non-staggered continuous treatments.

Using the panel-data version of the local-projection method, the authors find significant short-run effects, which shortly vanish. Our findings also sharply differ from those one would obtain with a distributed-lag regression.

The paper is organized as follows. Section 2 introduces the set up and our assumptions. Section 3 shows that commonly used methods to estimate instantaneous and dynamic effects with general designs are not robust to heterogeneous treatment effects. Section 4 defines our parameters of interest, presents our estimators, and presents our placebo estimators. Section 5 presents our inference results. Section 6 uses our estimators to revisit Favara and Imbs (2015).

Related literature

Our decompositions of the regression coefficients used by applied researchers to estimate the dynamic effects in general designs are related to the literature showing that TWFE regressions are not robust to heterogeneous treatment effects (see, e.g., de Chaisemartin and D'Haultfœuille, 2020*a*; Borusyak et al., 2021; Goodman-Bacon, 2021), and in particular to Sun and Abraham (2021) and de Chaisemartin and D'Haultfoeuille (2020*b*), who respectively study event-study regressions and regressions with several treatments.

In designs with a binary treatment, where groups are all untreated at period 1 but where the treatment can switch on and off, Deryugina (2017) and Sun and Abraham (2021) have proposed to redefine the treatment as "having ever been treated" and proceed as if the design was staggered. This idea is related to our idea of "binarizing and staggerizing" general designs. However, we consider more general designs than those two papers, and redefining the event as "changing treatment for the first time" in those more complicated designs is, to our knowledge, a contribution of this paper. Moreover, unlike us those two papers do not propose a way to convert the reduced-form estimates one obtains after staggerizing a non-staggered treatment into estimates with a clear economic interpretation.

The estimators we propose build upon two ideas recently proposed in the DID literature. First, with a binary-and-staggered treatment, Callaway and Sant'Anna (2021) have proposed DID estimators of the instantaneous and dynamic treatment effects that use the not-yet-treated groups as controls. de Chaisemartin and D'Haultfœuille (2020*a*) have also proposed to use the not-yet-treated as controls in staggered designs, to form DID estimators of the instantaneous treatment effect.[2] Our estimators extend that idea, by proposing to use the not-yet-switchers as controls. With a non-binary treatment and a non-staggered design, but ruling out dynamic effects, de Chaisemartin and D'Haultfœuille (2018) and de Chaisemartin and D'Haultfœuille (2020*a*) have proposed DID estimators comparing a group switching treatment between $t-1$

[2]Using the not-yet-treated as controls has also been proposed in the duration literature, see Sianesi (2004).

and t to a control group whose treatment does not change between $t-1$ and t, and with the same treatment as the switching group in $t-1$. The estimators we propose in this paper also build upon that second idea, by proposing DID estimators using as controls the not-yet-switchers with the same treatment as the switching group at the start of the panel.

Though binary-and-staggered designs are not the main focus of our paper, our estimators can also be used in such designs, and we now compare them with the estimators proposed by Callaway and Sant'Anna (2021). In such designs, the groups used as control groups by our estimators are actually the not-yet-treated. The estimators in Callaway and Sant'Anna (2021) either use the never-treated or the not-yet-treated as controls. When the not-yet-treated are used and there are no covariates in the estimation, our and their estimators are numerically equivalent, and our and their estimator of the instantaneous treatment effect is also numerically equivalent to the DID_M estimator in de Chaisemartin and D'Haultfœuille (2020*a*). With covariates, our and their estimators differ: while they account non-parametrically for the effect of time-invariant covariates, we account linearly for the effect of time-varying covariates. Those two approaches rely on non-nested assumptions, and could in principle be combined. Our approach can accomodate group-specific linear-trends in the estimation, unlike that in Callaway and Sant'Anna (2021). We refer the reader to de Chaisemartin and D'Haultfoeuille (2022) for a more thorough discussion of the alternative estimators available with a binary-and-staggered treatment, and in particular those proposed by Sun and Abraham (2021) and Borusyak et al. (2021).

Finally, in work posterior to ours, Callaway et al. (2021) have derived identification results specifically for the case with a continuous and staggered treatment. Though our estimators are applicable to continuous and staggered treatments, that case is not our main focus our interest, so their analysis of this case is much more detailed than ours. Still, their identification results under parallel trends are similar to ours: in that case our two papers propose estimands comparing the outcome evolution of treated and not-yet-treated. They also consider identification results under a stronger assumption. There is no connection between those results and ours.

Alternative approaches

Instead of the approach we propose, one could separately estimate the effects, relative to the status quo, of all the treatment trajectories observed in the data. This would amount to aggregating the $\text{DID}_{g,\ell}$ estimators only across groups with the same treatments from period 1 to $F_g+\ell$, and would result in more easily interpretable estimators. Callaway and Sant'Anna (2021) successfully pursue that strategy in the binary-and-staggered case. But in that case, there are at most $T+1$ treatment trajectories. In the non-staggered case, even with a binary treatment there may be up to 2^T treatment trajectories. Trying to separately estimate the effects of all trajectories may often yield noisy estimates, but could prove successful when the treatment can

take a low number of values and T is low.

Rather than imposing a parallel trends assumption like we do, an important literature has proposed estimators of instantaneous and dynamic effects of treatments that can both increase or decrease over time, under a sequential ignorability assumption. This assumption requires that at each period, the treatment is independent of current or future potential outcomes, conditional on past treatments and outcomes (see Robins, 1986; Murphy et al., 2001; Bojinov et al., 2021).[3] This condition is appealing, because it allows future treatments to depend on past outcomes, while parallel trends could be violated if groups switch in or out of the treatment after experiencing positive or negative outcome shocks (see Ashenfelter, 1978). However, the estimators proposed under sequential ignorability assume that groups' treatment probability conditional on past treatments and outcomes are known (see Murphy et al., 2001; Bojinov et al., 2021). This condition is met in sequential randomized experiments, where the experimenter fully controls the treatment assignment mechanism. On the other, hand, this condition is not met in observational studies. Then, the propensity score could be estimated, but one may face a curse of dimensionality: at period t, one would have to estimate one propensity score per treatment and outcome trajectory from period 1 to $t-1$. Overall, those two approaches complement each other: while the sequential ignorability approach is better suited to sequential experiments, our approach may be better suited to observational studies, keeping in mind that parallel trends should not be taken for granted and should be tested, for instance using the placebo estimators we propose. Sensitivity analyses assessing results' robustness to violations of parallel trends may also be warranted (see Manski and Pepper, 2018; Rambachan and Roth, 2019).

2 Set-up and identifying assumptions

One considers observations that can be divided into G groups and T periods. Time periods are indexed by $t \in \{1, ..., T\}$. Groups are indexed by $g \in \{1, ..., G\}$. There are $N_{g,t} > 0$ observations in group g at period t. The data may be an individual-level panel or repeated cross-section data set where groups are, say, individuals' county of birth. The data could also be a cross-section where cohort of birth plays the role of time. It is also possible that for all (g, t), $N_{g,t} = 1$, e.g. a group is one individual or firm.

One is interested in measuring the effect of a treatment on some outcome. For every $(i, g, t) \in \{1, ..., N_{g,t}\} \times \{1, ..., G\} \times \{1, ..., T\}$, let $D_{i,g,t}$ denote the treatment status of observation i in group g at period t. Let $\mathcal{D}$ denote the set of values $D_{i,g,t}$ can take. For all $\boldsymbol{d} \in \mathcal{D}^T$, let $Y_{i,g,t}(\boldsymbol{d})$ denote the potential outcome of observation i in group g at period t, if her treatments from

[3] Han (2021) extend this approach to endogenous treatments, relying instead on instruments. Murphy (2003) and Han (2019) discuss optimal treatment allocations in such set-ups.

period 1 to T are equal to $\boldsymbol{d}$. This dynamic potential outcome framework is similar to that in Robins (1986). It allows for the possibility that observations' outcome at time t be affected by their past and future treatments. Some observations may have already been treated prior to period 1, the first period in the data, and those treatments may still affect some of their period-1-to-T outcomes, the so-called initial conditions problem. However, we cannot estimate such dynamic effects, as treatments and outcomes are not observed for those periods, so we do not account for this potential dependency in our notation. We discuss potential strategies to account for this dependency in Section 4.4 of the Web Appendix.

We focus on sharp designs, where the treatment does not vary within (g,t) cells.

Assumption 1 *(Sharp design)* $\forall (i,g,t) \in \{1,...,N_{g,t}\} \times \{1,...,G\} \times \{1,...,T\}$, $D_{i,g,t} = D_{g,t}$.[4]

Assumption 1 is for instance satisfied when the treatment is a group-level variable, as a county- or a state-law, or when $N_{g,t} = 1$. Then, let $\boldsymbol{D}_g = (D_{g,1},...,D_{g,T})$ be a $1 \times T$ vector stacking the treatments of group g from period 1 to T, and let $\boldsymbol{D} = (\boldsymbol{D}_1,...,\boldsymbol{D}_G)$ be a vector stacking the treatments of all groups at every period. For all $\boldsymbol{d} \in \{0,1\}^T$, let also $Y_{g,t}(\boldsymbol{d}) = 1/N_{g,t} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}(\boldsymbol{d})$ denote the average potential outcome of group g at period t, if the treatments of group g from period 1 to T are equal to $\boldsymbol{d}$. Finally, we let $Y_{g,t} = Y_{g,t}(\boldsymbol{D}_g)$ denote the observed average outcome in group g at period t.

We now introduce our key identifying assumptions.

Assumption 2 *(No Anticipation) For all g, for all $\boldsymbol{d} \in \{0,1\}^T$, $Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(d_1,...,d_t)$.*

Assumption 2 requires that a group's current outcome do not depend on her future treatments, the so-called no-anticipation hypothesis. Abbring and Van den Berg (2003) have discussed that assumption in the context of duration models, and Malani and Reif (2015), Botosaru and Gutierrez (2018), and Sun and Abraham (2021) have discussed it in the context of DID models.

Let $\mathbf{0}$ denote an $1 \times T$ vector of zeros. Hereafter, we refer to $Y_{g,t}(\mathbf{0})$ as group g's never-treated potential outcome at period t.

Assumption 3 *(Independent groups, strong exogeneity, and parallel trends for never-treated outcome) $\forall t \geq 2$ and $\forall g \in \{1,...,G\}$, $E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})|\boldsymbol{D})$ does not vary across g.*

To ease the interpretation of Assumption 3, we state below three conditions that, together, are sufficient for Assumption 3 to hold, and that are easily interpretable:

1. The vectors $((Y_{g,t}(\mathbf{0}))_{1 \leq t \leq T}, \boldsymbol{D}_g)$ are mutually independent.
2. $E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})|\boldsymbol{D}_g) = E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}))$.

[4] In this assumption and others below, equalities and inequalities involving random variables are implicitly assumed to hold with probability one.

3. $\forall t \geq 2$, $E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}))$ does not vary across g.

Point 1 requires that the potential outcomes and treatments of different groups be independent, a commonly-made assumption in DID analysis, where standard errors are usually clustered at the group level (see Bertrand et al., 2004). Point 2 is related to the strong exogeneity condition in panel data models. It requires that the shocks affecting group g's never-treated outcome be mean independent of group g's treatments. For instance, this rules out cases where a group gets treated because it experiences negative shocks, the so-called Ashenfelter's dip (see Ashenfelter, 1978). Point 3 requires that in every group, the expectation of the never-treated outcome follow the same evolution over time. It is a generalization of the standard parallel trends assumption in DID models (see, e.g., Abadie, 2005) to our set-up allowing for dynamic effects. Sun and Abraham (2021), Athey and Imbens (2022), and Callaway and Sant'Anna (2021) also consider that assumption.

3 Current Practice

We conducted a survey of highly-influential papers published by the AER from 2015 to 2019 and that have used TWFE regressions. To do so, we first ran a Google Scholar (GS) search of all papers published by the AER in 2015, sorted according to GS's relevance criteria, which is nearly equivalent to sorting on GS citations (Beel and Gipp, 2009). We systematically reviewed the first 20 papers, and identified three that have estimated at least one TWFE regression. We repeated the same process for 2016, 2017, 2018, and 2019. In total, we have reviewed 100 papers, and have found 26 that have estimated at least one TWFE regression. The list of papers can be found in Table 2 in the Web Appendix.

Of those 26 papers, two have a binary treatment with no variation in treatment timing,[5] and four have a binary treatment and a staggered adoption design. Four of those six papers estimate dynamic treatment effects, using the event-study regression considered by Sun and Abraham (2021). Among the remaining 20 papers, 11 estimate only static TWFE regressions, thus implicitly ruling out dynamic treatment effects. The remaining nine papers estimate dynamic effects, using one of the three estimations methods described below.

When $D_{g,t} = I_g 1\{t \geq F\}$, meaning that treated groups all start getting treated at the same date F with group-specific intensities I_g, researchers have sometimes estimated TWFE regressions of the outcome on group and period FEs, and the treatment intensity interacted with period FEs.

[5] With a binary treatment and no variation in treatment timing, TWFE regressions estimate the average treatment effect on the treated, see de Chaisemartin and D'Haultfoeuille (2022).

Regression 1 *(TWFE regression with treatment intensity interacted with period FEs) Assume that* $D_{g,t} = I_g 1\{t \geq F\}$*, for* $2 \leq F \leq T$*. For every* $\ell \in \{1, ..., T\}, \ell \neq F-1$*, let* $\widehat{\beta}_{fe,\ell}$ *be the coefficient on* $I_g 1\{t = \ell\}$*, in a regression of* $Y_{i,g,t}$ *on group and period FEs and* $(I_g 1\{t = \ell\})_{\ell \in \{1,...,T\}, \ell \neq F-1}$.

For $\ell \in \{F, ..., T\}$, the coefficients $\widehat{\beta}_{fe,\ell}$ are supposed to estimate the instantaneous and dynamic effects of the treatment. For $\ell \in \{1, ..., F-2\}$, the coefficients $\widehat{\beta}_{fe,\ell}$ are supposed to be placebos testing the parallel trends assumption. Instead of Regression 1, researchers have also estimated regressions of $Y_{g,\ell} - Y_{g,F-1}$, the outcome change from the period before treatment to later periods, on I_g. One can show that when $N_{g,t}$ does not vary with t, as assumed below, the coefficient on I_g in that regression is equal to $\widehat{\beta}_{fe,\ell}$.

We now study what the coefficients $\widehat{\beta}_{fe,\ell}$ identify. For any real number $x \neq 0$ and integer $k \geq 1$, let $\Delta_{g,t}(k, x) = (Y_{g,t}(\mathbf{0}_{t-k}, \boldsymbol{x}_k) - Y_{g,t}(\mathbf{0}))/x$ denote group g's effect of having received x rather than 0 units of treatment from period $t - k + 1$ to t, normalized by x.

Proposition 1 *If Assumptions 1-3 hold,* $N_{g,t} = N_g$ *for all* (g, t)*, and* $D_{g,t} = I_g 1\{t \geq F\}$ *for* $F \in \{2, ..., T\}$*, then:*[6]

1. *For all* $\ell \in \{F, ..., T\}$*,*

$$E\left(\widehat{\beta}_{fe,\ell}\right) = E\left(\sum_{g: I_g \neq 0} w_g^{fe} \Delta_{g,\ell}(\ell - (F-1), I_g)\right), \tag{1}$$

   where w_g^{fe} *are weights proportional to* $N_g I_g (I_g - \overline{I})$*, with* $\overline{I} = \sum_{g=1}^{G} N_g I_g / \sum_{g=1}^{G} N_g$*, and where* $\sum_{g: I_g \neq 0} w_g^{fe} = 1$*.*

2. *If* $F > 2$*, for all* $\ell \in \{1, ..., F-2\}$*,*

$$E\left(\widehat{\beta}_{fe,\ell}\right) = 0. \tag{2}$$

Point 1 of Proposition 1 shows that for $\ell \in \{F, ..., T\}$, $\widehat{\beta}_{fe,\ell}$ estimates a weighted sum across groups of $\Delta_{g,\ell}(\ell - (F-1), I_g)$, group g's effect of having received I_g rather than 0 units of treatment for $\ell - F + 1$ periods, normalized by I_g. The weights w_g^{fe} sum to one but they are not proportional to the proportion that group g accounts for in the population, so $\widehat{\beta}_{fe,\ell}$ may be biased for the average across groups of these normalized dynamic effects. Perhaps more worryingly, some of the weights may be negative, so $E\left(\widehat{\beta}_{fe,\ell}\right)$ does not satisfy the so-called no-sign-reversal property: it could be, say, negative, even if all the group-specific normalized dynamic effects are positive. If the average treatment intensity $\overline{I}$ is positive (resp. negative), negative weights arise if there are groups with treatment intensity strictly included between 0

[6] In Propositions 1-3, we implicitly assume that the coefficients under consideration are well-defined, meaning that the standard rank condition holds for the corresponding OLS regression.

and $\overline{I}$ (resp. between $\overline{I}$ and 0). Treatment effects of groups with a treatment intensity of the same sign as, but closer to zero than $\overline{I}$ are weighted negatively by $\widehat{\beta}_{fe,\ell}$. Those effects may differ from those weighted positively by $\widehat{\beta}_{fe,\ell}$, for instance if groups choose their treatment intensity based on their treatment effect. Point 2 of Proposition 1 shows that for $\ell \in \{1, ..., F-2\}$, $E\left(\widehat{\beta}_{fe,\ell}\right) = 0$ under Assumptions 2 and 3. Accordingly, testing if $\widehat{\beta}_{fe,\ell}$ significantly differs from 0 is a valid test of Assumptions 2 and 3, even in the presence of heterogeneous treatment effects. The proof of Proposition 1 exploits similar ideas as the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020*a*), and leverages specific features of Regression 1.

When one does not have that $D_{g,t} = I_g 1\{t \geq F\}$, there may be variation in treatment timing and a group's treatment may increase or decrease multiple times. Then, researchers have sometimes estimated TWFE regressions of the outcome on the treatment and its first K lags, the so-called distributed-lag regression.

Regression 2 *(Distributed-lag regression) For every $\ell \in \{0, ..., K\}$, let $\widehat{\beta}_{dl,\ell}$ denote the coefficients on $D_{g,t-\ell}$ in a regression of $Y_{i,g,t}$ on group and period FEs and $(D_{g,t-\ell})_{\ell \in \{0,...,K\}}$, in the subsample of observations made at periods $t \geq K+1$.*

Regression 2 is discussed as a tool for dynamic effects estimation in Angrist and Pischke (2008), a hugely popular applied econometrics textbook. In practice, researchers may slightly augment or modify Regression 2. They may include treatment leads in the regression, to test for parallel trends. They may define the lagged treatments as equal to 0 at time periods when they are not observed, and estimate the regression in the full sample. They may also estimate the regression in first difference and without group fixed effects. Results similar to Proposition 2 below apply to all those variations on Regression 2.

Proposition 2 *If Assumptions 1-3 hold, $D_{g,t}$ is binary, and for all g and $t \geq K+1$ there exists real numbers $\left(\gamma^{\ell}_{g,t}\right)_{\ell \in \{0,...,K\}}$ such that for all $\boldsymbol{d} \in \{0,1\}^T$, $Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(\mathbf{0}_t) + \sum_{\ell=0}^{K} \gamma^{\ell}_{g,t} d_{t-\ell}$, then for all $\ell \in \{0, ..., K\}$,*

$$E\left(\widehat{\beta}_{dl,\ell}\right) = E\left(\sum_{\substack{(g,t): D_{g,t-\ell} \neq 0, \\ t \geq K+1}} w^{dl,\ell}_{g,t} \gamma^{\ell}_{g,t} + \sum_{\substack{\ell'=0 \\ \ell' \neq \ell}}^{K} \sum_{\substack{(g,t): D_{g,t-\ell'} \neq 0, \\ t \geq K+1}} w^{dl,\ell}_{g,t} \gamma^{\ell'}_{g,t}\right), \tag{3}$$

where $w^{dl,\ell}_{g,t}$ are weights proportional to the residuals in a regression of $D_{g,t-\ell}$ on $(D_{g,t-\ell'})_{\ell' \in \{0,...,K\}, \ell' \neq \ell}$, and such that $\sum_{\substack{(g,t): D_{g,t-\ell} \neq 0, \\ t \geq K+1}} w^{dl,\ell}_{g,t} = 1$ and $\sum_{\substack{(g,t): D_{g,t-\ell'} \neq 0, \\ t \geq K+1}} w^{dl,\ell}_{g,t} = 0$ for every $\ell' \neq \ell$.

Proposition 2 assumes that $Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(\mathbf{0}_t) + \sum_{\ell=0}^{K} \gamma^{\ell}_{g,t} d_{t-\ell}$, meaning that the distributed-lag regression is correctly specified: only the first K treatment lags affect the outcome, and those lags do not interact. Even under those strong assumptions, we show that $\widehat{\beta}_{dl,\ell}$, the coefficient

on the ℓth treatment lag, may not estimate a well-defined causal effect. Specifically, Proposition 2 shows that this coefficient estimates the sum of $K+1$ terms. The first term is a weighted sum of the effect of the ℓth treatment lag, across all (g,t) cells for which that lag is not equal to 0, with weights that sum to one but that may be negative. This term may be biased for the average effect of the ℓth treatment lag, if that effect varies across (g,t) cells. The remaining K terms are weighted sums of the effects of other treatment lags, with weights that sum to zero. If the effects of the other treatment lags vary across (g,t) cells, those terms may differ from zero and may contaminate $\widehat{\beta}_{dl,\ell}$. Unlike the weights in Proposition 1, those in Proposition 2 do not have a simple closed-form expression. However, they can easily be computed in applications. Proposition 2 considers a binary treatment that may switch on and off, but a similar, negative result can also be derived for a non-binary treatment. Proposition 2 is a special case of Corollary 2 in de Chaisemartin and D'Haultfoeuille (2020*b*), applied to Regression 2. It is also closely related to the results of Sun and Abraham (2021) for the event-study regression with a binary-and-staggered treatment.

When one does not have that $D_{g,t}=I_g1\{t\geq F\}$, researchers have also estimated regressions of leads of the outcome on group and period FEs and the treatment. Such regressions have sometimes been described as a panel-data version of the local-projection method originally proposed by Jordà (2005) for time-series data.

Regression 3 *(Local-projection panel regressions) For every $\ell\in\{0,...,K\}$, let $\widehat{\beta}_{lp,\ell}$ denote the coefficient on $D_{g,t}$ in a regression of $Y_{i,g,t+\ell}$ on group and period FEs and $D_{g,t}$, in the subsample such that $1\leq t\leq T-\ell$.*

Researchers may estimate Regression 3 in first-difference, without group fixed effects. A result similar to Proposition 3 also applies to that specification.

For all $\ell\in\{0,...,K\}$, let

$$\begin{aligned}D^{\ell}_{g,\cdot}&=\sum_{t=1}^{T-\ell}N_{g,t}D_{g,t}/\sum_{t=1}^{T-\ell}N_{g,t}\\ D_{\cdot,t}&=\sum_{g=1}^{G}N_{g,t}D_{g,t}/\sum_{g=1}^{G}N_{g,t}\\ D^{\ell}_{\cdot,\cdot}&=\sum_{g=1}^{G}\sum_{t=1}^{T-\ell}N_{g,t}D_{g,t}/\sum_{g=1}^{G}\sum_{t=1}^{T-\ell}N_{g,t}\end{aligned}$$

respectively denote group g's average treatment from period 1 to $T-\ell$, the average treatment at period t across groups, and the average treatment across groups from period 1 to $T-\ell$.

Proposition 3 *If Assumptions 1-3 hold, $D_{g,t}=I_g1\{t\geq F_g\}$ for $F_g\in\{2,...,T+1\}$, and $N_{g,t}=N_g$ for all (g,t), then for all $\ell\in\{1,...,K\}$,*

$$E[\widehat{\beta}_{lp,\ell}]=E\left[\sum_{(g,t):I_g\neq 0,t+\ell\geq F_g}w^{lp,\ell}_{g,t}\Delta_{g,t+\ell}(t+\ell-F_g+1,I_g)\right], \tag{4}$$

where the weights $w^{lp,\ell}_{g,t}$ *satisfy*

$$w^{lp,\ell}_{g,t} = \frac{N_g I_g (I_g 1\{t \geq F_g\} - D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.})}{\sum_{(g',t'):I_{g'}\neq 0, t' \geq F_{g'}} N_{g'} I_{g'} (I_{g'} - D^{\ell}_{g',.} - D_{.,t'} + D^{\ell}_{.,.})}.$$

For all $\ell \geq 1$, $\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g} w^{lp,\ell}_{g,t} \neq 1$, *unless* $\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g > t} N_g(-D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.}) = 0$. *Moreover,* $\sum_{(g,t):I_g\neq 0, t \geq F_g} w^{lp,0}_{g,t} = 1$. *Finally, if* $I_g = 1$ *for all* g *and* F_g *is not constant,* $\min_{g,t} w^{lp,\ell}_{g,t} < 0$ *for all* $\ell \geq 1$.

Equation (4) shows that for all $\ell \in \{1, ..., K\}$, $\widehat{\beta}_{lp,\ell}$ estimates a weighted sum, across (g,t) cells, of $\Delta_{g,t+\ell}(t+\ell-F_g+1, I_g)$, the effect of having been exposed to I_g units of treatment for $t+\ell-F_g+1$ periods. For (g,t)s such that group g started receiving the treatment at t, $F_g = t$ so the effect of $\ell+1$ periods of exposure enters in $\widehat{\beta}_{lp,\ell}$. But for (g,t)s such that group g started receiving the treatment at $t-1$, $F_g = t-1$ so the effect of $\ell+2$ periods of exposure enters in $\widehat{\beta}_{lp,\ell}$, etc. Accordingly, $\widehat{\beta}_{lp,\ell}$ does not estimate an average across groups of the effect of $\ell+1$ periods of exposure. Intuitively, this issue arises because groups with $D_{g,t} = 1$ may have started receiving the treatment before period t, while some groups with $D_{g,t} = 0$ may have started receiving it between periods $t+1$ and $t+\ell$. Perhaps more worryingly, for $\ell \geq 1$ the weights $w^{lp,\ell}_{g,t}$ do not sum to one, unless

$$\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g > t} N_g(-D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.}) = 0,$$

a condition that is very unlikely to hold in practice. This implies that even if the treatment does not have dynamic effects and the effect of the contemporaneous treatment is homogeneous between groups and over time (i.e.: $\Delta_{g,t}(k,x) = \Delta_{g,t}(1,x) = \theta$), $E\left(\widehat{\beta}_{lp,\ell}\right) \neq \theta$. This is because the regression is misspecified: it considers groups with $D_{g,t} = 0$ as untreated, while some of them may actually be treated at $t+\ell$. Finally, if the treatment is binary, some of the weights $w^{lp,\ell}_{g,t}$ are negative. We suspect this will also often be the case with a non-binary treatment, though establishing a formal result of that kind seems elusive.

Note that the issues we highlight in Proposition 3 arise with panel data and variation in treatment timing. Those issues are absent with time-series data where the unit receives a single treatment shock, as considered by Jordà (2005). Finally, Proposition 3 assumes that the treatment is potentially non-binary but staggered ($D_{g,t} = I_g 1\{t \geq F_g\}$). We conjecture that similar and perhaps more negative results can be obtained in more complicated designs. The proof of Proposition 3 exploits similar ideas as the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020*a*), and leverages specific features of Regression 3.

4 Parameters of interest, estimators, and placebo estimators

In this section, we define our parameters of interest and our estimators. To simplify the exposition, we consider the case of a binary treatment: $\mathcal{D} = \{0, 1\}$. Our results can be generalized to discrete treatments, and to continuous and staggered treatments such that $D_{g,t} = I_g 1\{t \geq F_g\}$.[7] We discuss those extensions at the end of this section, and formally cover them in Sections 2 and 3 of the Web Appendix. Below, we define parameters of interest and estimators for initially-untreated groups. Parameters of interest and estimators for initially-treated groups can be defined symmetrically. Again, they are discussed at the end of this section, and formally defined in Section 1 of the Web Appendix.

4.1 Parameters of interest, for initially-untreated groups

We take the perspective of a social planner, seeking to conduct a cost-benefit analysis comparing groups' actual treatments to the counterfactual "status-quo" scenario where every group would have kept throughout the same treatment as in period 1. In other words, the planner wants to know if the treatment/policy changes that took place over the duration of the panel led to a better situation than the one that would have prevailed if no policy change had been undertaken, an arguably natural policy question. As the planner wants to compare groups' actual treatments $\boldsymbol{D}$ to the status-quo treatments, our parameters of interest and all of our analysis are conditional on $\boldsymbol{D}$. Unconditional versions of our target parameters would compare the actual and status-quo treatments, in expectation across all actual treatment paths that groups may have experienced. In observational studies where the researcher has no control or knowledge of the assignment mechanism, this parameter may be harder to interpret than the conditional one. Conditioning on the design is, of course, immaterial for all our unbiasedness results: the estimators we propose below are conditionally unbiased for our parameters of interest, so they are also unconditionally unbiased. Conditional on the design, potential outcomes are the only source of randomness left. Expectations below are taken with respect to their probability distribution.[8]

For any $g \in \{1, ..., G\}$, let $F_g = \min\{t : t \geq 2, D_{g,t} \neq D_{g,t-1}\}$ denote the first period at which group g's treatment changes, with the convention that $F_g = T + 1$ if group g never changes treatment. In Assumption 4 below, we require that at least one group goes from untreated to treated at a period where another group has been untreated all along. This only rules out pathological designs, where groups untreated at period 1 all remain untreated till period T or all get treated for the first time at the same period.

[7] de Chaisemartin et al. (2022) extend our estimators to non-staggered, continuous treatments.

[8] While Propositions 1-3 above are shown unconditionnally, those results also hold conditional on $\boldsymbol{D}$, so there is no asymmetry between those results and those below.

Assumption 4 *(Non-pathological design)* $\exists(g,g') \in \{1,...,G\}^2$ *such that* $D_{g,1} = D_{g',1} = 0$ *and* $F_g < F_{g'}$.

Let $T_u = \max_{g:D_{g,1}=0} F_g - 1$ denote the last period where there is still a group that has been untreated since period 1. T_u is the last period where initially-untreated groups' treatment effects can be estimated under our assumptions: after T_u, there are no never treated groups that can be used as controls. For any initially-untreated group g such that $F_g \leq T_u$, and for any $\ell \in \{0,...,T_u - F_g\}$, let

$$\begin{aligned}\delta_{g,\ell} =& E(Y_{g,F_g+\ell}(\boldsymbol{D}_g) - Y_{g,F_g+\ell}(\mathbf{0})|\boldsymbol{D}) \\ =& E(Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, 1, D_{g,F_g+1}, ..., D_{g,F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})|\boldsymbol{D}),\end{aligned}$$

where the second equality follows from Assumption 2 and the definition of F_g. $\delta_{g,\ell}$ is the expected difference between group g's actual outcome at period $F_g+\ell$, ℓ periods after it got treated for the first time, and the counterfactual "status quo" outcome it would have obtained at that period if it had remained untreated from period one to $F_g + \ell$. In staggered designs, $\delta_{g,\ell}$ is the effect of having been treated rather than untreated for $\ell + 1$ periods. In non-staggered designs, $\delta_{g,\ell}$ is just the effect of having switched from untreated to treated for the first time ℓ periods ago. Depending on g's treatments after period F_g, this may correspond to the effect of having been treated rather than untreated from at least one to at most $\ell + 1$ periods.

We can now define our main parameter of interest. Let $\beta \in (0,1]$ denote the planner's discount rate. We assume that the outcome can be converted into a monetary equivalent, as is for instance the case with income, test scores, etc. Then, $\delta_{g,\ell}$ is the expected monetary benefit or loss, in group g and at period $F_g+\ell$, of having received the actual rather than the status-quo treatments since period F_g, the first period when group g deviated from its status-quo treatment. We also assume that the treatment is costly, and we let $c_{g,\ell} \geq 0$ denote the cost to treat one observation in group g at period $F_g+\ell$. To simplify, we assume that $c_{g,\ell}$ is non-stochastic, and that it can be readily computed, for instance using the accounts of the organization delivering the treatment.[9] In groups untreated at period 1, the actual treatments are beneficial relative to the status quo, up to period T_u, if and only if

$$\sum_{g:D_{g,1}=0,F_g\leq T_u} \sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}\delta_{g,\ell} - \sum_{g:D_{g,1}=0,F_g\leq T_u} \sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}D_{g,F_g+\ell} > 0. \quad (5)$$

Let

$$c_+ = \frac{\sum_{g:D_{g,1}=0,F_g\leq T_u} \sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}D_{g,F_g+\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u} \sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}D_{g,F_g+\ell}}$$

[9]The treatment cost may have to be estimated, by comparing the actual expenses of the organization delivering the treatment to its counterfactual expenses had it not delivered the treatment. In that case, c_+ defined below can be estimated, using the estimators we propose, but defining the outcome as the expenses of the organization delivering the treatment in group g at period t.

denote the average discounted cost of the treatment, across all the discounted treatments received by groups initially untreated. Assumption 4 ensures that c_+'s denominator is strictly positive. Dividing the left- and right-hand sides of (5) by that denominator, we finally get that the actual treatments are beneficial up to period T_u if and only if

$$\delta_+ := \frac{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}\delta_{g,\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}D_{g,F_g+\ell}} > c_+.$$

Our main parameter of interest is δ_+. By comparing δ_+ to c_+, the planner can assess if groups' actual treatments dominate the status quo.

δ_+ may be interpreted as an average total effect per unit of treatment. To see this, let us first consider a simple example with two groups and four periods, such that $N_{g,t}=1$, and almost surely $D_{1,1}=0$, $D_{1,2}=1$, $D_{1,3}=1$, and $D_{1,4}=0$, while group 2 is never treated. Then, if $\beta=1$,

$$\begin{aligned}\delta_+ &= E\left(\frac{Y_{1,2}(0,1)-Y_{1,2}(0,0)+Y_{1,3}(0,1,1)-Y_{1,3}(0,0,0)+Y_{1,4}(0,1,1,0)-Y_{1,4}(0,0,0,0)}{1+1+0}\right)\\ &= \frac{1}{2}E\left(Y_{1,2}(0,1)-Y_{1,2}(0,0)+Y_{1,3}(0,1,0)-Y_{1,3}(0,0,0)+Y_{1,4}(0,1,0,0)-Y_{1,4}(0,0,0,0)\right)\\ &+ \frac{1}{2}E\left(Y_{1,3}(0,1,1)-Y_{1,3}(0,1,0)+Y_{1,4}(0,1,1,0)-Y_{1,4}(0,1,0,0)\right). \qquad (6)\end{aligned}$$

The first expectation in (6) is the total effect produced by group 1's period-2 treatment, at periods 2, 3, and 4, relative to the situation where it would have always remained untreated. The second expectation in (6) is the total effect produced by group 1's period-3 treatment, at periods 3 and 4, conditional on its period-2 treatment and relative to the situation where it would have been untreated at periods 3 and 4. Accordingly, δ_+ is the average total effect of those two treatments.

A similar interpretation holds beyond this simple example. Let us define, for $k\in\{0,...,\ell\}$,

$$\begin{aligned}\delta_{g,k,\ell} = E(&Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, D_{g,F_g}, ..., D_{g,F_g+k}, \mathbf{0}_{\ell-k})\\ &-Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, D_{g,F_g}, ..., D_{g,F_g+k-1}, \mathbf{0}_{\ell-(k-1)})|\boldsymbol{D}),\end{aligned}$$

with the convention that $\delta_{g,0,\ell} = E(Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, D_{g,F_g}, \mathbf{0}_\ell) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})|\boldsymbol{D})$. $\delta_{g,k,\ell}$ is the effect, at period $F_g+\ell$, of switching g's period F_g+k treatment from 0 to D_{g,F_g+k}, while keeping its period-one-to-F_g+k-1 treatments at their actual values, and its period-F_g+k+1-to-$F_g+\ell$ treatments at 0. Observe that $\delta_{g,\ell}=\sum_{k=0}^{\ell}\delta_{g,k,\ell}$. Therefore, if $\beta=1$,

$$\begin{aligned}\delta_+ =&\frac{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\sum_{k=0}^{\ell}\delta_{g,k,\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}D_{g,F_g+\ell}}\\ =&\frac{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{k=0}^{T_u-F_g}\sum_{\ell=k}^{T_u-F_g} N_{g,F_g+\ell}\delta_{g,k,\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{k=0}^{T_u-F_g} N_{g,F_g+k}D_{g,F_g+k}}. \qquad (7)\end{aligned}$$

$\sum_{\ell=k}^{T_u-F_g} N_{g,F_g+\ell}\delta_{g,k,\ell}$ is the total effect, from period F_g+k to T_u, of switching g's period F_g+k treatment from 0 to D_{g,F_g+k}. The sum of total effects in the numerator of δ_+ is scaled by the sum of all the treatments D_{g,F_g+k}. Accordingly, δ_+ may be interpreted as an average total effect per unit of treatment. Of course, the horizons over which the total effects are cumulated are conditional on the design, and vary across the treatments D_{g,F_g+k}. To help interpret δ_+, one may compute the average, across all (g,k) such that $D_{g,F_g+k}=1$, of T_u-F_g-k+1, the number of effects cumulated in $\sum_{\ell=k}^{T_u-F_g} N_{g,F_g+\ell}\delta_{g,k,\ell}$: δ_+'s interpretation may differ if that parameter sums the effects of a treatment across 2 periods on average, and if it sums the effects of a treatment across 10 periods on average.

Beyond this general interpretation, δ_+ reduces to a standard average treatment effect in some special cases. For instance, if $T=2$,

$$\delta_+ = E\left(\frac{1}{N_+}\sum_{(i,g):D_{i,g,2}>D_{i,g,1}}(Y_{i,g,t}(0,1)-Y_{i,g,t}(0,0))\middle|\boldsymbol{D}\right),$$

where $N_+=\#\{(i,g):D_{i,g,2}>D_{i,g,1}\}$. δ_+ is the average effect of having been treated rather than untreated at period 2, among all observations going from untreated to treated from period 1 to 2. Then, δ_+ is similar to the switcher's average treatment effect considered in de Chaisemartin and D'Haultfœuille (2020a). Another special case where δ_+ may reduce to a standard average treatment effect are staggered designs. Then, if $\beta=1$, no groups are always treated, and $T_u=T$,

$$\delta_+ = E\left(\frac{1}{\sum_{(i,g,t)}D_{i,g,t}}\sum_{(i,g,t):D_{i,g,t}=1}\left(Y_{i,g,t}\left(\mathbf{0}_{F_g-1},\mathbf{1}_{t-F_g+1}\right)-Y_{i,g,t}\left(\mathbf{0}_t\right)\right)\middle|\boldsymbol{D}\right),$$

so δ_+ is the average of all the instantaneous and dynamic treatment effects, across treated units.

We now show that δ_+ is actually equal to a weighted average of event-study reduced-form effects, divided by a weighted average of event-study first-stage effects. Let $L_u=T_u-\min_{g:D_{g,1}=0}F_g$ denote the difference between the last period at which a group has been untreated all along and the first period where a group goes from untreated to treated. Under Assumption 4, $L_u\geq 0$. For every $\ell\in\{0,...,L_u\}$, let $N_\ell^1=\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell}>0$ be the discounted number of units in groups reaching ℓ periods after their first treatment at or before T_u. Let

$$\begin{aligned}\delta_{+,\ell} &= \sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N_\ell^1}\delta_{g,\ell} \qquad (8)\\ \delta_{+,\ell}^D &= \sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N_\ell^1}D_{g,F_g+\ell}.\end{aligned}$$

In staggered designs, $\delta_{+,\ell}$ is the average cumulative effect of $\ell+1$ treatment periods across all groups reaching $\ell+1$ treatment periods before T_u. Outside of staggered designs, $\delta_{+,\ell}$ is just the average effect of having switched treatment for the first time ℓ periods ago, across all initially-untreated groups that got treated for the first time at least ℓ periods before T_u. Irrespective of

the design, $\delta_{+,\ell}$ is always an effect of having experienced a weakly higher amount of treatment for $\ell+1$ periods. $\delta^D_{+,\ell}$ is simply the average treatment of groups that got treated for the first time ℓ periods ago. Finally, let $w_{+,\ell} = N^1_\ell / \sum_{\ell'=0}^{L_u} N^1_{\ell'}$.

Lemma 1 *If Assumptions 1 and 4 hold,*

$$\delta_+ = \frac{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta_{+,\ell}}{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta^D_{+,\ell}}. \tag{9}$$

Proof of Lemma 1: We have

$$\begin{aligned}
\frac{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta_{+,\ell}}{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta^D_{+,\ell}} =& \frac{\sum_{\ell=0}^{L_u}\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell}\delta_{g,\ell}}{\sum_{\ell=0}^{L_u}\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell}D_{g,F_g+\ell}} \\
=& \frac{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g}\beta^{F_g+\ell}N_{g,F_g+\ell}\delta_{g,\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g}\beta^{F_g+\ell}N_{g,F_g+\ell}D_{g,F_g+\ell}} \\
=& \delta_+.
\end{aligned}$$

The first equality follows from the definitions of $w_{+,\ell}$, $\delta_{+,\ell}$, and $\delta^D_{+,\ell}$. The second equality follows from the fact that by the definitions of F_g, T_u, and L_u,

$$\{(g,\ell) : 0\leq\ell\leq L_u, D_{g,1}=0, F_g\leq T_u-\ell\} = \{(g,\ell) : 0\leq\ell\leq T_u-F_g, D_{g,1}=0, F_g\leq T_u\} \ \square$$

Lemma 1 shows that δ_+, a parameter with a clear economic interpretation, is equal to a weighted average of the event-study reduced-form effects $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$, divided by a weighted average of the event-study first-stage effects $(\delta^D_{+,\ell})_{0\leq\ell\leq L_u}$.

Following Lemma 1, one may also define parameters of interest averaging the reduced-form and first-stage effects over a shorter horizon than δ_+. For every $k\in\{0,...,L_u\}$, let

$$\delta_{+,0:k} = \frac{\sum_{\ell=0}^{k} w_{+,k,\ell}\delta_{+,\ell}}{\sum_{\ell=0}^{k} w_{+,k,\ell}\delta^D_{+,\ell}}, \tag{10}$$

where $w_{+,k,\ell} = N^1_\ell / \sum_{\ell'=0}^{k} N^1_{\ell'}$. $\delta_{+,0:k}$ only takes into account the reduced-form and first-stage effects up to k periods after the first switch. For values of ℓ close to L_u, the estimators of the reduced-form effects $\delta_{+,\ell}$ may sometimes be very noisy: in the last periods before T_u, there may be very few groups left that have been untreated all along and can be used as controls. Then, trimming those last effects, as $\delta_{+,0:k}$ does, may be warranted to increase precision. Proposing a principled rule to determine the number of effects to be trimmed goes beyond the scope of this paper. We refer the reader to de Chaisemartin (2021) for a proposal in that direction.

Finally, the reduced-form effects $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ can provide useful information on simple average treatment effects, under some conditions. Specifically, assume that for all (g,t), $Y_{g,t}(d_1,...,d_t)$ is

increasing in each of its arguments. Then, if $\beta = 1$,

$$0 \leq \sum_{g:D_{g,1}=0,F_g\leq T_u-\ell} \frac{N_{g,F_g+\ell}}{N^1_\ell} E(Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, 1, \mathbf{0}_\ell) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})|\boldsymbol{D})$$
$$\leq \delta_{+,\ell} \leq \sum_{g:D_{g,1}=0,F_g\leq T_u-\ell} \frac{N_{g,F_g+\ell}}{N^1_\ell} E(Y_{g,F_g+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_{\ell+1}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})|\boldsymbol{D}). \quad (11)$$

The left-hand side in the inequality is the average effect, for groups whose treatment changes for the first time before $T_u - \ell$, of having been treated rather than untreated at period F_g. The right-hand side is the average effect, for the same groups, of having been treated from period F_g to $F_g + \ell$. Accordingly, $\delta_{+,\ell}$ is an upper bound of an average effect of having been treated for 1 period, and a lower bound of an average effect of having been treated for $\ell + 1$ periods. A similar bracketing result holds if $Y_{g,t}(d_1, ..., d_t)$ is decreasing in each of its arguments, with the inequalities reversed. These bracketing results also imply that $\delta_{+,\ell}$ satisfies the no-sign-reversal property, unlike the regression coefficients considered in Section 3. If $Y_{g,t}(d_1, ..., d_t)$ is increasing in each of its arguments, the parameters $\Delta_{g,\ell}(\ell - (F-1), I_g)$, $\gamma^\ell_{g,t}$, and $\Delta_{g,t+\ell}(t + \ell - F_g + 1, I_g)$ in Propositions 1-3 are all positive, but the expectations of $\widehat{\beta}_{fe,\ell}$, $\widehat{\beta}_{dl,\ell}$, and $\widehat{\beta}_{lp,\ell}$ could still be negative if there are negative weights in their decompositions in Propositions 1-3. On the other hand, Equation (11) shows that if for every (g, t) $Y_{g,t}(d_1, ..., d_t)$ is increasing in each of its argument, $\delta_{+,\ell} \geq 0$: $\delta_{+,\ell}$ preserves the sign of the treatment effect if that sign is the same for all (g, t) cells, and for all the current and past treatments.

4.2 Estimators

We start by proposing conditionally unbiased estimators of the $\delta_{g,\ell}$ parameters. Let $N^u_t = \sum_{g:D_{g,1}=0,F_g>t} N_{g,t}$ denote the number of observations at period t in groups untreated from period 1 to t. For every g such that $D_{g,1} = 0$ and $F_g \leq T_u$ and every $\ell \in \{0, ..., T_u - F_g\}$, note that $N^u_{F_g+\ell} > 0$. Then, let

$$\text{DID}_{g,\ell} = Y_{g,F_g+\ell} - Y_{g,F_g-1} - \sum_{g':D_{g',1}=0,F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N^u_{F_g+\ell}} (Y_{g',F_g+\ell} - Y_{g',F_g-1}).$$

$\text{DID}_{g,\ell}$ compares the $F_g - 1$-to-$F_g + \ell$ outcome evolution, in group g and in groups untreated from period 1 to $F_g + \ell$.

Lemma 2 *If Assumptions 1-3 hold, then for every $g : D_{g,1} = 0, F_g \leq T_u$ and every $\ell \in \{0, ..., T_u - F_g\}$, $E[DID_{g,\ell}|\boldsymbol{D}] = \delta_{g,\ell}$.*

Proof of Lemma 2: We have

$$
\begin{aligned}
&E\left[\mathrm{DID}_{g,\ell}|\boldsymbol{D}\right]\\
=&E\left[Y_{g,F_g+\ell}-Y_{g,F_g-1}(\mathbf{0})\Big|\boldsymbol{D}\right]-E\left[\sum_{g':D_{g',1}=0,F_{g'}>F_g+\ell}\frac{N_{g',F_g+\ell}}{N^u_{F_g+\ell}}(Y_{g',F_g+\ell}(\mathbf{0})-Y_{g',F_g-1}(\mathbf{0}))\Bigg|\boldsymbol{D}\right]\\
=&\delta_{g,\ell}+E\left[Y_{g,F_g+\ell}(\mathbf{0})-Y_{g,F_g-1}(\mathbf{0})\Big|\boldsymbol{D}\right]-E\left[\sum_{g':D_{g',1}=0,F_{g'}>F_g+\ell}\frac{N_{g',F_g+\ell}}{N^u_{F_g+\ell}}(Y_{g',F_g+\ell}(\mathbf{0})-Y_{g',F_g-1}(\mathbf{0}))\Bigg|\boldsymbol{D}\right]\\
=&\delta_{g,\ell}.
\end{aligned}
$$

The first equality follows from the definition of F_g and Assumption 2. The second equality follows from adding and subtracting $Y_{g,F_g+\ell}(\mathbf{0})$ and from the definition of $\delta_{g,\ell}$. The third equality follows from Assumption 3 and the definition of $N^u_{F_g+\ell}$ □

Then, we propose conditionally unbiased estimators of $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$, by replacing the $\delta_{g,\ell}$ parameters by their estimators in Equation (8). Specifically, for every $\ell\in\{0,...,L_u\}$, let

$$\mathrm{DID}_{+,\ell}=\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N^1_\ell}\mathrm{DID}_{g,\ell}. \tag{12}$$

Finally, we propose a conditionally unbiased estimator of δ_+, by replacing the $\delta_{+,\ell}$ parameters by their estimators in Equation (9). Specifically, let

$$\widehat{\delta}_+=\frac{\sum_{\ell=0}^{L_u}w_{+,\ell}\mathrm{DID}_{+,\ell}}{\sum_{\ell=0}^{L_u}w_{+,\ell}\delta^D_{+,\ell}}.$$

Our estimators' unbiasedness, stated in Theorem 1 below, directly follows from the fact that $(F_g)_{1\leq g\leq G}$, T_u, $(N^1_\ell)_{0\leq\ell\leq L_u}$, $(w_{+,\ell})_{0\leq\ell\leq L_u}$, and $(\delta^D_{+,\ell})_{0\leq\ell\leq L_u}$ are functions of $\boldsymbol{D}$, from the linearity of the conditional expectation operator, and from Lemma 2 and Equations (8) and (9).

Theorem 1 *If Assumptions 1-4 hold, then for all $\ell\in\{0,...,L_u\}$, we have $E\left[DID_{+,\ell}|\boldsymbol{D}\right]=\delta_{+,\ell}$ and $E\left[\widehat{\delta}_+|\boldsymbol{D}\right]=\delta_+$.*

A few comments on our estimators are in order. First, the fact that the IV-like estimator $\widehat{\delta}_+$ is conditionally and unconditionally unbiased may be surprising. This comes from the fact that we, essentially, make a strong-exogeneity assumption (see Point 2 below Assumption 3). In our derivation of $E\left[\widehat{\delta}_+|\boldsymbol{D}\right]$, this allows us to pull out the denominator of $\widehat{\delta}_+$ before invoking the parallel trends assumption.[10]

Second, $\mathrm{DID}_{g,\ell}$ uses groups' F_g-1 outcome as the baseline. Instead, one could average their period-1-to-F_g-1 outcomes, thus giving rise to the following estimator:

$$\tilde{\mathrm{DID}}_{g,\ell}=Y_{g,F_g+\ell}-\frac{1}{F_g-1}\sum_{t=1}^{F_g-1}Y_{g,t}-\sum_{g':D_{g',1}=0,F_{g'}>F_g+\ell}\frac{N_{g',F_g+\ell}}{N^u_{F_g+\ell}}\left(Y_{g',F_g+\ell}-\frac{1}{F_g-1}\sum_{t=1}^{F_g-1}Y_{g',t}\right).$$

[10] In a standard cross-sectional IV, the equivalent of this strong-exogeneity assumption would amount to assuming that potential outcomes do not depend on potential treatments, an often implausible assumption.

Under Assumptions 2-3, $\tilde{\text{DID}}_{g,\ell}$ is also unbiased for $\delta_{g,\ell}$, and it may be more precise than $\text{DID}_{g,\ell}$. We prefer to use $\text{DID}_{g,\ell}$, for two reasons. First, in the next section we propose a placebo estimator $\text{DID}^{\text{pl}}_{g,\ell}$ symmetric to $\text{DID}_{g,\ell}$. Proposing a placebo symmetric to $\tilde{\text{DID}}_{g,\ell}$ is not possible, precisely because $\tilde{\text{DID}}_{g,\ell}$ leverages all pre-treatment periods. As we discuss in further details in the next section, the symmetry between our actual and placebo estimators makes them readily amenable to the sensitivity analysis proposed by Rambachan and Roth (2019), that one may use to assess the sensitivity of one's results to violations of Assumption 3. Second, if Assumption 3 does not exactly hold and the discrepancy between groups' trends gets larger over longer horizons, as would for instance happen when there are group-specific linear trends, leveraging earlier pre-treatment periods increases the bias of a DID estimator (see Roth, forthcoming). Then, $\tilde{\text{DID}}_{g,\ell}$ is more biased than $\text{DID}_{g,\ell}$. $\tilde{\text{DID}}_{g,\ell}$ could be less biased than $\text{DID}_{g,\ell}$ if Assumption 3 holds but Assumption 2 fails due to anticipation effects arising a few periods before a group gets treated. However, one may immunize $\text{DID}_{g,\ell}$ against anticipation effects, by redefining F_g as the date when it was announced that group g would receive the treatment for the first time. By contrast, it may be more difficult to immunize $\tilde{\text{DID}}_{g,\ell}$ against differential trends widening over time.

A third remark on $\text{DID}_{g,\ell}$ is that it can easily be extended to implement a triple-difference estimation strategy. Assume each (g,t) cell can be divided into two subgroups, one that is eligible for treatment and that will receive it if cell (g,t) is treated, and another subgroup that is ineligible and remains untreated even when cell (g,t) is treated. For instance, it could be that females are eligible for the treatment while males are ineligibles. If one wants to leverage this feature to implement a triple-difference estimator, one can just redefine the outcome $Y_{g,t}$ as the difference between the average outcomes of the eligible and ineligible subgroups in cell (g,t). Instead of parallel trends, the resulting estimators rely on the assumption that differential trends across groups over time are the same in the eligible and ineligible groups.

Finally, for every $k \in \{0, ..., L_u\}$, let

$$\widehat{\delta}_{+,0:k} = \frac{\sum_{\ell=0}^{k} w_{+,k,\ell} \text{DID}_{+,\ell}}{\sum_{\ell=0}^{k} w_{+,k,\ell} \delta^{D}_{+,\ell}}$$

be estimators of the trimmed parameters $\delta_{+,0:k}$. It directly follows from (10) and Theorem 1 that those estimators are conditionally unbiased.

4.3 Placebo estimators

We now propose placebo estimators of Assumptions 2 and 3. First, for any g such that $D_{g,1} = 0$ and $3 \leq F_g \leq T_u$ and for any $\ell \in \{0, ..., \min(T_u - F_g, F_g - 3)\}$, let

$$\text{DID}^{\text{pl}}_{g,\ell} = Y_{g,F_g-\ell-2} - Y_{g,F_g-1} - \sum_{g':D_{g',1}=0, F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N^{u}_{F_g+\ell}} (Y_{g',F_g-\ell-2} - Y_{g',F_g-1}).$$

$\text{DID}^{\text{pl}}_{g,\ell}$ is a placebo estimator mimicking $\text{DID}_{g,\ell}$. Like $\text{DID}_{g,\ell}$, it compares group g's outcome evolution to that of groups untreated from period one to $F_g+\ell$. But unlike $\text{DID}_{g,\ell}$, it compares those groups' outcome evolutions from period $F_g-\ell-2$ to period F_g-1, namely before group g gets treated for the first time. Accordingly, $\text{DID}^{\text{pl}}_{g,\ell}$ assesses if g and groups not yet treated at $F_g+\ell$ are on parallel trends when untreated, for $\ell+1$ periods, the number of periods over which parallel trends has to hold for $\text{DID}_{g,\ell}$ to be unbiased. Note that while $\text{DID}_{g,\ell}$ goes from the past (period F_g-1) to the future (period $F_g+\ell$), $\text{DID}^{\text{pl}}_{g,\ell}$ goes from the future (period F_g-1) to the past (period $F_g-\ell-2$). This is to follow the standard practice in event-study regressions, where the reference period is also the one before treatment.

Then, let $L^{\text{pl}}_u=\max_{g:D_{g,1}=0}\{\min(T_u-F_g,F_g-3)\}$. L^{pl}_u is the largest ℓ such that there is a group for which the effect of switching from untreated to treated for the first time ℓ periods ago can be estimated ($\ell\leq T_u-F_g$), and for which one can form a placebo estimator comparing that group's outcome evolution to that of the not-yet-treated groups over the $\ell+1$ periods before it first got treated ($F_g-\ell-2\geq 1\Leftrightarrow\ell\leq F_g-3$). One may have $L^{\text{pl}}_u=-1$, if all groups that switch from untreated to treated for the first time do so at period 2. Then, none of the placebos defined below can be computed, as one does not observe the outcome evolution of any group before it switches from untreated to treated. Outside of that special case, $L^{\text{pl}}_u\geq 0$, so one can at least compute one of the $\text{DID}^{\text{pl}}_{+,\ell}$ estimators below.

For all $\ell\in\{0,...,L^{\text{pl}}_u\}$, let $N^{1,\text{pl}}_\ell=\sum_{g:D_{g,1}=0,\ell+3\leq F_g\leq T_u-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell}>0$ be the discounted number of observations in groups for which both $\text{DID}_{g,\ell}$ and $\text{DID}^{\text{pl}}_{g,\ell}$ can be computed. Then, let

$$\text{DID}^{\text{pl}}_{+,\ell}=\sum_{g:D_{g,1}=0,\ell+3\leq F_g\leq T_u-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N^{1,\text{pl}}_\ell}\text{DID}^{\text{pl}}_{g,\ell}.$$

Theorem 2 *If Assumptions 1-4 hold and $L^{pl}_u\geq 0$, then $\forall\ell\in\{0,...,L^{pl}_u\}$, $E\left[DID^{pl}_{+,\ell}|\boldsymbol{D}\right]=0$.*

Theorem 2 shows that $E\left[\text{DID}^{\text{pl}}_{+,\ell}|\boldsymbol{D}\right]=0$ is a testable implication of Assumptions 2 and 3.

The placebo estimators $\text{DID}^{\text{pl}}_{+,\ell}$ mimick our actual estimators, but accordingly they do not exhaust all the testable implications of our assumptions. There are at least two advantages to having placebos mimicking the actual estimators. First, when we reject $E\left[\text{DID}^{\text{pl}}_{+,\ell}|\boldsymbol{D}\right]=0$, the value of $\text{DID}^{\text{pl}}_{+,\ell}$ may be used to sign $\text{DID}_{+,\ell}$'s bias. Specifically, assume that for all g such that $\text{DID}_{g,\ell}$ and $\text{DID}^{\text{pl}}_{g,\ell}$ can be computed, and for all g' such that $F_{g'}>F_g+\ell$, the sign of

$$E\left[Y_{g,F_g+\ell}(\mathbf{0})-Y_{g,F_g-1}(\mathbf{0})|\boldsymbol{D}\right]-E\left[Y_{g',F_g+\ell}(\mathbf{0})-Y_{g',F_g-1}(\mathbf{0})|\boldsymbol{D}\right]$$

does not depend on (g,g'). In words, the sign of the difference in trends between a switcher and a not-yet-treated group has to be the same, for every pair composed of a switcher and a not-yet-treated. Then, the sign of the bias of $\text{DID}_{+,\ell}$ is equal to the sign of $-E[\text{DID}^{\text{pl}}_{+,\ell}|\boldsymbol{D}]$. The second, related advantage of having placebo estimators mimicking the actual estimators is

that it makes them more amenable to the sensitivity analysis proposed by Rambachan and Roth (2019), which allows researchers to assess how sensitive their results are to violations of parallel trends. This approach assumes that trends are not exactly parallel, and that one has placebo estimators whose magnitude is informative of the magnitude of switchers and not-yet treated differential $Y_{g,t}(\mathbf{0})$-trends after the switchers get treated. This requirement is more plausible when placebos closely mimick the actual estimators.

The "long-difference" placebos we propose here differ from the "first-difference" ones we proposed in de Chaisemartin and D'Haultfœuille (2020*a*) and that we extend in the Web Appendix to allow for dynamic effects. The long-difference placebos test if trends are parallel over several periods, while the first-difference placebos test if trends are parallel over pairs of consecutive periods. If treated and untreated groups follow different linear trends, differential trends will be larger, and easier to detect, over several periods than over two consecutive periods. Then, the long-difference placebos may lead to a more powerful test of parallel trends. As discussed above, they may also be informative on the bias of $\text{DID}_{+,\ell}$. On the other hand, the first-difference placebos may be useful to specifically test Assumption 2, the no-anticipation assumption. Assume for instance that those placebos are statistically insignificant, except between the two last periods before the switching groups switch. This may suggest that Assumption 3 holds, but Assumption 2 fails. Then, as explained above, one may just recompute our estimators, redefining F_g as the date when it was announced that group g would receive the treatment for the first time.

4.4 Extensions

4.4.1 Estimators for initially-treated groups

In Section 1 of the Web Appendix, we consider groups treated at the start of the panel. For those groups, the status-quo treatment is to always be treated, which is by definition always higher than their actual treatments. Accordingly, we define parameters of interest $\delta_{-,\ell}$ symmetric to $\delta_{+,\ell}$, except that $\delta_{-,\ell}$ is the average difference between initially-treated groups' status-quo and actual outcomes, rather than the opposite. Like $\delta_{+,\ell}$, $\delta_{-,\ell}$ is an average effect of having received a weakly higher amount of treatment for $\ell+1$ periods. As above, a weighted average of the reduced-form effects $\delta_{-,\ell}$ divided by a weighted average of the first-stage effects for initially-treated groups is equal to the cost-benefit ratio δ_- a planner may use to compare those groups' actual and status-quo treatments. Again, that parameter can also be interpreted as an average total effect per unit of treatment.

We first propose estimators $\text{DID}_{-,\ell}$ of the parameters $\delta_{-,\ell}$, which compare the F_g-1-to-$F_g+\ell$ outcome evolution of groups treated from period one to $F_g+\ell$ and groups untreated for the first time at period F_g. Those $\text{DID}_{-,\ell}$ estimators are unbiased under Assumption 2 and an assumption similar to Assumption 3, for the always-treated potential outcome. Those assumptions can be

tested, using placebo estimators similar to those defined above. We then define estimators $\widehat{\delta}_-$ of δ_- as a weighted sum of the $\text{DID}_{-,\ell}$ estimators.

Finally, we propose DID_ℓ estimators, which average the $\text{DID}_{+,\ell}$ and $\text{DID}_{-,\ell}$ estimators, with weights proportional to the number of observations each estimator applies to. The DID_ℓ estimator is unbiased for an average effect of having received a weakly higher amount of treatment for $\ell+1$ periods, across both initially-untreated and initially-treated groups. Similarly, we propose an estimator $\widehat{\delta}$, that aggregates $\widehat{\delta}_+$ and $\widehat{\delta}_-$, and estimates an average total effect per unit of treatment, across both initially-untreated and initially-treated groups.

4.4.2 Non-binary treatments

In Section 2 of the Web Appendix, we extend our analysis to discrete treatments. In that case, we let $\delta_{g,\ell} = E(Y_{g,F_g+\ell} - Y_{g,F_g+\ell}(D_{g,1},...,D_{g,1}))$ be the expected difference between group g's actual outcome at $F_g+\ell$ and the counterfactual "status quo" outcome it would have obtained if its treatment had remained equal to its period-one value from period one to $F_g+\ell$. Under parallel trends assumptions, we show that the $\delta_{g,\ell}$ parameters can be unbiasedly estimated by $\text{DID}_{g,\ell}$, a DID comparing g's outcome evolution from F_g-1 to $F_g+\ell$ to the same evolution among groups with the same treatment as g at period one, and whose treatment has not changed yet at period $F_g+\ell$.

There may be values of (g,ℓ) such that neither $\delta_{g,\ell}$ nor $-\delta_{g,\ell}$ are effects of having experienced a weakly higher amount of treatment for ℓ periods. Assume for instance that a group g_0 is such that $D_{g_0,1}=1, D_{g_0,2}=2, D_{g_0,3}=0$. Then,

$$\begin{aligned}\delta_{g_0,1} =& E(Y_{g_0,3}(1,2,0) - Y_{g_0,3}(1,1,1)) \\ =& E(Y_{g_0,3}(1,2,0) - Y_{g_0,3}(1,1,0)) - E(Y_{g_0,3}(1,1,1) - Y_{g_0,3}(1,1,0))\end{aligned}$$

is the difference between the effect of increasing g_0's period-2 treatment from 1 to 2, and the effect of increasing g_0's period-3 treatment from 0 to 1. One could have that both effects are positive but $\delta_{g_0,1}$ is negative, so $\delta_{g_0,1}$ does not satisfy the no-sign reversal property. Accordingly, we discard those (g,ℓ)s, and focus on the (g,ℓ)s such that from period one to $F_g+\ell$, group g's treatment is either always weakly higher than its period-one treatment, or always weakly lower. As in the binary-treatment case, those $\delta_{g,\ell}$s can be aggregated into reduced-form event-study effects measuring the effect of having experienced a weakly larger treatment for $\ell+1$ periods. Those reduced-form effects can then be aggregated into economically interpretable cost-benefit ratios. All those parameters can be estimated by linear combinations of the $\text{DID}_{g,\ell}$.

Finally, in Section 3 of the Web Appendix, we consider the case where the treatment is continuous and staggered, namely $D_{g,t} = I_g 1\{t \geq F_g\}$ for some group-specific intensities I_g. Our estimators for initially-untreated groups in the binary-treatment case can still be readily used in that case, though their interpretation slightly changes as explained therein.

4.4.3 Other extensions

Our web appendix also contains other extensions. First, we show how covariates $X_{g,t}$ can be included in our set-up. We replace Assumption 3 by the requirement that $E[Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}) - (X_{g,t} - X_{g,t-1})'\theta_0 | \boldsymbol{D}]$ does not vary across g, meaning that groups can experience differential trends provided those differential trends are fully explained by changes in their covariates. The idea, then, is to estimate θ_0 by regressing $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ in the sample of not-yet-treated (g,t)s, and define the same estimators of the $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ and δ_+ parameters as above, replacing $Y_{g,t} - Y_{g,t-1}$ by $Y_{g,t} - Y_{g,t-1} - (X_{g,t} - X_{g,t-1})'\widehat{\theta}_0$. Second, we show how to allow for different trends across sets of groups. This mirrors the common practice of, e.g., allowing for state-specific trends in county-level TWFE regressions. Third, we show that our results extend to fuzzy designs, provided some groups are fully untreated at $t = 1$. Fourth, we outline some of the benefits of ruling out the effect of past treatments beyond k lags, for some $k \geq 0$. For instance, under this assumption we propose a solution to the initial conditions problem. Fifth, we define first-difference placebo estimators.

5 Inference

In this section, we establish the asymptotic properties of $\text{DID}_{+,\ell}$ and $\widehat{\delta}_+$ when the number of groups tends to infinity while their size $N_{g,t}$ remains fixed.[11] We also propose confidence intervals for $\delta_{+,\ell}$ and δ_+, and show their asymptotic validity. To define these confidence intervals, let $N^1_{t,\ell} = \sum_{g:D_{g,1}=0, F_g=t-\ell} N_{g,t}$ and

$$U_{G,g,\ell} = \frac{G}{N^1_\ell} \sum_{t=\ell+2}^{T_u} \beta^t N_{g,t} \left[\mathbb{1}\{F_g = t-\ell\} - \frac{N^1_{t,\ell}}{N^u_t} \mathbb{1}\{F_g > t\} \right] (Y_{g,t} - Y_{g,t-\ell-1}),$$

$$U_{G,g} = \frac{\sum_{\ell=0}^{L_u} w_{+,\ell} U_{G,g,\ell}}{\sum_{\ell=0}^{L_u} w_{+,\ell} \text{DID}^D_{+,\ell}}.$$

Then, we define

$$\widehat{\sigma}^2_\ell = \frac{1}{G} \sum_{g:D_{g,1}=0} \left(U_{G,g,\ell} - \text{DID}_{+,\ell}\right)^2,$$

$$\widehat{\sigma}^2 = \frac{1}{G} \sum_{g:D_{g,1}=0} \left(U_{G,g} - \widehat{\delta}_+\right)^2.$$

The confidence intervals of nominal level $1-\alpha$ on $\delta_{+,\ell}$ and δ_+ that we consider are respectively

$$\text{CI}_{1-\alpha}(\delta_{+,\ell}) = \left[\text{DID}_{+,\ell} \pm z_{1-\alpha/2} \widehat{\sigma}_\ell G^{-1/2}\right],$$

$$\text{CI}_{1-\alpha}(\delta_+) = \left[\widehat{\delta}_+ \pm z_{1-\alpha/2} \widehat{\sigma} G^{-1/2}\right],$$

[11] Another possibility, following Conley and Taber (2011), would be to consider a fixed number of groups, and let the size of each group $N_{g,t}$ go to $+\infty$ under some spatial dependence restrictions.

where $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of the standard normal distribution.

Our asymptotic results are based on the three assumptions below. Hereafter, we let $\Sigma_g := V(Y_{g,1},...,Y_{g,T}|\boldsymbol{D})$ and for any symmetric semidefinite positive matrix Σ, $\underline{\rho}(\Sigma)$ denotes its smallest eigenvalue.

Assumption 5 *(Independent groups) The vectors $(Y_{g,1},...,Y_{g,T},\boldsymbol{D}_g)_{g\geq 1}$ are mutually independent.*

Assumption 6 *(Asymptotically non-pathological design) The support of $F_g|D_{g,1}=0$ does not depend on g and includes at least two values. Calling $\mathcal{S}$ this set, there exists $\underline{p}>0$ such that for all g and $t\in\mathcal{S}$, $Pr(D_{g,1}=0,F_g=t)\geq\underline{p}$.*

Assumption 7 *(Regularity conditions for asymptotic normality) We have, for some $\delta>0$,*

$$\sup_{g,t} N_{g,t}<\infty,\ \inf_{G,g\leq G}\underline{\rho}\left(\Sigma_g\right)>0\ \ a.s.\ \text{and}\ \sup_{G,g\leq G,\boldsymbol{d}\in\mathcal{D}} E[|Y_{g,t}|^{2+\delta}|\boldsymbol{D}=\boldsymbol{d}]<\infty\ \ a.s.$$

Assumption 5 is commonly made in DID studies, where standard errors are often clustered at the group level (Bertrand et al., 2004). Importantly, this assumption allows for serial correlation of the treatments and outcomes within each group. Assumption 6 may be seen as an asymptotic version of Assumption 4. It provides conditions on the treatment assignment ensuring that a growing number of "treatment" and "control" groups can be used to estimate each of the $\text{DID}_{+,\ell}$s. Finally, Assumption 7 ensures one can apply the Lyapunov central limit theorem in our set-up with independent but not identically distributed variables. The second condition therein prevents degenerate situations where the (non-trivial) linear combinations of the $(Y_{g,1},...,Y_{g,T})$ in $\text{DID}_{+,\ell}$ and $\widehat{\delta}_+$ would actually be constant.

Theorem 3 *Suppose that Assumptions 1-3 and 5-7 hold. Then, conditional on $\boldsymbol{D}$ and almost surely, for all $\ell\in\{0,...,L_u\}$,*

$$\sqrt{G}\frac{DID_{+,\ell}-\delta_{+,\ell}}{\left(\frac{1}{G}\sum_{g:D_{g,1}=0}V(U_{G,g,\ell}|\boldsymbol{D})\right)^{1/2}}\xrightarrow{d}\mathcal{N}(0,1), \tag{13}$$

$$\sqrt{G}\frac{\widehat{\delta}_+-\delta_+}{\left(\frac{1}{G}\sum_{g:D_{g,1}=0}V(U_{G,g}|\boldsymbol{D})\right)^{1/2}}\xrightarrow{d}\mathcal{N}(0,1). \tag{14}$$

Moreover, we have, almost surely,

$$\liminf_{G\to\infty}\Pr\left[\delta_{+,\ell}\in CI_{1-\alpha}(\delta_{+,\ell})|\boldsymbol{D}\right]\geq 1-\alpha,$$
$$\liminf_{G\to\infty}\Pr\left[\delta_{+}\in CI_{1-\alpha}(\delta_{+})|\boldsymbol{D}\right]\geq 1-\alpha.$$

As in Section 4, we reason conditional on the random vectors $\boldsymbol{D} = (\boldsymbol{D}_g)_{g\geq 1}$ here. Specifically, Theorem 3 shows that for almost all sequences $\boldsymbol{D}$, $\text{DID}_{+,\ell}$ and $\widehat{\delta}_+$ are asymptotically normal. The second part of the theorem ensures that confidence intervals are asymptotically conservative. If groups are identically distributed, the inequalities become equalities and the confidence intervals reach their nominal level asymptotically.

Finally, the results of Theorem 3 would easily extend to the estimators $(\widehat{\delta}_{+,0:k})_{k\in\{0,\ldots,L_u\}}$ of the trimmed parameters $(\delta_{+,0:k})_{k\in\{0,\ldots,L_u\}}$.

6 Application to banking deregulations and the housing market

In 1994, the Interstate Banking and Branching Efficiency Act (IBBEA) allowed US Banks to operate across state borders without formal authorization from state authorities. Initially, all states still imposed restrictions on:

- de novo branching without explicit agreement by state authorities;
- the minimum age of the target institution in case of mergers;
- the acquisition of individual branches without acquiring the entire bank;
- the total amount of statewide deposits controlled by a single bank or bank holding company.

Rice and Strahan (2010) compute the number of restrictions in place in every state and in each year from 1994 to 2005. Favara and Imbs (2015) reverse their index, so it can be interpreted as the number of restrictions lifted, ranging from 0 to 4. They then use 1994-to-2005 county-level data to estimate the effect of the number of regulations lifted on the volume of mortgages originated by banks and on housing prices, using a panel data version of the local-projection regressions proposed by Jordà (2005).[12] We use their data to revisit the same questions using our estimators.[13] The treatment is at the state×year level, while mortgage volume and housing prices are at the county×year level. All our estimators are computed with a discount factor of 1. We follow Favara and Imbs (2015) and cluster standard errors at the state level.

The treatment is equal to 0 in every state in 1994. Since no state had conducted any deregulation prior to 1994, the initial conditions problem is not an issue in this application. There are eight states that never deregulate and that can be used as controls till the end of the panel. Of the 42 states that lift at least one restriction from 1994 to 2005, 38 do so for the first time in 1995,

[12]The aggregation of these qualitatively different deregulations into a scalar variable seems uncontroversial in the finance literature.

[13]Favara and Imbs (2015) also study the effect of banking deregulation on other measures of banks' credit activities (number of loans originated, etc.). We obtain similar results on those other measures.

1996, 1997, or 1998. Accordingly, up to $\ell = 2005 - 1998 = 7$, the $\delta_{+,\ell}$ parameters apply to almost the same counties: $\delta_{+,0}$ and $\delta_{+,7}$ respectively apply to 821 and 706 counties.[14] $\delta_{+,8}$, $\delta_{+,9}$, and $\delta_{+,10}$ on the other hand only apply to a smaller subsample of counties (348, 234, and 1 county, respectively). Hence, we do not report the estimates of those parameters. For all the outcomes we consider, the estimates of $\delta_{+,8}$ and $\delta_{+,9}$ are even larger than those of $\delta_{+,7}$. We also focus on the trimmed cost-benefit parameter $\delta_{+,0:7}$ instead of δ_+, which means that we account for treatment effects up to 7 years after the first switch. Similarly, six placebo estimators can be computed, but only four apply to more than 50% of the 821 counties whose treatment changes at least once. The fifth and sixth placebos only apply to 115 and 69 counties respectively, so we do not report them.[15]

We start by showing the estimates of the $\delta_{+,\ell}$ parameters. To the right of zero, the blue line on Figure 1 below shows the $\text{DID}_{+,\ell}$ estimates of the effects of a first banking-deregulation episode on the logarithm of the volume of loans originated by banks, the year of the first deregulation ($\ell = 0$), and in later years ($\ell > 0$). $\text{DID}_{+,0} = 0.079$ (s.e.=0.049): in the year of the first deregulation, loan volume increase by 7.9% more in counties that deregulate than in counties that do not, a marginally insignificant difference at the 10% level (t-stat=1.614). This effect builds up over time, and becomes significant at the 10% level after one year ($\text{DID}_{+,1} = 0.161$, s.e.=0.092) and at the 5% level after three years ($\text{DID}_{+,3} = 0.469$, s.e.=0.215). To the left of zero, placebo estimates are shown. The $\text{DID}_{+,\ell}$ and $\text{DID}^{\text{pl}}_{+,\ell}$ estimators control for state-specific linear trends: without them, placebos are large and significant. With state-specific linear trends, the placebos are small (all are below 0.045 in absolute value) and insignificant: an F-test cannot reject the null that all placebos are insignificant (p-value=0.408). This lends credibility to the parallel trends assumption, at least over a few years. Even accounting for the placebos' confidence intervals, differential pre-trends cannot account for a large fraction of the difference in trends between treated and control counties in the first three years after a deregulation (see Roth, forthcoming). On the other hand, we cannot test if the parallel trends assumption is satisfied over eight years, the duration over which parallel trends has to hold for $\text{DID}_{+,7}$ to be unbiased.

Those results differ starkly from those one would obtain using the distributed-lag TWFE regression in Definition 2, with seven lags of the treatment and controlling for state-specific linear trends. Table 1 in the Web Appendix shows that the coefficients on the current and first three lags of the treatment are small and negative, while the coefficients on the fourth to seventh lags are small and positive. All those coefficients are statistically insignificant, suggesting that banking deregulations have no effect on loan volume.

[14] As $D_{g,t}$ is not binary, $\delta_{+,\ell}$ and δ_+ have to be slightly redefined, see our extension to non-binary treatments in the Web Appendix for more details.

[15] Numbers are for the volume-of-mortgages-originated outcome; numbers would slightly differ for the housing-price outcome: counties with missing values are not exactly the same for the two outcomes.

However, heterogeneous treatment effects may bias this distributed-lag TWFE regression. We follow Proposition 2 and compute the weights attached to this regression. Under the assumptions of this proposition, the coefficient on the current treatment $D_{g,t}$ estimates the sum of eight terms. The first term is a weighted sum of the effects of the current treatment across treated (g,t) cells, where 90 effects receive a positive weight, 120 effects receive a negative weight, and where positive and negative weights respectively sum to 3.40 and -2.40. The second term is a weighted sum of the effects of the first lag of the treatment across (g,t) cells treated one period ago, where 87 effects receive a positive weight, 120 receive a negative weight, and where positive and negative weights respectively sum to 2.39 and -2.39. The remaining six terms are weighted sums of the effects of the second, third, ..., and seventh lag of the treatment, with again large sums of positive and negative weights. Similar results hold for the coefficients on the lagged treatments $D_{g,t-\ell}$. Then, the dynamic TWFE coefficients could be heavily biased, if the effects of the current and lagged treatments vary across (g,t) cells.

We conduct a test, whose results suggest that effects are indeed heterogeneous in this application. This may explain why the distributed-lag TWFE regression leads to such different results from those we obtain with our estimators. Let us momentarily assume that the distributed-lag TWFE regression is correctly specified:

$$Y_{i,g,t}(\boldsymbol{D}_g) = Y_{i,g,t}(\mathbf{0}) + \sum_{\ell=0}^{7} \gamma_\ell D_{g,t-k}, \tag{15}$$

meaning that the effects of the current and past treatments are constant, both across groups and over time. Then, if Assumptions 1-3 hold, it follows from standard results on OLS regressions that the distributed-lag TWFE regression estimators are unbiased for $(\gamma_\ell)_{\ell\in\{0,\ldots,7\}}$ conditional on $\boldsymbol{D}$. Accordingly, one can use those estimators to estimate $(\delta_{+,\ell})_{\ell\in\{0,\ldots,7\}}$. Indeed, if Equation (15) holds, it directly follows from Equation (8) that $\delta_{+,\ell}$ only depends on $(\gamma_\ell)_{\ell\in\{0,\ldots,7\}}$ and $\boldsymbol{D}$. The green line on Figure 1 shows the resulting estimates of $(\delta_{+,\ell})_{\ell\in\{0,\ldots,7\}}$. They are negative, statistically insignificant, and very different from the DID estimators we propose in this paper, shown on the blue line. The TWFE- and DID-based estimates of $\delta_{+,5}$, $\delta_{+,6}$, and $\delta_{+,7}$ are significantly different at the 5% level, with t-stats respectively equal to 1.981, 2.086, and 2.136. The TWFE- and DID-based estimates of $\delta_{+,4}$ and $\delta_{+,3}$ are significantly different at the 10% level. Overall, these results imply that under Assumptions 1-3, we can reject the constant effect assumption in Equation (15). The blue and green lines on Figure 1 also show that our method and the distributed-lag TWFE regression can lead to qualitatively different conclusions.

The blue line on Figure 1 validates the research design, by showing that differential pre-trends are much smaller than differential trends after a treatment switch. It also shows that being exposed to a weakly higher number of deregulations for $\ell+1$ periods has a positive effect on loan volume, and that this effect is increasing with ℓ. However, the $\delta_{+,\ell}$ parameters cannot be interpreted as effects per regulation dropped, because they are not normalized by the number of

deregulations taking place at the time of the first deregulation episode, and because they do not take into account further deregulations or reregulations that occur after the first deregulation. More than 20% of the 42 deregulating states experience at least two changes in their regulations: one state reinstates a regulation after having dropped it, while eight states drop further regulations after their first deregulation. Rather than estimating separately the effects of those subsequent deregulations, which would require imposing restrictions on the dynamic effects of initial deregulations (see de Chaisemartin and D'Haultfoeuille, 2020*b*), our approach estimates the combined effects of initial and subsequent deregulations on the full outcome path.

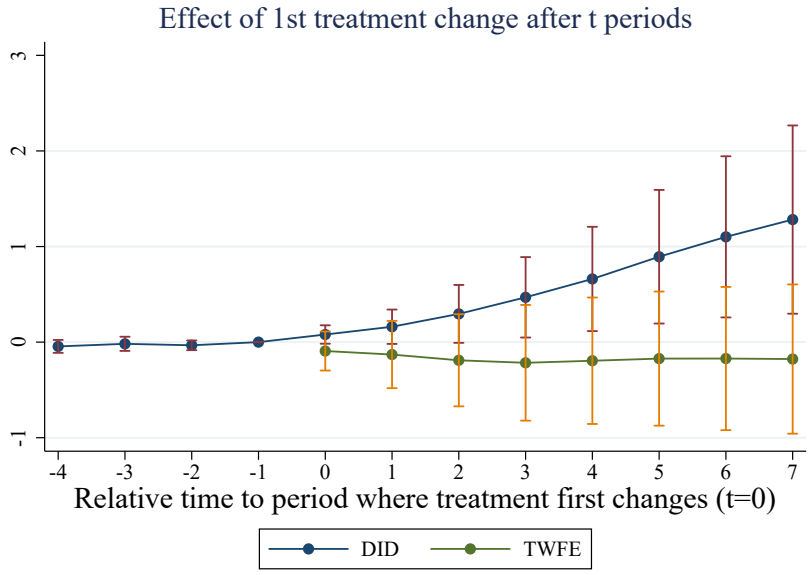

Figure 1: Effect of banking deregulations on loan volume.

Notes: To the right of zero, the blue line on the figure shows the $\text{DID}_{+,\ell}$ estimates of the effect of a first banking-deregulation episode on the logarithm of the volume of loans originated by banks, the year of that first deregulation episode, and in later years. To the left of zero, the blue line shows the $\text{DID}^{\text{pl}}_{+,\ell}$ placebo estimates. At $x=-1$, the placebo is normalized to 0. $\text{DID}^{\text{pl}}_{+,0}$ is shown at $x=-2$, etc. The estimates on the blue line are computed by the Stata `did_multiplegt` command, using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015). To the right of zero, the green line shows estimates of $(\delta_{+,\ell})_{\ell\in\{0,\ldots,7\}}$ based on a distributed-lag TWFE regression of the logarithm of the volume of loans on county and year fixed effects, the current treatment, and seven lags of the treatment. Standard errors are estimated using 100 bootstrap replications clustered at the state level. 95% confidence intervals relying on a normal approximation are shown in red for the $\text{DID}_{+,\ell}$ and $\text{DID}^{\text{pl}}_{+,\ell}$ estimates, and in orange for the TWFE estimators. Both sets of estimators have state-specific linear trends, and are weighted by the inverse of the number of counties per state as in Favara and Imbs (2015).

To help interpret the reduced-form effects on Figure 1, we show the first-stage effects attached

to them. To the right of zero, Figure 2 below shows the $\text{DID}^D_{+,\ell}$ estimates, which measure the number of regulations dropped when a state first deregulates ($\ell = 0$) and in following years ($\ell > 0$). $\text{DID}^D_{+,0} = 2.095$: 2.095 regulations are dropped on average when a state first deregulates. $\text{DID}^D_{+,1} = 2.095$ as well, and the $\text{DID}^D_{+,\ell}$ slightly increase after that: $\text{DID}^D_{+,7} = 2.433$, meaning that seven years after the first deregulation episode, 2.433 regulations have been dropped. Figure 2 for instance implies that $\text{DID}_{+,1}$ estimates an average effect produced by increasing both the lagged and current number of deregulations by 2.095 units on average. One can use Figure 2 similarly to interpret the other $\text{DID}_{+,\ell}$ estimates.

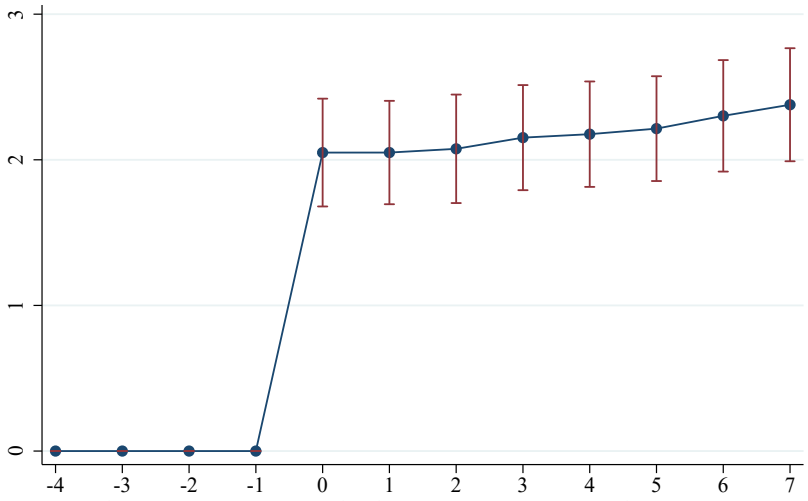

Figure 2: Banking regulations dropped, before & after first deregulation.

Notes: The figure shows the average number of banking regulations dropped, before and after the first deregulation episode. The number of banking deregulations was constructed by Rice and Strahan (2010), and ranges from 0 (fully regulated) to 4 (fully deregulated). The estimators are computed by the Stata `did_multiplegt` command, using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015). Standard errors are estimated using 100 bootstrap replications clustered at the state level. 95% confidence intervals relying on a normal approximation are shown in red. The estimation is weighted by the inverse of the number of counties per state, as in Favara and Imbs (2015).

Finally, to estimate our main parameter of interest $\delta_{+,0:7}$, we average the $\text{DID}_{+,\ell}$ estimates in Figure 1, and we divide that average by the average of the $\text{DID}^D_{+,\ell}$ estimates in Figure 2. This leads to $\widehat{\delta}_{+,0:7} = 0.271$ (s.e.=0.115). Across all deregulations, one deregulation increases loan volume by 27.1% on average, when one sums the deregulation's instantaneous and dynamic effects.

Favara and Imbs (2015) also estimate the effect of banking deregulations on housing prices. Figure 1 in our Web Appendix shows the $\text{DID}_{+,\ell}$ estimates for that outcome. We again find small short-run effects and large long-run effects. That figure also shows that for that outcome,

our estimates of the $(\delta_{+,\ell})_{\ell\in\{0,...,7\}}$ parameters are close to, and insignificantly different from, those one would obtain from a distributed-lag TWFE regression. However, estimators based on the distributed-lag TWFE regression have much larger standard errors than our estimators, and are all insignificant.

Our conclusions regarding the dynamics of the effect of banking regulations starkly differ from those in Favara and Imbs (2015). Based on their Figure 1 Panel B, the authors conclude that the "effects of deregulation on credit supply are temporary: they peak on impact, (...) and peter down until they become insignificant four years after the shock." Similarly, they find significant short-run effects of banking deregulations on housing prices, which vanish after 5 years. We find the opposite.

Favara and Imbs (2015) base their conclusions on a local-projection TWFE regression. To estimate the treatment effect on, say, the growth rate of loan volume, they regress, for every $\ell \in \{0, 1, ..., 8\}$, $\ln(L_{g,t+\ell}) - \ln(L_{g,t+\ell-1})$, the growth rate of loans in county g between years $t+\ell-1$ to $t+\ell$, on county and year FE, the number of deregulations in county g and year $t-1$, and some controls. Proposition 3 does not apply directly to their specification, for two reasons. First, the assumption that $D_{g,t} = I_g \mathbb{1}\{t \geq F_g\}$, under which we derive Proposition 3, is not met in this application: nine states experience at least two changes in their regulations. Second, the authors actually define their treatment variable as $D_{g,t-1}$, the lagged number of regulations, rather than $D_{g,t}$. However, the two key problems uncovered in Proposition 3 are also present in their specification, so it is unlikely that their coefficients $\widehat{\beta}_{lp,\ell}$ unbiasedly estimate the effect of deregulations that took place $\ell+2$ periods ago, accounting for the lagged treatment in their specification. First, many counties with $D_{g,t-1} = 1$ have started deregulating before $t-1$, so $\widehat{\beta}_{lp,\ell}$ estimates effects of deregulations that took place strictly more than $\ell+2$ periods ago for those counties. For instance, for $\ell = 0$, (g,t)s that have $D_{g,t-1} = 1$ have started receiving the treatment 4.27 years before t on average. Second, many counties with $D_{g,t-1} = 0$ have deregulated between periods t to $t+\ell$, so those regressions are misspecified, and may be biased even under fully homogeneous effects. Because the treatment is lagged, the regression is potentially misspecified even for $\ell = 0$. In practice, the amount of misspecification is substantial: for $\ell = 0$, 19.4% of (g,t)s that have $D_{g,t-1} = 0$ have $D_{g,t} = 1$; for $\ell = 8$, 86.9% of (g,t)s that have $D_{g,t-1} = 0$ have $D_{g,t+8} = 1$. Overall, the discrepancy between our results and those of Favara and Imbs (2015) could be due to the fact that those local projection regressions are misspecified and conflate effects of deregulations that took place at different lags.

An alternative explanation is that our specifications differ on two important dimensions. First, their outcome variables are in first-difference, to study the effect of deregulations on the growth rate of loan volume and housing prices. Our specifications on the other hand are in levels. Second, their treatment variable is lagged while ours is not. We recompute our estimators mimicking their specification, and still find much larger long- than short-run effects. Figure 2

in our Web Appendix shows those results for housing prices, results are similar for loan volume. Conversely, we also re-estimate their local projection regressions in levels and with the current treatment: doing so, we actually find some significantly negative effects of deregulations.

Overall, our results show that banking deregulations have small short-term effects on loan volume and housing prices, and large long-term effects. Our results also suggest that heterogeneous effects may severely bias distributed-lag TWFE regressions for the loan-volume outcome. Finally, compared to our estimates, the TWFE local-projection regressions used by Favara and Imbs (2015) lead to opposite conclusions regarding the dynamic of the effect of banking deregulations.

7 Conclusion

We propose DID estimators of instantaneous and dynamic treatment effects, that are robust to heterogeneous treatment effects, unlike commonly used TWFE regressions. Relative to other heterogeneity-robust DID estimators, our estimators are the first that can be used with general designs where the treatment may neither be binary nor staggered. Our main idea is to propose a generalization of the event-study approach to such designs, by defining the event as the period where a group's treatment changes for the first time. Our approach leads to an event-study graph, with the distance to the first treatment change on the x-axis, and estimators of "intention-to-treat" effects of having received a weakly higher amount of treatment for $\ell + 1$ periods on the y-axis. We show that those intention-to-treat effects can be aggregated into a parameter with a clear economic interpretation, that a planner may use to compare groups' actual outcomes to the outcomes they would have obtained if their treatment had remained the same as in period one throughout the panel. In other words, this parameter can be used to determine if the treatment/policy changes that took place over the duration of the panel led to a better situation than the one that would have prevailed if no policy change had been undertaken, an arguably natural policy question. Importantly, that parameter can also be interpreted as an average total effect per unit of treatment, where "total effect" refers to the sum of the instantaneous and dynamic effects of a treatment.

References

Abadie, A. (2005), 'Semiparametric difference-in-differences estimators', *Review of Economic Studies* **72**(1), 1–19.

Abbring, J. H. and Van den Berg, G. J. (2003), 'The nonparametric identification of treatment effects in duration models', *Econometrica* **71**(5), 1491–1517.

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly harmless econometrics*, Princeton university press.

Ashenfelter, O. (1978), 'Estimating the effect of training programs on earnings', *The Review of Economics and Statistics* **60**, 47–57.

Athey, S. and Imbens, G. W. (2022), 'Design-based analysis in difference-in-differences settings with staggered adoption', *Journal of Econometrics* **226**(1), 62–79.

Beel, J. and Gipp, B. (2009), Google scholar's ranking algorithm: an introductory overview, *in* 'Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)', Vol. 1, Rio de Janeiro (Brazil), pp. 230–241.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *The Quarterly Journal of Economics* **119**(1), 249–275.

Bojinov, I., Rambachan, A. and Shephard, N. (2021), 'Panel experiments and dynamic causal effects: A finite population perspective', *Quantitative Economics* **12**(4), 1171–1196.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), Revisiting event study designs: Robust and efficient estimation. arXiv preprint arXiv:2108.12419.

Botosaru, I. and Gutierrez, F. H. (2018), 'Difference-in-differences when the treatment status is observed in only one period', *Journal of Applied Econometrics* **33**(1), 73–90.

Callaway, B., Goodman-Bacon, A. and Sant'Anna, P. H. (2021), Difference-in-differences with a continuous treatment. arXiv preprint arXiv:2107.02637.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Conley, T. G. and Taber, C. R. (2011), 'Inference with "difference in differences" with a small number of policy changes', *The Review of Economics and Statistics* **93**(1), 113–125.

de Chaisemartin, C. (2021), The minimax estimator of the average treatment effect, among linear combinations of estimators of bounded conditional average treatment effects. arXiv preprint arXiv:2105.08766.

de Chaisemartin, C. and D'Haultfœuille, X. (2018), 'Fuzzy differences-in-differences', *The Review of Economic Studies* **85**(2), 999–1028.

de Chaisemartin, C. and D'Haultfœuille, X. (2020*a*), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020*b*), Two-way fixed effects regressions with several treatments. arXiv preprint arXiv:2012.10077.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022), Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. National Bureau of Economic Research Working Paper 29691.

de Chaisemartin, C., D'Haultfœuille, X., Pasquier, F. and Vazquez-Bare, G. (2022), Difference-in-differences estimators for treatments continuously distributed at every period. arXiv preprint arXiv:2201.06898.

Deryugina, T. (2017), 'The fiscal cost of hurricanes: Disaster aid versus social insurance', *American Economic Journal: Economic Policy* **9**(3), 168–98.

Favara, G. and Imbs, J. (2015), 'Credit supply and the price of housing', *American Economic Review* **105**(3), 958–92.

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

Gut, A. (1992), 'The weak law of large numbers for arrays', *Statistics & probability letters* **14**(1), 49–52.

Gut, A. (2013), *Probability: a graduate course*, Vol. 75, Springer Science & Business Media.

Han, S. (2019), Optimal dynamic treatment regimes and partial welfare ordering. arXiv preprint arXiv:1912.10014.

Han, S. (2021), 'Identification in nonparametric models for dynamic treatment effects', *Journal of Econometrics* **225**(2), 132–147.

Jordà, Ò. (2005), 'Estimation and inference of impulse responses by local projections', *American economic review* **95**(1), 161–182.

Malani, A. and Reif, J. (2015), 'Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform', *Journal of Public Economics* **124**, 1–17.

Manski, C. F. and Pepper, J. V. (2018), 'How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions', *Review of Economics and Statistics* **100**(2), 232–244.

Murphy, S. A. (2003), 'Optimal dynamic treatment regimes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 331–355.

Murphy, S. A., van der Laan, M. J., Robins, J. M. and Group, C. P. P. R. (2001), 'Marginal mean models for dynamic regimes', *Journal of the American Statistical Association* **96**(456), 1410–1423.

Rambachan, A. and Roth, J. (2019), An honest approach to parallel trends. Working paper.

Rice, T. and Strahan, P. E. (2010), 'Does credit competition affect small-firm finance?', *The Journal of Finance* **65**(3), 861–889.

Robins, J. (1986), 'A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect', *Mathematical modelling* **7**(9-12), 1393–1512.

Roth, J. (forthcoming), 'Pre-test with caution: Event-study estimates after testing for parallel trends', *American Economic Review: Insights* .

Sianesi, B. (2004), 'An evaluation of the swedish system of active labor market programs in the 1990s', *Review of Economics and statistics* **86**(1), 133–155.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

A Appendix: proofs

A.1 Proof of Proposition 1

First, by Frisch-Waugh's theorem,

$$\widehat{\beta}_{fe,\ell} = \frac{\sum_{g=1}^{G}\sum_{t=1}^{T} N_g \varepsilon_{g,t} Y_{g,t}}{\sum_{g=1}^{G}\sum_{t=1}^{T} N_g \varepsilon_{g,t} I_g \mathbb{1}\{t=\ell\}}, \tag{16}$$

where $\varepsilon_{g,t}$ is the residual of the regression of $I_g\mathbb{1}\{t=\ell\}$ on group and time fixed effects and the $(I_g\mathbb{1}\{t=k\})_{k\notin\{F-1,\ell\}}$, with cell (g,t) weighted by N_g. Now, reasoning as in Lemma 1 in de Chaisemartin and D'Haultfœuille (2020*a*), we obtain that for all g, g' and $t \geq F > t'$,

$$E[Y_{g,t} - Y_{g,t'} - (Y_{g',t} - Y_{g',t'})\,|\boldsymbol{D}] = I_g E\left[\Delta_{g,t}(t-F+1, I_g)|\boldsymbol{D}\right] - I_{g'} E\left[\Delta_{g',t}(t-F+1, I_{g'})|\boldsymbol{D}\right].$$

Then, by the same reasoning as in the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020*a*), we obtain

$$E\left[\sum_{g=1}^{G}\sum_{t=1}^{T} N_g \varepsilon_{g,t} Y_{g,t}\middle|\boldsymbol{D}\right] = E\left[\sum_{g:I_g\neq 0}\sum_{t=F}^{T} N_g \varepsilon_{g,t} I_g \Delta_{g,t}(t-F+1, I_g)\middle|\boldsymbol{D}\right]. \tag{17}$$

Now, we compute $\varepsilon_{g,t}$. Let $(\alpha_g)_{g=1\ldots G}, (\gamma_t)_{t=1\ldots T}$ and $(\delta_t)_{t\notin\{F-1,\ell\}}$ denote the coefficients of the regression of $I_g\mathbb{1}\{t=\ell\}$ on group and time fixed effects and the variables $(I_g\mathbb{1}\{t=\ell\})_{t\notin\{F-1,\ell\}}$, where cell (g,t) is weighted by N_g. Then

$$\varepsilon_{g,t} = I_g\mathbb{1}\{t=\ell\} - \alpha_g - \gamma_t - I_g\mathbb{1}\{t\notin\{F-1,\ell\}\}\,\delta_t.$$

Assume without loss of generality that $\gamma_{F-1}=0$. Then, using the first-order conditions (FOC) with respect to time fixed effects, we obtain, for all $t\in\{1,...,T\}$,

$$\gamma_t = \bar{I}\left(\mathbb{1}\{t=\ell\} - \delta_t\mathbb{1}\{t\notin\{F-1,\ell\}\}\right).$$

Second, the FOC with respect to group fixed effects yield, for all $g\in\{1,...,G\}$,

$$\alpha_g = \frac{I_g - \bar{I}}{T}\left[1 - \sum_{t\notin\{F-1,\ell\}}\delta_t\right].$$

Third, the the FOC with respect to the variables $(I_g\mathbb{1}\{t=\ell\})_{t\notin\{F-1,\ell\}}$ imply that for all $t\notin\{F-1,\ell\}$,

$$\delta_t V_e(I) + \frac{V_e(I)}{T}\left(1 - \sum_{t\notin\{F-1,\ell\}}\delta_t\right) = 0,$$

with $V_e(I) = \sum_{g=1}^{G} N_g(I_g-\bar{I})^2/\sum_{g=1}^{G}N_g$. This last condition implies that $\delta_t = -1/2$. Hence, $\alpha_g = (I_g-\bar{I})/2$ and

$$\gamma_t = \bar{I}\left(\mathbb{1}\{t=\ell\} + \frac{\mathbb{1}\{t\notin\{F-1,\ell\}\}}{2}\right).$$

As a result, for all $t \geq F$, $t \neq \ell$, we obtain $\varepsilon_{g,t} = 0$. Point 2 of the proposition follows by (16), (17) and the law of iterated expectation. Moreover, for any $\ell \geq F$, by what precedes, $\varepsilon_{g,\ell} = (I_g - \bar{I})/2$. By, again, (16), (17) and the law of iterated expectation, we get

$$E\left[\widehat{\beta}_{fe,\ell}\right] = E\left[\frac{\sum_{g:I_g\neq 0} N_g I_g (I_g - \bar{I}) \Delta_{g,\ell}(\ell - F + 1, I_g)}{\sum_{g:I_g\neq 0} N_g I_g (I_g - \bar{I})}\right].$$

Point 1 follows.

A.2 Proof of Proposition 3

First, by Frisch-Waugh's theorem,

$$\widehat{\beta}_{lp,\ell} = \frac{\sum_{g=1}^{G}\sum_{t=1}^{T-\ell} N_g \varepsilon_{g,t} Y_{g,t+\ell}}{\sum_{g=1}^{G}\sum_{t=1}^{T-\ell} N_g \varepsilon_{g,t} D_{g,t}}, \tag{18}$$

where $\varepsilon_{g,t}$ is the residual of the regression of $D_{g,t}$ on period and group fixed effects in the subsample such that $1 \leq t \leq T-\ell$. Because $N_{g,t} = N_g$ for all (g,t), $\varepsilon_{g,t} = D_{g,t} - D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.}$. Reasoning as in Lemma 1 in de Chaisemartin and D'Haultfœuille (2020a), we have, for any g, g',

$$\begin{aligned} E[Y_{g,t+\ell} - Y_{g,1} - (Y_{g',t+\ell} - Y_{g',1})|\boldsymbol{D}] =& I_g \mathbb{1}\left\{t \geq F_g\right\} E[\Delta_{g,t+\ell}(t+\ell-F_g+1, I_g)|\boldsymbol{D}] \\ & - I_{g'} \mathbb{1}\left\{t \geq F_{g'}\right\} E[\Delta_{g',t+\ell}(t+\ell-F_{g'}+1, I_{g'})|\boldsymbol{D}]. \end{aligned}$$

A similar reasoning as in the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020a) yields

$$\begin{aligned} & E\left[\sum_{g=1}^{G}\sum_{t=1}^{T-\ell} N_g \varepsilon_{g,t} Y_{g,t+\ell} \Big| \boldsymbol{D}\right] \\ =& \sum_{(g,t):I_g\neq 0, t+\ell\geq F_g} N_g \varepsilon_{g,t} I_g E[\Delta_{g,t+\ell}(t+\ell-F_g+1, I_g)|\boldsymbol{D}] \\ =& \sum_{(g,t):I_g\neq 0, t+\ell\geq F_g} N_g I_g \left(D_{g,t} - D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.}\right) E[\Delta_{g,t+\ell}(t+\ell-F_g+1, I_g)|\boldsymbol{D}]. \end{aligned} \tag{19}$$

The denominator of $\widehat{\beta}_{lp,\ell}$ satisfies

$$\sum_{g=1}^{G}\sum_{t=1}^{T-\ell} N_g \varepsilon_{g,t} D_{g,t} = \sum_{(g,t):I_g\neq 0, t\geq F_g} N_g I_g (I_g \mathbb{1}\left\{t \geq F_g\right\} - D^{\ell}_{g,.} - D_{.,t} + D^{\ell}_{.,.}).$$

Thus, we obtain

$$E[\widehat{\beta}_{lp,\ell}] = E\left[\sum_{(g,t):I_g\neq 0, t+\ell\geq F_g} w^{lp,\ell}_{g,t} \Delta_{g,t+\ell}(t+\ell-F_g+1, I_g)\right],$$

where the weights $w_{g,t}^{lp,\ell}$ satisfy

$$w_{g,t}^{lp,\ell} = \frac{N_g I_g (I_g \mathbb{1}\{t \geq F_g\} - D_{g,.}^{\ell} - D_{.,t} + D_{.,.}^{\ell})}{\sum_{(g',t'):I_{g'}\neq 0, t' \geq F_{g'}} N_{g'} I_{g'} (I_{g'} - D_{g',.}^{\ell} - D_{.,t'} + D_{.,.}^{\ell})}.$$

Hence, for all $\ell \geq 0$, using the convention that a sum over an empty set is empty,

$$\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g} w_{g,t}^{lp,\ell} = 1 + \frac{\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g > t} N_g I_g (-D_{g,.}^{\ell} - D_{.,t} + D_{.,.}^{\ell})}{\sum_{(g',t'):I_{g'}\neq 0, t' \geq F_{g'}} N_{g'} I_{g'} (I_{g'} - D_{g',.}^{\ell} - D_{.,t'} + D_{.,.}^{\ell})}.$$

This implies that for all $\ell \geq 1$, $\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g} w_{g,t}^{lp,\ell} \neq 1$ unless $\sum_{(g,t):I_g\neq 0, t+\ell \geq F_g > t} N_g(-D_{g,.}^{\ell} - D_{.,t} + D_{.,.}^{\ell}) = 0$. Moreover, $\sum_{(g,t):I_g\neq 0, t \geq F_g} w_{g,t}^{lp,0} = 1$.

Finally, assume that $I_g = 1$ for all g. Let $g^* = \arg\min_g F_g$ and let $t^* = F_{g*} - 1$. Notice that $\ell \leq T - F_{g^*}$. If $\ell > T - F_{g^*}$, $D_{g,t} = 0$ for all the cells in the sample on which the regression is estimated, so $\widehat{\beta}_{lp,\ell}$ is not well defined.[16] Notice also that the denominator of the weights is positive. Then, the sign of $w_{g^*,t^*}^{lp,\ell}$ is the same as that of $-D_{g^*,.}^{\ell} - D_{.,t^*} + D_{.,.}^{\ell}$. Moreover,

$$\begin{aligned} D_{g^*,.}^{\ell} &= \frac{\sum_{t=1}^{T-\ell} N_{g^*} D_{g^*,t}}{\sum_{t=1}^{T-\ell} N_{g^*}} = \frac{T - \ell - F_{g^*} + 1}{T - \ell}, \\ D_{.,.}^{\ell} &= \frac{\sum_{g=1}^{G} \sum_{t=1}^{T-\ell} N_g D_{g,t}}{\sum_{g=1}^{G} \sum_{t=1}^{T-\ell} N_g} = \frac{\sum_{g=1}^{G} N_g \max(0, T - \ell - F_g + 1)}{(T-\ell)\sum_{g=1}^{G} N_g}. \end{aligned}$$

By construction, $T-\ell-F_{g^*}+1 \geq T-\ell-F_g+1$. Moreover, $T-\ell-F_{g^*}+1 > 0$ since $\ell \leq T-F_{g^*}$. Then, because F_g is not constant, $T - \ell - F_{g^*} + 1 > T - \ell - F_g + 1$ for at least one g. Thus, $\sum_{g=1}^{G} N_g \max(0, T - \ell - F_g + 1) < \sum_{g=1}^{G} N_g (T - \ell - F_{g^*} + 1)$, implying that $D_{.,.}^{\ell} < D_{g^*,.}^{\ell}$. As a result,

$$\min_{g,t} w_{g,t}^{lp,\ell} \leq w_{g^*,t^*}^{lp,\ell} < 0.$$

A.3 Proof of Theorem 2

$$\begin{aligned} & E\left[\mathrm{DID}_{g,\ell}^{\mathrm{pl}} \middle| \boldsymbol{D}\right] \\ =& E\left[Y_{g,F_g-\ell-2}(\mathbf{0}) - Y_{g,F_g-1}(\mathbf{0}) \middle| \boldsymbol{D}\right] - E\left[\sum_{g':D_{g',1}=0, F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N_{F_g+\ell}^{u}} (Y_{g',F_g-\ell-2}(\mathbf{0}) - Y_{g',F_g-1}(\mathbf{0})) \middle| \boldsymbol{D}\right] \\ =& 0 \end{aligned} \tag{20}$$

The first equality follows from the definition of F_g and Assumption 2. The second equality follows from Assumption 3 and the definition of $N_{F_g+\ell}^{u}$. Then, the result follows from the definition of $\mathrm{DID}_{+,\ell}^{\mathrm{pl}}$, the fact that $(F_g)_{1\leq g\leq G}$, T_u, and $(N_\ell^{1,\mathrm{pl}})_{0\leq \ell \leq L_u^{\mathrm{pl}}}$are functions of $\boldsymbol{D}$, the linearity of the conditional expectation operator, and Equation (20).

[16] As stated in Footnote 6, the proposition implicitly assumes that $\widehat{\beta}_{lp,\ell}$ is well defined.

A.4 Proof of Theorem 3

Recall that $\mathcal{S}$ is the support of $F_g|D_{g,1}=0$, which does not depend on g by Assumption 6. For every $t \in \mathcal{S}$, by the strong law of large numbers for independent but not identically distributed variables (e.g., Theorem 5.4 in Gut, 2013), we have, almost surely,

$$\frac{1}{G}\sum_{g=1}^{G}\left(\mathbb{1}\left\{D_{g,1}=0, F_g=t\right\} - \Pr(D_{g,1}=0, F_g=t)\right) \xrightarrow{a.s.} 0.$$

Thus, by Assumption 6 again,

$$\sum_{g=1}^{G}\mathbb{1}\left\{D_{g,1}=0, F_g=t\right\} \geq G(\underline{p}+\varepsilon_{t,G}), \tag{21}$$

where $\varepsilon_{t,G} \xrightarrow{a.s.} 0$. Hence, almost surely, $\#\{g : D_{g,1}=0, F_g=t\} \to \infty$. As a result, there is an event $\mathcal{E}$ of probability one for which there exists $\underline{G}$ such that for all $G \geq \underline{G}$, $\#\{g : D_{g,1}=0, F_g=t\} > 0$ for all $t \in \mathcal{S}$. In what follows, we reason without loss of generality conditional on $\mathcal{E}$, assuming that $G \geq \underline{G}$. Note that under these conditions, T_u is not random: $T_u = \max \mathcal{S} - 1$. Similarly, $\underline{T} := \min_{g=1...G} F_g$ and L_u are also not random: $\underline{T} = \min \mathcal{S}$, and $L_u = \max \mathcal{S} - 1 - \min \mathcal{S}$. Finally, remark that by Assumption 6, $L_u \geq 0$.

Asymptotic normality of $DID_{+,\ell}$

We consider an arbitrary $\ell \in \{0, ..., L_u\}$. We have

$$\text{DID}_{+,\ell} = \frac{1}{G}\sum_{g:D_{g,1}=0} U_{G,g,\ell}.$$

We show below that

$$\lim_{G\to\infty} \frac{\sum_{g:D_{g,1}=0} E\left[\left|U_{G,g,\ell} - E[U_{G,g,\ell}|\boldsymbol{D}]\right|^{2+\delta}\middle|\, |\boldsymbol{D}\right]}{\left(\sum_{g:D_{g,1}=0} V(U_{G,g,\ell}\,|\boldsymbol{D})\right)^{1+\delta/2}} = 0 \text{ a.s.} \tag{22}$$

Then, by Lyapunov's central limit theorem for triangular arrays and $\delta_{+,\ell} = E[\text{DID}_{+,\ell}|\boldsymbol{D}]$ (in view of Theorem 1), Equation (13) will follow.

First, remark that

$$U_{G,g,\ell} = \frac{1}{N^1_\ell/G}\sum_{t=1}^{T_u}\lambda_{g,t,\ell}Y_{g,t},$$

with

$$\begin{aligned}\lambda_{g,t,\ell} :=& \beta^t N_{g,t}\left(\mathbb{1}\left\{F_g = t-\ell\right\} - \frac{N^1_{t,\ell}}{N^u_t}\mathbb{1}\left\{F_g > t\right\}\right)\mathbb{1}\left\{t \geq \ell+2\right\} - \beta^{t+\ell+1}N_{g,t+\ell+1} \\ &\times\left(\mathbb{1}\left\{F_g = t+1\right\} - \frac{N^1_{t+\ell+1,\ell}}{N^u_{t+\ell+1}}\mathbb{1}\left\{F_g > t+\ell+1\right\}\right)\mathbb{1}\left\{t \leq T_u - \ell - 1\right\}.\end{aligned} \tag{23}$$

If $F_g \leq T_u - \ell$, then

$$\lambda_{g,F_g+\ell,\ell} = \beta^{F_g+\ell} N_{g,F_g+\ell}.$$

Then, letting $\boldsymbol{\lambda}_{g,\ell} := (\lambda_{g,1,\ell}, ..., \lambda_{g,T_u,\ell})'$, we have, by Assumption 7,

$$\begin{aligned}
\left(N_\ell^1/G\right)^2 \sum_{g:D_{g,1}=0} V(U_{G,g,\ell}|\boldsymbol{D}) &= \sum_{g:D_{g,1}=0} \boldsymbol{\lambda}_{g,\ell}' \Sigma_g \boldsymbol{\lambda}_{g,\ell} \\
&\geq \underline{\rho} \sum_{g:D_{g,1}=0} \|\boldsymbol{\lambda}_{g,\ell}\|_2^2 \\
&\geq \underline{\rho} \sum_{g:D_{g,1}=0, F_g \leq T_u - \ell} \left(\beta^{F_g+\ell} N_{g,F_g+\ell}\right)^2 \qquad (24) \\
&\geq \underline{\rho} \beta^{2T_u} \#\{g : D_{g,1} = 0, F_g \leq T_u - \ell\}. \qquad (25)
\end{aligned}$$

Next, by the triangle inequality, convexity of $x \mapsto x^{2+\delta}$ and Jensen's inequality,

$$\left|\sum_{t=1}^{T_u} \lambda_{g,t,\ell}(Y_{g,t} - E[Y_{g,t}|\boldsymbol{D}])\right|^{2+\delta} \leq T^{1+\delta} 2^{1+\delta} \sum_{t=1}^{T_u} |\lambda_{g,t,\ell}|^{2+\delta} \left(|Y_{g,t}|^{2+\delta} + E[|Y_{g,t}|^{2+\delta}|\boldsymbol{D}]\right).$$

Therefore, by Assumption 7, there exists $C_0 > 0$ such that

$$\left(N_\ell^1/G\right)^{2+\delta} E\left[|U_{G,g,\ell} - E[U_{G,g,\ell}|\boldsymbol{D}]|^{2+\delta} \,|\boldsymbol{D}\right] \leq C_0 \left[\max_{t=1...T_u} |\lambda_{g,t,\ell}|\right]^{2+\delta}. \qquad (26)$$

Moreover, by (23) and using again Assumption 7,

$$\max_{t=1...T_u} |\lambda_{g,t,\ell}| \leq 2 \sup_{g,t} N_{g,t} \times \left[1 + \max_{t=1...T_u} \frac{N_{t,\ell}^1}{N_t^u}\right]. \qquad (27)$$

Also, by definition of $N_{t,\ell}^1$,

$$\begin{aligned}
N_{t,\ell}^1 &\leq \sup_{g,t} N_{g,t} \times \#\{g : D_{g,1} = 0, F_g = t - \ell \geq 2\} \\
&\leq \sup_{g,t} N_{g,t} \times \#\{g : D_{g,1} = 0, F_g \leq T_u - \ell\}. \qquad (28)
\end{aligned}$$

Further, by definition of N_t^u,

$$N_t^u \geq N_{T_u}^u = \#\{g : D_{g,1} = 0, F_g = T_u + 1\}. \qquad (29)$$

Gathering together (25), (26), (28) and (29) and using the convexity of $x \mapsto x^{2+\delta}$, we obtain that there exists $C_1, C_2 > 0$ such that

$$\begin{aligned}
&\frac{\sum_{g:D_{g,1}=0} E\left[|U_{G,g,\ell} - E[U_{G,g,\ell}|\boldsymbol{D}]|^{2+\delta}| \,|\boldsymbol{D}\right]}{\left(\sum_{g:D_{g,1}=0} V(U_{G,g,\ell}\,|\boldsymbol{D})\right)^{1+\delta/2}} \\
\leq &\frac{C_1 G}{\#\{g : D_{g,1} = 0, F_g \leq T_u - \ell\}^{1+\delta/2}} + C_2 G \left(\frac{\#\{g : D_{g,1} = 0, F_g \leq T_u - \ell\}}{\#\{g : D_{g,1} = 0, F_g = T_u + 1\}^2}\right)^{1+\delta/2}. \qquad (30)
\end{aligned}$$

Remark that $T_u + 1 \in \mathcal{S}$ and $\{2, ..., T_u - \ell\} \subset \mathcal{S}$. Hence, by (21), we have, almost surely,

$$\begin{aligned} G \geq \#\{g : D_{g,1} = 0, F_g \leq T_u - \ell\} &\geq G(\underline{p} + \varepsilon_{2,G}), \\ \#\{g : D_{g,1} = 0, F_g = T_u + 1\} &\geq G(\underline{p} + \varepsilon_{T_u+1,G}), \end{aligned}$$

where $\varepsilon_{2,G}$ and $\varepsilon_{T_u+1,G}$ both converge to 0 almost surely. Therefore, both terms on the right-hand side of (30) tend to 0. Equation (22), and thus (13), follow.

Asymptotic normality of $\widehat{\delta}_+$

The reasoning is similar to above so we omit some details here. Because $\widehat{\delta}_+ = \sum_{g:D_{g,1}=0} U_{G,g}/G$ and $\delta_+ = E[\widehat{\delta}_+|\boldsymbol{D}]$ (by Theorem 1), it suffices to prove

$$\lim_{G\to\infty} \frac{\sum_{g:D_{g,1}=0} E\left[|U_{G,g} - E[U_{G,g}|\boldsymbol{D}]|^{2+\delta}| \, |\boldsymbol{D}\right]}{\left(\sum_{g:D_{g,1}=0} V(U_{G,g} \, |\boldsymbol{D})\right)^{1+\delta/2}} = 0 \text{ a.s.} \tag{31}$$

To this end, note that $U_{G,g} = \sum_{t=1}^{T_u} \mu_{g,t} Y_{g,t}/K_G$, with $\mu_{g,t} := \sum_{\ell=0}^{L_u} \lambda_{g,t,\ell}$ and

$$K_G := \frac{1}{G} \sum_{\ell=0}^{L_u} \sum_{t'=\ell+2}^{T_u} \beta^{t'} \sum_{g':F_{g'}=t'-\ell} N_{g',t'} D_{g',t'}.$$

Moreover, for all $t \geq 2$, we have $\mu_{g,t} = \beta^t N_{g,t}(T_u + 1 - F_g)$ if $F_g \in \{\max(2, t + \underline{T} - T_u), ..., t\}$. Then, reasoning as above, we obtain

$$K_G^2 \sum_{g:D_{g,1}=0} V(U_{G,g}|\boldsymbol{D}) \geq \underline{\rho}\beta^{2T_u} \sum_{g:g:D_{g,1}=0, F_g \geq \underline{T}} (T_u + 1 - F_g).$$

Moreover, in view of (27)-(29), there exists $C_3 > 0$ such that

$$\begin{aligned} \max_{t=1...T_u} |\mu_{g,t}| &\leq \sum_{\ell=0}^{L_u} \max_{t=1...T_u} |\lambda_{g,t,\ell}| \\ &\leq C_3 \left[1 + \frac{\sum_{g:D_{g,1}=0, F_g \geq \underline{T}} (T_u + 1 - F_g)}{\#\{g : D_{g,1} = 0, F_g = T_u + 1\}}\right]. \end{aligned}$$

Then, reasoning as above, there exists $C_4, C_5 > 0$ such that

$$\begin{aligned} \frac{\sum_{g:D_{g,1}=0} E\left[|U_{G,g} - E[U_{G,g}|\boldsymbol{D}]|^{2+\delta}| \, |\boldsymbol{D}\right]}{\left(\sum_{g:D_{g,1}=0} V(U_{G,g} \, |\boldsymbol{D})\right)^{1+\delta/2}} \leq & \frac{C_4 G}{\left(\sum_{g:D_{g,1}=0, F_g \geq \underline{T}} (T_u + 1 - F_g)\right)^{1+\delta/2}} \\ & + C_5 G \left(\frac{\sum_{g:D_{g,1}=0, F_g \geq \underline{T}} (T_u + 1 - F_g)}{\#\{g : D_{g,1} = 0, F_g = T_u + 1\}^2}\right)^{1+\delta/2}. \end{aligned}$$

Equation (31) then follows by, again, the strong law of large numbers.

Confidence intervals have asymptotic coverage of at least $1-\alpha$

We consider only $\text{CI}_{1-\alpha}(\delta_{+,\ell})$, the reasoning being the same for $\text{CI}_{1-\alpha}(\delta_{+})$. Let us define

$$\overline{\sigma}^2_{G,\ell} := \frac{1}{G}\sum_{g:D_{g,1}=0} E\left[\left(U_{G,g,\ell}-\delta_{+,\ell}\right)^2|\boldsymbol{D}\right].$$

By the weak law of large numbers of Gut (1992) and since $E[|U_{G,g,\ell}|^{2+\delta}|\boldsymbol{D}]<\infty$, we have, almost surely,

$$\frac{1}{G}\sum_{g=1}^{G}\mathbb{1}\left\{D_{g,1}=0\right\}\left[U^2_{G,g,\ell}-E\left[U^2_{G,g,\ell}|\boldsymbol{D}\right]\right]\xrightarrow{P}0. \tag{32}$$

By (13), $\text{DID}_{+,\ell}-\delta_{+,\ell}\xrightarrow{P}0$. Moreover,

$$\begin{aligned}|\delta_{+,\ell}| &= \left|\frac{1}{G}\sum_{g:D_{g,1}=0}E[U_{G,g,\ell}|\boldsymbol{D}]\right| \\ &\leq \frac{1}{G}\sum_{g=1}^{G}E\left[|U_{G,g,\ell}|^{2+\delta}|\boldsymbol{D}\right]^{1/(2+\delta)} \\ &\leq C_6,\end{aligned}$$

where the last inequality holds for some $C_6>0$ by Assumption 7. As a result,

$$\begin{aligned}\left|\text{DID}^2_{+,\ell}-\delta^2_{+,\ell}\right| \leq\ & \left(\text{DID}_{+,\ell}-\delta_{+,\ell}\right)^2+2|\delta_{+,\ell}|\times|\text{DID}_{+,\ell}-\delta_{+,\ell}| \\ &\xrightarrow{P}0.\end{aligned} \tag{33}$$

Then, by (32) and (33), we have, almost surely,

$$\widehat{\sigma}^2_{\ell}-\overline{\sigma}^2_{G,\ell}\xrightarrow{P}0. \tag{34}$$

Next, by definition of the conditional expectation,

$$E\left[\left(U_{G,g,\ell}-\delta_{+,\ell}\right)^2|\boldsymbol{D}\right]\geq E\left[\left(U_{G,g,\ell}-E(U_{G,g,\ell}|\boldsymbol{D})\right)^2|\boldsymbol{D}\right]=V(U_{G,g,\ell}|\boldsymbol{D}).$$

Thus,

$$\overline{\sigma}^2_{G,\ell}\geq\frac{1}{G}\sum_{g:D_{g,1}=0}V(U_{G,g,\ell}|\boldsymbol{D}). \tag{35}$$

Moreover,

$$\begin{aligned}\frac{1}{G}\sum_{g:D_{g,1}=0}V(U_{G,g,\ell}|\boldsymbol{D}) &\geq \underline{\rho}\frac{\frac{1}{G}\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\left(\beta^{F_g+\ell}N_{g,F_g+\ell}\right)^2}{(N^1_\ell/G)^2} \\ &\geq \underline{\rho}\frac{\frac{1}{G}\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\left(\beta^{F_g+\ell}N_{g,F_g+\ell}\right)^2}{\left(\frac{1}{G}\sum_{g:D_{g,1}=0,F_g\leq T_u-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell}\right)^2} \\ &\geq \underline{\rho},\end{aligned}$$

where the first inequality follows by (24), the second by definition of N_ℓ^1 and the third by convexity of $x \mapsto x^2$. Hence, In view of (34) and (35), $\widehat{\sigma}_\ell^2 > \underline{\rho}/2 > 0$ with probability approaching one. Then, under this event, we get

$$\left|\frac{\overline{\sigma}_{G,\ell}}{\widehat{\sigma}_\ell} - 1\right| = \left|\frac{\widehat{\sigma}_\ell^2 - \overline{\sigma}_{G,\ell}^2}{\widehat{\sigma_\ell}(\overline{\sigma}_{G,\ell} + \widehat{\sigma}_\ell)}\right| \leq \frac{\left|\widehat{\sigma}_\ell^2 - \overline{\sigma}_{G,\ell}^2\right|}{\underline{\rho}}.$$

Therefore, almost surely, $|\overline{\sigma}_{G,\ell}/\widehat{\sigma}_\ell - 1| \xrightarrow{P} 0$. Then,

$$G^{1/2}\frac{\mathrm{DID}_{+,\ell} - \delta_{+,\ell}}{\widehat{\sigma}_\ell} = \frac{\left(\frac{1}{G}\sum_{g:D_{g,1}=0} V(U_{G,g,\ell}|\boldsymbol{D})\right)^{1/2}}{\overline{\sigma}_{G,\ell}} \times \left[\frac{\overline{\sigma}_{G,\ell}}{\widehat{\sigma}_\ell} \times G^{1/2}\frac{\mathrm{DID}_{+,\ell} - \delta_{+,\ell}}{\left(\frac{1}{G}\sum_{g:D_{g,1}=0} V(U_{G,g,\ell}|\boldsymbol{D})\right)^{1/2}}\right].$$

By Slutski's lemma and (13), the term into brackets converges in distribution (and almost surely) to $Z \sim \mathcal{N}(0,1)$. Then, using (35), we obtain

$$\begin{aligned}\Pr(\delta_{+,\ell} \in \mathrm{CI}_{1-\alpha}(\delta_{+,\ell})|\boldsymbol{D}) &\geq \Pr\left(\left|\frac{\overline{\sigma}_{G,\ell}}{\widehat{\sigma}_\ell} \times G^{1/2}\frac{\mathrm{DID}_{+,\ell} - \delta_{+,\ell}}{\left(\frac{1}{G}\sum_{g:D_{g,1}=0} V(U_{G,g,\ell}|\boldsymbol{D})\right)^{1/2}}\right| \leq z_{1-\alpha/2}\middle|\boldsymbol{D}\right) \\ &\xrightarrow{a.s.} \Pr(|Z| \leq z_{1-\alpha/2}) = 1-\alpha.\end{aligned}$$

The result follows.

Difference-in-Differences Estimators of Intertemporal Treatment Effects
Web Appendix

Clément de Chaisemartin* Xavier D'Haultfœuille†

Abstract

In this web appendix, we consider several extensions to our theoretical results. In the first section, we propose parameters of interest and estimators for initially-treated groups. In the second section, we consider discrete treatments. In the third section, we consider continuous and staggered treatments. In the fourth section, we consider other extensions. Finally, the last section collects additional material on the application.

1 Initially-treated groups

In this section, we extend the analysis in Section 4 of the paper to groups that are treated at the first period. We thus consider designs where some initially-treated groups become untreated, at a date where some initially-treated groups have been treated all along, as stated below.

Assumption 8 *(Non-pathological design for initially-treated groups)* $\exists(g,g') \in \{1,...,G\}^2$ *such that* $D_{g,1}=D_{g',1}=1$ *and* $F_g < F_{g'}$.

1.1 Parameters of interest for initially-treated groups, and aggregated parameters

Let $T_a = \max_{g:D_{g,1}=1} F_g - 1$ denote the last period where there is still a group that has been treated since period 1. T_a is the last period where initially-treated groups' treatment effects can be estimated under our assumptions. For any initially-treated group g such that $F_g \leq T_a$, and for any $\ell \in \{0,...,T_a - F_g\}$, let

$$\delta_{g,\ell} = E(Y_{g,F_g+\ell}(\boldsymbol{D}_g) - Y_{g,F_g+\ell}(\mathbf{1})|\boldsymbol{D}) = E(Y_{g,F_g+\ell}(\mathbf{1}_{F_g-1},0,D_{g,F_g+1},...,D_{g,F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{1}_{F_g+\ell})|\boldsymbol{D}).$$

*Economics Department, Sciences Po, clement.dechaisemartin@sciencespo.fr
†CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr

$\delta_{g,\ell}$ is the expected difference between group g's actual outcome at period $F_g+\ell$, ℓ periods after it got untreated for the first time, and the counterfactual "status quo" outcome it would have obtained at that period if it had remained treated from period one to $F_g+\ell$. For every g, $\delta_{g,\ell}$ is an effect of having received a weakly lower amount of treatment for $\ell+1$ periods. Depending on g's treatments after period F_g, $\delta_{g,\ell}$ may correspond to the effect of having been untreated rather than treated from at least one to at most $\ell+1$ periods.

We can now define our main parameter of interest for initially-treated groups. In groups treated at period 1, the actual treatments are beneficial relative to the status quo, up to period T_a, if and only if

$$\sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}\delta_{g,\ell} - \sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}(D_{g,F_g+\ell}-1)>0. \quad (36)$$

Let

$$c_- = \frac{\sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}(1-D_{g,F_g+\ell})}{\sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(1-D_{g,F_g+\ell})}$$

denote the average discounted cost of the treatment, across all the discounted treatments not received by groups initially treated. Assumption 8 ensures that c_-'s denominator is strictly positive. Dividing the left- and right-hand side of (36) by the negative of that denominator, we finally get that the actual treatments are beneficial up to period T_a if and only if

$$\delta_- := \frac{\sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(-\delta_{g,\ell})}{\sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(1-D_{g,F_g+\ell})} < c_-. \quad (37)$$

Like δ_+, δ_- can be written as an average total effect per unit of treatment. One may then average δ_+ and δ_-, to form an average total effect per unit of treatment, across initially-untreated and initially-treated groups. Accordingly, let

$$\delta = w_+\delta_+ + (1-w_+)\delta_-,$$

where

$$w_+ = \frac{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}D_{g,F_g+\ell}}{\sum_{g:D_{g,1}=0,F_g\leq T_u}\sum_{\ell=0}^{T_u-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}D_{g,F_g+\ell} + \sum_{g:D_{g,1}=1,F_g\leq T_a}\sum_{\ell=0}^{T_a-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(1-D_{g,F_g+\ell})}.$$

The weight on δ_+ is proportional to the number of periods of exposure to treatment, among observations in initially-untreated groups. The weight on δ_- is proportional to the number of periods of non-exposure to treatment, among observations in initially-treated groups.

Like δ_+, δ_- is equal to a weighted average of event-study reduced-form effects, divided by a weighted average of event-study first-stage effects. Let $L_a = T_a - \min_{g:D_{g,1}=1} F_g$ denote the difference between the last period at which a group has been treated all along and the earliest

period at which a group goes from treated to untreated. Under Assumption 8, $L_a \geq 0$. For every $\ell \in \{0, ..., L_a\}$, let $N^0_\ell = \sum_{g:D_{g,1}=1, F_g \leq T_a - \ell} \beta^{F_g+\ell} N_{g,F_g+\ell} > 0$ be the discounted number of units in groups reaching ℓ periods after their first untreated period at or before T_a. Let

$$\delta_{-,\ell} = \sum_{g:D_{g,1}=1, F_g \leq T_a - \ell} \frac{\beta^{F_g+\ell} N_{g,F_g+\ell}}{N^0_\ell} (-\delta_{g,\ell}) \tag{38}$$
$$\delta^D_{-,\ell} = \sum_{g:D_{g,1}=1, F_g \leq T_a - \ell} \frac{\beta^{F_g+\ell} N_{g,F_g+\ell}}{N^0_\ell} (1 - D_{g,F_g+\ell}).$$

$\delta_{-,\ell}$ is the negative of the average effect of having switched treatment for the first time ℓ periods ago, across all initially-treated groups that got untreated for the first time at least ℓ periods before T_a. $\delta_{-,\ell}$ is an effect of having experienced a weakly higher amount of treatment for $\ell + 1$ periods. $\delta^D_{-,\ell}$ is simply equal to one minus the average treatment of groups that got untreated for the first time ℓ periods ago. Finally, let $w_{-,\ell} = N^0_\ell / \sum_{\ell'=0}^{L_a} N^0_{\ell'}$. Following the same steps as those used to show Lemma 1, one can show that

$$\delta_- = \frac{\sum_{\ell=0}^{L_a} w_{-,\ell} \delta_{-,\ell}}{\sum_{\ell=0}^{L_a} w_{-,\ell} \delta^D_{-,\ell}}. \tag{39}$$

As $\delta_{+,\ell}$ and $\delta_{-,\ell}$ both measure the average effect of having been exposed to a weakly higher amount of treatment for $\ell + 1$ periods, they may be aggregated. Let

$$\delta_\ell = \frac{N^1_\ell}{N^1_\ell + N^0_\ell} \delta_{+,\ell} + \frac{N^0_\ell}{N^1_\ell + N^0_\ell} \delta_{-,\ell}.$$

1.2 Identifying assumption and estimators for initially-treated groups

Let $\mathbf{1}$ denote a $1 \times T$ vector of ones. For initially-treated groups, our estimators rely on the following assumption on $Y_{g,t}(\mathbf{1})$, the always-treated potential outcome.

Assumption 9 *(Independent groups, strong exogeneity, and parallel trends for always-treated outcome)* $\forall t \geq 2$ *and* $\forall g \in \{1, ..., G\}$, $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}) | \boldsymbol{D})$ *does not vary across* g.

Assumption 9 is the equivalent of Assumption 3, for the always-treated potential outcome. Together, Assumptions 3 and 9 imply that the effect of being always versus never treated follows the same evolution over time in every group: $E(Y_{g,t}(\mathbf{1}) - Y_{g,t}(\mathbf{0}) | \boldsymbol{D}) - E(Y_{g,t-1}(\mathbf{1}) - Y_{g,t-1}(\mathbf{0}) | \boldsymbol{D})$ does not vary across g. Imposing jointly Assumptions 3 and 9 is necessary if one wants to estimate the treatment effects both for initially-treated and initially-untreated groups. If one only wants to estimate, say, the treatment effects for initially-untreated groups, imposing Assumption 3 is enough. Placebos similar to those defined in Section 4.3 of the paper could be proposed to test Assumption 9. They are omitted to preserve space.

We start by proposing conditionally unbiased estimators of the $\delta_{g,\ell}$ parameters for initially-treated groups. Let $N_t^a = \sum_{g:D_{g,1}=1,F_g>t} N_{g,t}$ denote the number of observations at period t in groups treated from period 1 to t. For every g such that $D_{g,1} = 1$ and $F_g \leq T_a$ and every $\ell \in \{0, ..., T_a - F_g\}$, note that $N_{F_g+\ell}^a > 0$. Then, let

$$\text{DID}_{g,\ell} = Y_{g,F_g+\ell} - Y_{g,F_g-1} - \sum_{g':D_{g',1}=1,F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N_{F_g+\ell}^a}(Y_{g',F_g+\ell} - Y_{g',F_g-1}).$$

Following the same steps as those used to prove Lemma 2, one can show that if Assumptions 1-2 and 9 hold, then for every g such that $D_{g,1} = 1$ and $F_g \leq T_a$ and every $\ell \in \{0, ..., T_a - F_g\}$, we have $E\left[\text{DID}_{g,\ell}|\boldsymbol{D}\right] = \delta_{g,\ell}$.

Then, we propose conditionally unbiased estimators of $(\delta_{-,\ell})_{0\leq\ell\leq L_a}$, by replacing the $\delta_{g,\ell}$ parameters by their estimators in Equation (38). Specifically, for every $\ell \in \{0, ..., L_a\}$, let

$$\text{DID}_{-,\ell} = \sum_{g:D_{g,1}=1,F_g\leq T_a-\ell} \frac{\beta^{F_g+\ell} N_{g,F_g+\ell}}{N_\ell^0}(-\text{DID}_{g,\ell}). \tag{40}$$

We also propose a conditionally unbiased estimator of δ_-, by replacing the $\delta_{-,\ell}$ parameters by their estimators in Equation (39). Specifically, let

$$\widehat{\delta}_- = \frac{\sum_{\ell=0}^{L_a} w_{-,\ell}\text{DID}_{-,\ell}}{\sum_{\ell=0}^{L_a} w_{-,\ell}\delta_{-,\ell}^D}.$$

Using similar arguments as those used to show Theorem 1 in the paper, one can show that the estimators $(\text{DID}_{-,\ell})_{\ell\in\{0,...,L_a\}}$ and $\widehat{\delta}_-$ are conditionally unbiased under Assumptions 1-2 and 8-9

Turning to the parameters aggregating initially-untreated and initially-treated groups' effects, a conditionally unbiased estimator of δ_ℓ can be defined as

$$\text{DID}_\ell = \frac{N_\ell^1}{N_\ell^1 + N_\ell^0}\text{DID}_{+,\ell} + \frac{N_\ell^0}{N_\ell^1 + N_\ell^0}\text{DID}_{-,\ell}, \tag{41}$$

and a conditionally unbiased estimator of δ can be defined as

$$\widehat{\delta} = w_+\widehat{\delta}_+ + (1 - w_+)\widehat{\delta}_-. \tag{42}$$

Our estimators for initially treated groups are computed by the `did_multiplegt` Stata package, when the `switchers(out)` option is specified. If the `switchers(in)` option is specified, estimators for initially untreated groups are computed. If no option is specified, aggregated estimators across initially treated and untreated groups are computed.

2 Discrete and potentially non-staggered treatments

2.1 Parameters of interest

In this section, we consider the case where the treatment is not binary, but ordered and discrete: $\mathcal{D} = \{0, ..., \overline{d}\}$ for $\overline{d} \geq 1$. For every g, let $T_g = \max_{g':D_{g',1}=D_{g,1}} F_{g'} - 1$ denote the last period where there is still a group with the same treatment as g's in period one and whose treatment has not changed since the start of the panel. For any g such that $F_g \leq T_g$, and for any $\ell \in \{0, ..., T_g - F_g\}$, let $\delta_{g,\ell} = E(Y_{g,F_g+\ell} - Y_{g,F_g+\ell}(D_{g,1}, ..., D_{g,1}))$ be the expected difference between group g's actual outcome at $F_g + \ell$ and the counterfactual "status quo" outcome it would have obtained if its treatment had remained equal to its period-one value from period one to $F_g + \ell$.

We consider designs where Point 1 and either Point 2 or Point 3 (or both) in Assumption 10 below are met.

Assumption 10 *(Non-pathological design for discrete treatments)*

1. $\forall g \in \{1, ..., G\}$, *either* $D_{g,t} \geq D_{g,1}$ *for every* t, *or* $D_{g,t} \leq D_{g,1}$ *for every* t.
2. $\exists (g, g') \in \{1, ..., G\}^2$ *such that* $D_{g,1} = D_{g',1}$, $F_g < F_{g'}$, *and* $D_{g,F_g} > D_{g,1}$.
3. $\exists (g, g') \in \{1, ..., G\}^2$ *such that* $D_{g,1} = D_{g',1}$, $F_g < F_{g'}$, *and* $D_{g,F_g} < D_{g,1}$.

Point 1 requires that groups' treatments are either always weakly higher or always weakly lower than their period-one treatments. As discussed in the paper, for all (g, ℓ) such that ℓ periods after its first treatment change, g has experienced both a treatment strictly below and a treatment strictly above its period-one treatment, $\delta_{g,\ell}$ can be written as a linear combination, with negative weights, of the effects of increasing the treatment at different periods. Accordingly, $\delta_{g,\ell}$ could be negative even if increasing the treatment always increases the outcome. Under Point 1 of Assumption 10, such pairs (g, ℓ) do not exist. If Point 1 fails, we recommend discarding such pairs (g, ℓ). This can be achieved easily, by discarding from the sample all cells (g, t) such that at t, g has experienced both a treatment strictly below and a treatment strictly above its period-one treatment.

For all g such that $F_g \leq T$, let $R_g = 1\{D_{g,F_g} > D_{g,1}\}$ be an indicator equal to 1 (resp. 0) for groups whose first treatment change is an increase (resp. a decrease). Under Point 1 of Assumption 10, groups with $R_g = 1$ are such that $D_{g,t} \geq D_{g,1}$ for all t, so $\delta_{g,\ell}$ is an effect of having been exposed to a weakly higher amount of treatment for $\ell + 1$ periods. Conversely, under Point 1, groups with $R_g = 0$ are such that $D_{g,t} \leq D_{g,1}$ for all t, so $\delta_{g,\ell}$ is an effect of having been exposed to a weakly lower amount of treatment for $\ell + 1$ periods.

Point 2 (resp. 3) of Assumption 10 requires that there is at least one group g experiencing a treatment increase (resp. decrease) at a time period where there is at least another group g' with

the same period-one treatment as g and whose treatment has not changed since the start of the panel. Point 2 (resp. Point 3) is sufficient to ensure that our parameters of interest for groups with $R_g = 1$ (resp. $R_g = 0$) are well-defined and can be unbiasedly estimated. Accordingly, Points 2 and 3 do not need to jointly hold to have that some of our estimators are applicable.

We now define our main parameter of interest, first for groups with $R_g = 1$. For several of the notation we introduce now, the quantity defined below coincides with a quantity already defined in the paper when the treatment is binary. In that case, we use the same notation here as in the paper. In groups with $R_g = 1$, the actual treatments are beneficial relative to the status quo, up to the last period where their treatment effects can be estimated, if and only if

$$\sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}\delta_{g,\ell} - \sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}(D_{g,F_g+\ell}-D_{g,1}) > 0. \quad (43)$$

Let

$$c_+ = \frac{\sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}c_{g,\ell}(D_{g,F_g+\ell}-D_{g,1})}{\sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(D_{g,F_g+\ell}-D_{g,1})}$$

denote the average discounted cost of the treatment, across all the incremental units of treatment received by groups with $R_g = 1$ relative to their period-one treatment. Points 1 and 2 of Assumption 10 ensure that c_+'s denominator is strictly positive. Dividing the left- and right-hand side of (43) by that denominator, we finally get that the actual treatments are beneficial if and only if

$$\delta_+ := \frac{\sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}\delta_{g,\ell}}{\sum_{g:R_g=1,F_g\leq T_g}\sum_{\ell=0}^{T_g-F_g} N_{g,F_g+\ell}\beta^{F_g+\ell}(D_{g,F_g+\ell}-D_{g,1})} > c_+.$$

Here again, δ_+ can be written as an average total effect per unit of treatment.

Let $L_u = \max_{g:R_g=1}(T_g - F_g)$ denote the largest ℓ such that $\delta_{g,\ell}$ can be estimated, for groups g such that $R_g = 1$. Under Point 2 of Assumption 10, $L_u \geq 0$. For every $\ell \in \{0, ..., L_u\}$, let $N^1_\ell = \sum_{g:R_g=1,F_g\leq T_g-\ell}\beta^{F_g+\ell}N_{g,F_g+\ell} > 0$ be the discounted number of units in groups reaching ℓ periods after their first treatment change at or before T_g. Let

$$\delta_{+,\ell} = \sum_{g:R_g=1,F_g\leq T_g-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N^1_\ell}\delta_{g,\ell}, \quad (44)$$

$$\delta^D_{+,\ell} = \sum_{g:R_g=1,F_g\leq T_g-\ell}\frac{\beta^{F_g+\ell}N_{g,F_g+\ell}}{N^1_\ell}(D_{g,F_g+\ell}-D_{g,1}).$$

Following the same steps as those used to show Lemma 1 in the paper, one can show that if Assumption 1 and Point 2 of Assumption 10 hold,

$$\delta_+ = \frac{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta_{+,\ell}}{\sum_{\ell=0}^{L_u} w_{+,\ell}\delta^D_{+,\ell}}. \quad (45)$$

Parameters of interest can be defined symmetrically for groups with $R_g = 0$. Those definitions are omitted to preserve space.

2.2 Identifying assumption and estimators

For any $d \in \mathcal{D}$, let $\boldsymbol{d}$ denote a $1 \times T$ vector of ds. With a discrete treatment, our estimators rely on the following assumption on the potential outcomes $Y_{g,t}(\boldsymbol{d})$, that group g obtains at period t if its treatment is always equal to d.

Assumption 11 *(Independent groups, strong exogeneity, and parallel trends for $Y_{g,t}(\boldsymbol{d})$-potential outcomes) $\forall d \in \mathcal{D}$, $\forall t \geq 2$, and $\forall g \in \{1, ..., G\}$, $E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d})|\boldsymbol{D})$ does not vary across g.*

Assumption 11 generalizes Assumption 3 to all the $Y_{g,t}(\boldsymbol{d})$ potential outcomes. Placebos similar to those defined in Section 4.3 of the paper could be proposed to test Assumption 11. They are omitted to preserve space.

We start by proposing conditionally unbiased estimators of the $\delta_{g,\ell}$ parameters. For all (g,t), let $N_t^g = \sum_{g':D_{g',1}=D_{g,1},F_{g'}>t} N_{g',t}$ denote the number of observations at period t in groups g' with the same period-one treatment as g, and that kept the same treatment from period 1 to t. For every g such that $R_g = 1$ and $F_g \leq T_g$, and every $\ell \in \{0, ..., T_g - F_g\}$, we have $N^g_{F_g+\ell} > 0$. Then, let

$$\text{DID}_{g,\ell} = Y_{g,F_g+\ell} - Y_{g,F_g-1} - \sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N^g_{F_g+\ell}} (Y_{g',F_g+\ell} - Y_{g',F_g-1}).$$

$\text{DID}_{g,\ell}$ compares the $F_g - 1$-to-$F_g + \ell$ outcome evolution, in group g and in groups with g's period-one treatment from period 1 to $F_g + \ell$. Using the same steps as those used to prove Lemma 2 in the paper, one can show that if Assumptions 1-2 and 11 hold, then for every g such that $R_g = 1$ and $F_g \leq T_g$ and every $\ell \in \{0, ..., T_g - F_g\}$, $E\left[\text{DID}_{g,\ell}|\boldsymbol{D}\right] = \delta_{g,\ell}$. Then, unbiased estimators of $\delta_{+,\ell}$ can be formed by replacing $\delta_{g,\ell}$ by $\text{DID}_{g,\ell}$ in Equation (44). Similarly, an unbiased estimator of δ_+ can be formed by replacing the parameters $\delta_{+,\ell}$ by their estimators in Equation (45).

3 Continuous and staggered treatments

Our approach can accommodate some continuous treatments. For instance, it is sometimes the case that $D_{g,t} = I_g 1\{t \geq F_g\}$, as in Proposition 3 in the paper: groups start getting treated at different dates, with potentially continuous group-specific intensities I_g, but once a group gets treated its treatment intensity never changes. The estimators we propose in Section 4 of the paper in the binary-treatment case for initially-untreated groups can still be readily used in that case. In this context, the parameter δ_+ is still a cost-benefit ratio comparing the actual and status-quo treatments of initially untreated groups. The interpretation of the parameters $\delta_{+,\ell}$ slightly changes: those become average effects, across groups that become treated before period $T_u - \ell$, of having received I_g rather than 0 units of treatment for $\ell + 1$ periods. By estimating

the first-stage effects $\delta^D_{+,\ell}$, one can assess the average value of I_g across those groups, which may help interpret the magnitude of $\delta_{+,\ell}$.

A frequent special case of staggered designs with a continuous treatment is when $D_{g,t} = I_g 1\{t \geq F\}$, as in Proposition 1 in the paper: there is no variation in treatment timing, only variation in treatment intensity. Again, the estimators we propose in Section 4 of the paper in the binary-treatment case for initially-untreated groups can still be readily used in that case, provided there are never-treated groups such that $I_g = 0$. Otherwise, there is no valid control group one may use to identify other groups' effects.

4 Other extensions

4.1 Controlling for covariates

Oftentimes, researchers want to control for a vector of covariates $X_{g,t}$ in their estimation. In this section, we propose estimators controlling linearly for time-varying covariates. Another possibility would be to control for time-invariant covariates in a nonparametric way, following Callaway and Sant'Anna (2021). The two approaches rely on non-nested parallel trends assumptions and could in principle be combined. A practical advantage of the approach we propose is that it applies to the case where $X_{g,t}$ is a vector of group-specific linear trends, a popular set of controls among applied researchers. The approach proposed in Callaway and Sant'Anna (2021) cannot accommodate group-specific linear trends.

We need to slightly modify our assumptions. Hereafter, we let $\boldsymbol{X}_g = (X'_{g,1}, ..., X'_{g,T})$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_G)$.

Assumption 12 *(Independence between groups, strong exogeneity and parallel trends with covariates for the never-treated outcome) There is a vector θ_0 of same dimension as $X_{g,t}$ such that $\forall(g,t)$, $t \geq 2$, $E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}) - (X_{g,t} - X_{g,t-1})'\theta_0 | \boldsymbol{D}, \boldsymbol{X}\right)$ does not vary across g.*

Assumption 12 is similar to Assumption 3, except that it includes the covariates in the conditioning, and that the parallel trends assumption is on the "residualized" never-treated potential outcome $Y_{g,t}(\mathbf{0}) - X'_{g,t}\theta_0$. In de Chaisemartin and D'Haultfœuille (2020), we showed that Assumption 12 underlies two-way fixed effects regressions with covariates. Assumption 12 requires that there exist θ_0 and λ_t such that

$$E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}) | \boldsymbol{D}, \boldsymbol{X}\right) = (X_{g,t} - X_{g,t-1})'\theta_0 + \lambda_t.$$

Then, Assumption 12 allows for the possibility that groups experience different evolutions of their never-treated outcome over time, but it requires that those differential evolutions be accounted

for by a linear model in $X_{g,t} - X_{g,t-1}$, the change in a group's covariates. An interesting special case is when the control variables are group-specific linear trends. Then, Assumption 12 requires that

$$E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})|\boldsymbol{D}\right) = \theta_{0g} + \lambda_t,$$

for some constants θ_{0g} and λ_t. From $t-1$ to t, the evolution of the never-treated outcome in group g should deviate from its group-specific linear trend θ_{0g} by an amount λ_t common to all groups. Then, Assumption 12 is a "common deviation from linear trends" assumption, which may be more plausible than the standard parallel trends assumption.

Let $\widehat{\theta}_0$ denote the coefficient of $X_{g,t} - X_{g,t-1}$ in the OLS regression of $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ and time fixed effects, in the sample of all (g,t) such that $F_{g,1} > t$. For every $g : D_{g,1} = 0, F_g \leq T_u$ and every $\ell \in \{0, ..., T_u - F_g\}$, let

$$\begin{aligned}\text{DID}^X_{g,\ell} =& Y_{g,F_g+\ell} - Y_{g,F_g-1} - (X_{g,F_g+\ell} - X_{g,F_g-1})'\widehat{\theta}_0 \\ &- \sum_{g':D_{g',1}=0, F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N^u_{F_g+\ell}} (Y_{g',F_g+\ell} - Y_{g',F_g-1} - (X_{g',F_g+\ell} - X_{g',F_g-1})'\widehat{\theta}_0).\end{aligned}$$

$\text{DID}^X_{g,\ell}$ is similar to $\text{DID}_{g,\ell}$, except that instead of comparing groups' outcome evolution, it compares the part of that evolution that is not explained by a change in covariates.

Following the same steps as those used to prove Lemma 2, one can show that if Assumptions 1-2 and 12 hold, then for every g such that $D_{g,1} = 1$ and $F_g \leq T_u$, and every $\ell \in \{0, ..., T_u - F_g\}$, we have $E\left[\text{DID}^X_{g,\ell} \middle| \boldsymbol{D}, \boldsymbol{X}\right] = \delta_{g,\ell}$. Then, conditionally unbiased estimators of $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ controlling for the covariates can be proposed, by replacing the $\delta_{g,\ell}$ parameters by their estimators $\text{DID}^X_{g,\ell}$ in Equation (8). Similarly, a conditionally unbiased estimator of δ_+ can be proposed, by replacing the $\delta_{+,\ell}$ parameters by their estimators controlling for covariates in Equation (9). Using similar arguments as those used to show Theorem 1 in the paper, one can show that those estimators controlling for covariates are conditionally unbiased under Assumptions 1-2, 4, and 12. One can follow the exact same steps to propose placebo estimators to test Assumptions 2 and 12.

Our estimators controlling for covariates are computed by the `did_multiplegt` Stata package, when the `controls` option is specified. When the controls are group-specific linear trends, the `trends_lin` option needs to be specified.

4.2 Allowing for different trends across sets of groups

In some cases, controlling for covariates may be insufficient to account for differences in trends between groups. Then, a common remedy in static or dynamic two-way fixed effect regressions consists in including interactions between time FE and FE for sets of groups. For instance, if groups are US counties, one can allow for state-specific trends. A similar idea can be pursued

in our context. Let us index sets of groups by $s \in \{1, ..., S\}$. In this set-up, we modify our assumptions as follows.

Assumption 13 *(Parallel trends within sets of groups) For all* $s \in \{1, ..., S\}$, $E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})|\boldsymbol{D})$ *does not vary across* $g \in s$.

Assumption 13 is a weakening of Assumption 3, as it only requires that the never-treated potential outcome of groups in the same set of groups follow the same evolution over time.

Assumption 14 *(Non-pathological design, with parallel trends within sets of groups)*

1. *There is at least one* $s \in \{1, ..., S\}$ *such that there exists* $(g, g') \in s^2$ *such that* $D_{g,1} = D_{g',1} = 0$ *and* $F_{g,1} < F_{g',1}$.
2. $T_u^s = \max_{g:D_{g,1}=0, g\in s} F_g - 1$ *does not depend on* s.

Point 1 of Assumption 14 requires that there is at least one set of groups within which an initially-untreated group gets treated for the first time at a date where another initially-untreated group is still untreated. Point 2 requires that the last period where at least one group has been untreated since the start of the panel be the same in every set of groups. This may hold when there are many never-treated groups, so that each set of groups may have at least one never-treated group. Outside of that case, this condition may be more restrictive. The only purpose of this condition is to ensure that under parallel trends within sets of groups, one can still unbiasedly estimate the parameters $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ and δ_+, thus greatly alleviating the notational burden. If that condition fails, there are sets of groups for which $\delta_{g,\ell}$ cannot be unbiasedly estimated for all $\ell \in \{0, ..., L_u\}$, because $T_u^s < T_u$, so we would need to redefine the parameters $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ and δ_+. This extension is conceptually straightforward but notationally burdensome. We omit it to preserve space.

For every $s \in \{1, ..., S\}$ and $t \in \{1, ..., T\}$, let $N_t^{u,s} = \sum_{g\in s:D_{g,1}=0,F_g>t} N_{g,t}$ denote the number of observations at period t in groups $g \in s$ untreated from period 1 to t. For every $s \in \{1, ..., S\}$, for every $g \in s$ such that $D_{g,1} = 0$ and $F_g \leq T_u$, and every $\ell \in \{0, ..., T_u - F_g\}$, let

$$\text{DID}_{g,\ell}^s = Y_{g,F_g+\ell} - Y_{g,F_g-1} - \sum_{g'\in s:D_{g',1}=0,F_{g'}>F_g+\ell} \frac{N_{g',F_g+\ell}}{N_{F_g+\ell}^{u,s}}(Y_{g',F_g+\ell} - Y_{g',F_g-1}).$$

$\text{DID}_{g,\ell}^s$ is similar to $\text{DID}_{g,\ell}$, except that it only compares the outcome evolution of groups in the same set of groups s. For instance, if groups are US counties and sets of groups are US states, $\text{DID}_{g,\ell}^s$ compares the outcome evolution of counties in the same state.

Following the same steps as those used to prove Lemma 2, one can show that if Assumptions 1-2 and 13-14 hold, then for every $s \in \{1, ..., S\}$, $g \in s$ such that $D_{g,1} = 0, F_g \leq T_u$ and $\ell \in \{0, ..., T_u - F_g\}$, we have $E\left[\text{DID}_{g,\ell}^s \middle| \boldsymbol{D}\right] = \delta_{g,\ell}$. Then, conditionally unbiased estimators of

$(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ relying only on parallel trends within sets of groups can be proposed, by replacing the $\delta_{g,\ell}$ parameters by their estimators $\text{DID}^s_{g,\ell}$ in Equation (8). Similarly, a conditionally unbiased estimator of δ_+ can be proposed, by replacing the $\delta_{+,\ell}$ parameters by their estimators relying only on parallel trends within sets of groups in Equation (9). Using similar arguments as those used to show Theorem 1 in the paper, one can show that those estimators are conditionally unbiased under Assumptions 1-2 and 13-14. One can follow the exact same steps to propose placebo estimators to test Assumptions 2 and 13.

Our estimators relying on parallel trends within sets of groups are computed by the `did_multiplegt` Stata package, when the `trends_nonparam` option is specified.

4.3 Fuzzy designs

In this subsection, we briefly discuss fuzzy designs, where the treatment varies within (g,t) cells. The estimators we propose in this paper are readily applicable to groups that are fully untreated at period 1. Then, one just need to redefine the parameters $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ and δ_+, replacing $Y_{g,t}(\boldsymbol{D})$ by $\frac{1}{N_{g,t}}\sum_{i=1}^{N_{g,t}} Y_{i,g,t}(\boldsymbol{D}_{i,g})$, where $\boldsymbol{D}_{i,g} = (D_{i,g,1},...,D_{i,g,T})$ is a vector stacking the treatments of observation i in group g. Then, one can show that under Assumptions 2-4, $(\text{DID}_{+,\ell})_{0\leq\ell\leq L_u}$ and $\widehat{\delta}_+$ are unconditionally unbiased for $(\delta_{+,\ell})_{0\leq\ell\leq L_u}$ and δ_+.

In groups partly treated at period 1, the estimators proposed in this paper are not applicable. First, only groups where the proportion of treated units does not change over time can be used as controls, and such groups may not exist. When such groups exist, two estimation strategies, that differ from the ones considered in this paper, are available. First, one may separately estimate the outcome evolution of treated and untreated units in the control groups, and apply those counterfactual evolutions separately to treated and control units in groups whose proportion of treated units does change, and thus recover the counterfactual outcome those groups would have experienced if their proportion of treated units had not changed (see de Chaisemartin and D'Haultfœuille, 2018). Second, one may redefine the treatment as a continuous, group-level variable, and apply the estimators proposed by de Chaisemartin et al. (2022).

4.4 Ruling out lagged treatment effects

Up to now, we have made no restriction on the effects of past treatments. We now investigate the benefits of imposing such restrictions. Specifically, we consider the following assumption.

Assumption 15-k *(No effect of past treatments beyond k lags)*

For all (g,t), all t, and all $(d_1,...,d_t)\in\{0,1\}^t$, $Y_{g,t}((d_{t'})_{t'\leq t}) = Y_{g,t}(d_{t-k},...,d_t)$.

Assumption 15-k is plausible when the treatment is unlikely to have very long-run effects. It is commonly made in event-study regressions and extensively discussed in Borusyak et al. (2021) and in Schmidheiny and Siegloch (2020) as a possible way to identify these regressions.

In our context, imposing Assumption 15 can provide a solution to the "initial conditions" problem. So far, we have assumed that the treatments prior to the start of the panel $(D_{g,t})_{t\leq 0}$ do not affect potential outcomes. There are at least two situations where this assumption is innocuous. First, such treatments may not exist. Assume for instance that one seeks to estimate the effect of being unionized on earnings, using the NLSY panel. In this data set, $t = 1$ corresponds to the first year on the labor market, so $D_{g,t}$ is not defined for $t \leq 0$. Second, in staggered adoption designs, our main results still hold if potential outcomes depend on $(D_{g,t})_{t\leq 0}$. In groups g untreated at period 1, $D_{g,t} = 0$ for all $t \leq 0$, so our results still hold. Outside of those designs, our results do not apply when potential outcomes depend on $(D_{g,t})_{t\leq 0}$. However, under Assumption 15-k, our results apply to a restricted panel, including groups with a stable treatment from period 1 to $k + 1$, and starting at period $k + 1$. Note that a similar idea was already put forward by Schmidheiny and Siegloch (2020) in the context of event-study regressions, except that in the context of those regressions one only needs to drop the first $k+1$ periods of the panel, while we also need to drop groups whose treatment changes at some point between periods 1 and $k+1$.

4.5 First-difference placebo estimators

We now define the first-difference placebo estimators of Assumptions 2 and 3 discussed in Section 4.3 of the paper. First, for any g such that $D_{g,1} = 0$ and $3 \leq F_g \leq T_u$ and for any $\ell \in \{0, ..., F_g - 3\}$, let

$$\text{DID}^{\text{fdp}}_{g,\ell} = Y_{g,F_g-1-\ell} - Y_{g,F_g-2-\ell} - \sum_{g':D_{g',1}=0,F_{g'}>F_g} \frac{N_{g',F_g}}{N^u_{F_g}} (Y_{g',F_g-1-\ell} - Y_{g',F_g-2-\ell}).$$

"fdp" stands for first-difference placebo. $\text{DID}^{\text{fdp}}_{g,\ell}$ compares the outcome evolution of group g and of groups never treated from period 1 to F_g, between two consecutive periods, $\ell + 1$ periods before g got treated. Let $L^{\text{fdp}}_u = \max\{t \leq T_u : \exists g : F_g = t\} - 3$ denote the last time period at which a group switches from untreated to treated for the first time while there is still a group that has always been untreated, minus three. For every $\ell \in \{0, ..., L^{\text{fdp}}_u\}$, let $N^{1,\text{fdp}}_\ell = \sum_{g:F_g-3\geq\ell} \beta^{F_g} N_{g,F_g} > 0$, and let

$$\text{DID}^{\text{fdp}}_{+,\ell} = \sum_{g:F_g-3\geq\ell} \frac{\beta^{F_g} N_{g,F_g}}{N^{1,\text{fdp}}_\ell} \text{DID}^{\text{fdp}}_{g,\ell}.$$

As with the long-difference placebos, one can show that if Assumptions 1-3 hold,

$$E\left[\text{DID}^{\text{fdp}}_{+,\ell} \middle| \boldsymbol{D}\right] = 0 \quad \forall \ell \in \{0, ..., L^{\text{fdp}}_u\}.$$

First-difference placebos are computed by the `did_multiplegt` Stata package, when the `placebo` option is specified. If the `longdiff_placebo` is also specified, the command instead computes the long-difference placebo estimators defined in Section 4.3 of the paper.

5 Supplementary results on the reanalysis of Favara and Imbs (2015)

This section collects supplementary results on our reanalysis of Favara and Imbs (2015).

First, Table 1 below shows the coefficients from a distributed-lag TWFE regression of the logarithm of loans volume on county and year fixed effects, the current treatment, and seven lags of the treatment, controlling for state-specific linear trends. All coefficients are small and statistically insignificant.

Table 1: Coefficients from distributed-lag TWFE regression of the logarithm of loans volume on banking deregulation

Period-t deregulation	-0.0443
	(0.0504)
Period-$t-1$ deregulation	-0.0177
	(0.0578)
Period-$t-2$ deregulation	-0.0280
	(0.0498)
Period-$t-3$ deregulation	-0.0106
	(0.0391)
Period-$t-4$ deregulation	0.0118
	(0.0219)
Period-$t-5$ deregulation	0.0120
	(0.0160)
Period-$t-6$ deregulation	0.0040
	(0.0127)
Period-$t-7$ deregulation	0.0005
	(0.0120)

Notes: The table shows the coefficients from a distributed-lag TWFE regression of the logarithm of loans volume on county and year fixed effects, the current treatment, and seven lags of the treatment, controlling for state-specific linear trends. The regression is estimated using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015). Standard errors are estimated using 100 bootstrap replications clustered at the state level. The regression is weighted by the inverse of the number of counties per state as in Favara and Imbs (2015).

Second, we study the effect of banking deregulations on houses prices. The blue line on Figure 1 shows our DID estimates of the effect of a first banking-deregulation on the logarithm of houses prices. $\text{DID}_{+,0} = 0.004$ (s.e.=0.004): in the year of the first deregulation, house prices increase by 0.4% more in counties that deregulate than in counties that do not, an insignificant difference. This effect builds up over time, and becomes significant at the 10% level after 3 years ($\text{DID}_{+,3} = 0.045$, s.e.=0.024) and at the 5% level after 4 years ($\text{DID}_{+,4} = 0.073$, s.e.=0.031). To the left of zero, the placebo estimates are much smaller than the actual estimates (all are below 0.006 in absolute value) and insignificant: an F-test cannot reject the null that all placebos are insignificant (p-value=0.642). The green line on Figure 1 shows estimates of $(\delta_{+,\ell})_{\ell \in \{0,\ldots,7\}}$ using a distributed-lag TWFE regression of the logarithm of houses prices on county and year fixed effects, the current treatment, and seven lags of the treatment, controlling for state-specific linear trends. The TWFE- and DID-based estimates are close, and insignificantly different. The standard errors of the TWFE-based estimates are much larger than those of the DID-based estimates, and all the TWFE-based estimates are insignificant.

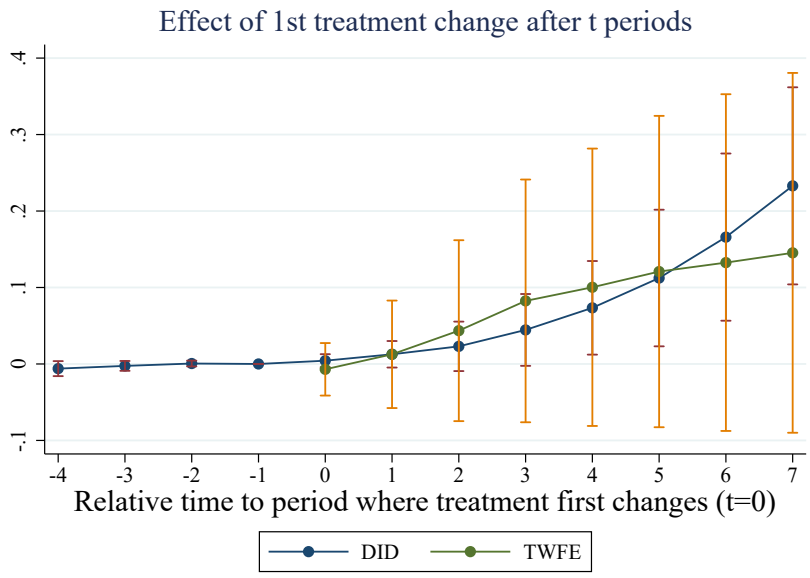

Figure 1: Effect of banking deregulations on housing prices.

Notes: To the right of zero, the blue line on the figure shows the $\text{DID}_{+,\ell}$ estimates of the effect of a first banking-deregulation episode on the logarithm of house prices, the year of that first deregulation episode, and in later years. To the left of zero, the blue line shows the $\text{DID}^{\text{pl}}_{+,\ell}$ placebo estimates. At $x = -1$, the placebo is normalized to 0. $\text{DID}^{\text{pl}}_{+,0}$ is shown at $x = -2$, etc. The estimates on the blue line are computed by the Stata `did_multiplegt` command, using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015). To the right of zero, the green line shows estimates of $(\delta_{+,\ell})_{\ell\in\{0,...,7\}}$ based on a distributed-lag TWFE regression of the logarithm of house prices on county and year fixed effects, the current treatment, and seven lags of the treatment. Standard errors are estimated using 100 bootstrap replications clustered at the state level. 95% confidence intervals relying on a normal approximation are shown in red for the $\text{DID}_{+,\ell}$ and $\text{DID}^{\text{pl}}_{+,\ell}$ estimates, and in orange for the TWFE estimates. Both sets of estimators have state-specific linear trends, and are weighted by the inverse of the number of counties per state as in Favara and Imbs (2015).

Combining the estimates shown on the blue line of Figure 1 to the first-stage estimated on Figure 2 in the paper leads to $\widehat{\delta}_{+,0:7} = 0.037$ (s.e.=0.014). Across all deregulations, one deregulation increases house prices by 3.7% on average, when one sums the deregulation's instantaneous and dynamic effects. Even though this goes beyond the scope of this paper, one may tentatively divide the $\widehat{\delta}_{+,0:7}$ estimator for housing prices by that for loan volume. Doing so, we obtain an elasticity of housing prices to loan volume of 0.137.

Third, we investigate whether differences between our specification and that in Favara and Imbs (2015) can explain why we reach such different conclusions. There are two differences between our specifications. First, their outcome variables are first-differences of the logarithm of loan volume and housing prices, while our outcome variables are in levels. Second, their treatment

variable is actually the lag of regulations in county g. We recompute our estimators mimicking their specification, and still find much larger long- than short-run effects. Figure 2 below shows the results for housing prices. Only six dynamic effects are shown, because the treatment is now lagged. Results are similar for loan volume.

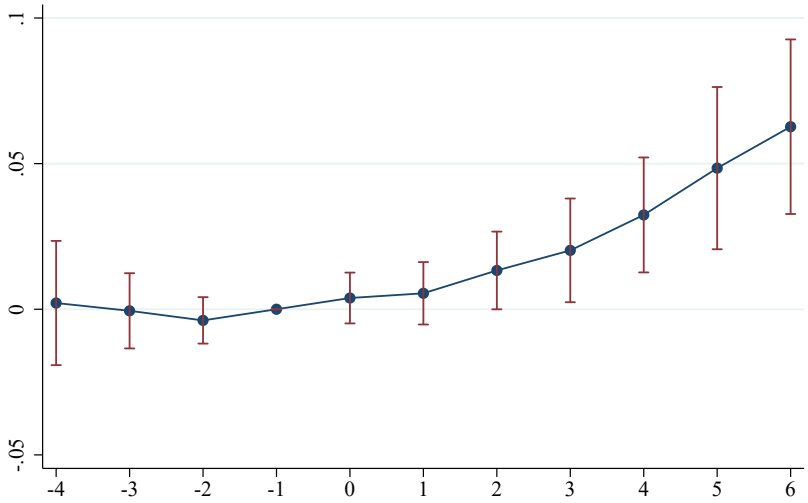

Figure 2: Effect of banking deregulations on housing prices, same specification as in Favara and Imbs (2015)

Notes: To the right of zero, the figure shows the $\text{DID}_{+,\ell}$ estimates of the effect of the lag of a first banking-deregulation episode on the first difference of the logarithm of housing prices. To the left, the figure shows the $\text{DID}^{\text{pl}}_{+,\ell}$ placebo estimates, that should not significantly differ from 0 if the parallel trends condition is satisfied. At $x = -1$, the placebo is normalized to 0. $\text{DID}^{\text{pl}}_{+,0}$ is shown at $x = -2$, etc. The estimates are computed by the Stata `did_multiplegt` command, using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015). Standard errors are estimated using 100 bootstrap replications clustered at the state level. 95% confidence intervals relying on a normal approximation are shown in red. The estimation is weighted by the inverse of the number of counties per state as in Favara and Imbs (2015).

6 Details on the literature review

Table 2: List of highly-cited papers published by the AER from 2015 to 2019 using TWFE regressions

Reference	Design	Estimate dynamic effects?
Dell (2015)	Not binary and/or not staggered	No
Burgess et al. (2015)	Not binary and/or not staggered	No
Favara and Imbs (2015)	Not binary and/or not staggered	Yes
Pierce and Schott (2016)	Not binary and/or not staggered	Yes
Hoynes et al. (2016)	Binary, staggered	Yes
Munshi and Rosenzweig (2016)	Not binary and/or not staggered	No
Atkin (2016)	Not binary and/or not staggered	No
Allcott et al. (2016)	Not binary and/or not staggered	No
Suárez Serrato and Zidar (2016)	Not binary and/or not staggered	Yes
Di Maggio et al. (2017)	Binary, staggered	Yes
Brandt et al. (2017)	Not binary and/or not staggered	No
Berman et al. (2017)	Not binary and/or not staggered	No
Handley and Limao (2017)	Not binary and/or not staggered	No
Dix-Carneiro and Kovak (2017)	Not binary and/or not staggered	Yes
Besley et al. (2017)	Not binary and/or not staggered	Yes
Donaldson (2018)	Not binary and/or not staggered*	No
Fuest et al. (2018)	Not binary and/or not staggered	Yes
Hershbein and Kahn (2018)	Not binary and/or not staggered	Yes
Monte et al. (2018)	Binary, staggered	No
Huber (2018)	Not binary and/or not staggered	No
Antecol et al. (2018)	Binary, staggered	Yes
Fetzer (2019)	Not binary and/or not staggered	Yes
Kaur (2019)	Not binary and/or not staggered	Yes
Naritomi (2019)	Binary, no variation in timing	Yes
Bloom et al. (2019)	Not binary and/or not staggered	No
Diamond et al. (2019)	Binary, no variation in timing	No

*: Except in one table in the paper, where the treatment is binary and staggered.

References

Allcott, H., Collard-Wexler, A. and O'Connell, S. D. (2016), 'How do electricity shortages affect industry? evidence from india', *American Economic Review* **106**(3), 587–624.

Antecol, H., Bedard, K. and Stearns, J. (2018), 'Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?', *American Economic Review* **108**(9), 2420–41.

Atkin, D. (2016), 'Endogenous skill acquisition and export manufacturing in mexico', *American Economic Review* **106**(8), 2046–85.

Berman, N., Couttenier, M., Rohner, D. and Thoenig, M. (2017), 'This mine is mine! how minerals fuel conflicts in africa', *American Economic Review* **107**(6), 1564–1610.

Besley, T., Folke, O., Persson, T. and Rickne, J. (2017), 'Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden', *American economic review* **107**(8), 2204–42.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I. and Van Reenen, J. (2019), 'What drives differences in management practices?', *American Economic Review* **109**(5), 1648–83.

Borusyak, K., Jaravel, X. and Spiess, J. (2021), Revisiting event study designs: Robust and efficient estimation. arXiv preprint arXiv:2108.12419.

Brandt, L., Van Biesebroeck, J., Wang, L. and Zhang, Y. (2017), 'Wto accession and performance of chinese manufacturing firms', *American Economic Review* **107**(9), 2784–2820.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A. and Padró i Miquel, G. (2015), 'The value of democracy: evidence from road building in kenya', *American Economic Review* **105**(6), 1817–51.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

de Chaisemartin, C. and D'Haultfœuille, X. (2018), 'Fuzzy differences-in-differences', *The Review of Economic Studies* **85**(2), 999–1028.

de Chaisemartin, C. and D'Haultfœuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.

de Chaisemartin, C., D'Haultfœuille, X., Pasquier, F. and Vazquez-Bare, G. (2022), Difference-in-differences estimators for treatments continuously distributed at every period. arXiv preprint arXiv:2201.06898.

Dell, M. (2015), 'Trafficking networks and the mexican drug war', *American Economic Review* **105**(6), 1738–79.

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A. and Yao, V. (2017), 'Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging', *American Economic Review* **107**(11), 3550–88.

Diamond, R., McQuade, T. and Qian, F. (2019), 'The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco', *American Economic Review* **109**(9), 3365–94.

Dix-Carneiro, R. and Kovak, B. K. (2017), 'Trade liberalization and regional dynamics', *American Economic Review* **107**(10), 2908–46.

Donaldson, D. (2018), 'Railroads of the raj: Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Favara, G. and Imbs, J. (2015), 'Credit supply and the price of housing', *American Economic Review* **105**(3), 958–92.

Fetzer, T. (2019), 'Did austerity cause brexit?', *American Economic Review* **109**(11), 3849–86.

Fuest, C., Peichl, A. and Siegloch, S. (2018), 'Do higher corporate taxes reduce wages? micro evidence from germany', *American Economic Review* **108**(2), 393–418.

Handley, K. and Limao, N. (2017), 'Policy uncertainty, trade, and welfare: Theory and evidence for china and the united states', *American Economic Review* **107**(9), 2731–83.

Hershbein, B. and Kahn, L. B. (2018), 'Do recessions accelerate routine-biased technological change? evidence from vacancy postings', *American Economic Review* **108**(7), 1737–72.

Hoynes, H., Schanzenbach, D. W. and Almond, D. (2016), 'Long-run impacts of childhood access to the safety net', *American Economic Review* **106**(4), 903–34.

Huber, K. (2018), 'Disentangling the effects of a banking crisis: Evidence from german firms and counties', *American Economic Review* **108**(3), 868–98.

Kaur, S. (2019), 'Nominal wage rigidity in village labor markets', *American Economic Review* **109**(10), 3585–3616.

Monte, F., Redding, S. J. and Rossi-Hansberg, E. (2018), 'Commuting, migration, and local employment elasticities', *American Economic Review* **108**(12), 3855–90.

Munshi, K. and Rosenzweig, M. (2016), 'Networks and misallocation: Insurance, migration, and the rural-urban wage gap', *American Economic Review* **106**(1), 46–98.

Naritomi, J. (2019), 'Consumers as tax auditors', *American Economic Review* **109**(9), 3031–72.

Pierce, J. R. and Schott, P. K. (2016), 'The surprisingly swift decline of us manufacturing employment', *American Economic Review* **106**(7), 1632–62.

Schmidheiny, K. and Siegloch, S. (2020), On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. ZEW Discussion Paper 20-01.

Suárez Serrato, J. C. and Zidar, O. (2016), 'Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms', *American Economic Review* **106**(9), 2582–2624.