NBER WORKING PAPER SERIES

DIFFERENCE-IN-DIFFERENCES ESTIMATORS OF INTERTEMPORAL TREATMENT EFFECTS

Clément de Chaisemartin
Xavier D'Haultfoeuille

Working Paper 29873
http://www.nber.org/papers/w29873

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2022, revised July 2023

We are very grateful to Mélitine Malézieux and Doulo Sow for their outstanding research assistance. We are also very grateful to Isabelle Méjean, Ashesh Rambachan, Jonathan Roth, Pedro Sant'Anna, Jesse Shapiro, the editor, and four anonymous referees for their helpful comments. Clément de Chaisemartin was funded by the European Union (ERC, REALLYCREDIBLE,GA N°101043899). Views and opinions expressed are those of the authors and do not reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Xavier D'Haultfoeuille gratefully acknowledges financial support from the research grants Otelo (ANR-17-CE26-0015-041). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Difference-in-Differences Estimators of Intertemporal Treatment Effects
Clément de Chaisemartin and Xavier D'Haultfoeuille
NBER Working Paper No. 29873
March 2022, revised July 2023
JEL No. C21,C23

ABSTRACT

We study treatment-effect estimation using panel data. The treatment may be nonbinary, non-absorbing, and the outcome may be affected by the treatment lags. We make parallel-trends assumptions, but do not restrict treatment effect heterogeneity, unlike commonly-used two-way-fixed-effects regressions. We propose reduced-form event-study estimators of the effect of being exposed to a weakly higher treatment dose for periods. We also propose normalized event-study estimators, that estimate a weighted average of the effects of the current treatment and its lags. Finally, we show that the reduced-form estimators can be combined into an economically interpretable cost-benefit ratio.

Clément de Chaisemartin
Department of Economics
University of California at Santa Barbara
Santa Barbara, CA 93106
clementdechaisemartin@ucsb.edu

Xavier D'Haultfoeuille
CREST
5 avenue Henry Le Chatelier
91764 Palaiseau cedex
FRANCE
xavier.dhaultfoeuille@ensae.fr

1 Introduction

We study treatment-effect estimation, using a panel of groups, indexed by g, observed at several time periods, indexed by t. $Y_{g,t}$, group g's period-t outcome, may be affected by $D_{g,t}$, group g's period-t treatment, but also by g's lagged treatments. In such instances, to estimate treatment effects, researchers often use two-way fixed effects (TWFE) regressions of $Y_{g,t}$ on group fixed effects, period fixed effects, and some measures of contemporaneous and past exposure to the treatment. A recent literature, reviewed below, has shown that TWFE regressions are not robust to heterogeneous treatment effects across groups and/or over time. Another strand of literature, also reviewed below, has proposed heterogeneity-robust difference-in-differences (DID) estimators, relying on parallel-trends assumptions, like TWFE estimators, but robust to heterogeneous treatment effects. In that second strand, most of the papers that allow lagged treatments to affect the outcome assume that treatment is binary and staggered, meaning absorbing. To our knowledge, this paper is the first to have proposed and studied heterogeneity-robust DID estimators, in applications where the treatment is either non-binary and/or non-staggered, and where the lagged treatments may affect the outcome. Applications with a non-binary and/or non-staggered treatment are common. Of the 100 most-cited papers published by the American Economic Review (AER) from 2015 to 2019, we find that 26 estimate a TWFE regression, but only four have a binary-and-staggered treatment. Our estimators are widely applicable: essentially, they can be used in any design where groups' period-one treatment takes a finite number of values.[1] They are computed by the `did_multiplegt_dyn` Stata package.[2]

Let $Y_{g,t}(d_1, ..., d_t)$ denote the potential outcome of g at t, if its treatments from period 1 to t are equal to $(d_1, ..., d_t)$. Let F_g denote the first period at which group g's treatment changes, and

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

be the expected difference between g's actual outcome at $F_g - 1 + \ell$ 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 $F_g - 1 + \ell$. To estimate $\delta_{g,\ell}$, we propose an estimator $\text{DID}_{g,\ell}$

[1] de Chaisemartin et al. (2022) extend our estimators to designs where the period-one treatment is continuous.

[2] `did_multiplegt_dyn` supersedes the `did_multiplegt` Stata package, to estimate treatment effects allowing for effects of lagged treatments on the outcome.

comparing the $F_g - 1$-to-$F_g - 1 + \ell$ outcome evolution between group g, and groups whose treatment has not changed yet at $F_g - 1 + \ell$, and with the same treatment as g at period one. This last requirement is important: we show that comparing switchers and non-switchers with different period-one treatments relies on a parallel-trends assumption that, essentially, rules out both effects of the lagged treatments on the outcome and time-varying treatment effects. To test the weaker parallel-trends assumption underlying our $\text{DID}_{g,\ell}$ estimator, we propose placebo estimators comparing the outcome trends of switchers and non-switchers with the same period-one treatment, before switchers switch.

We then aggregate the $\text{DID}_{g,\ell}$ estimators across groups. We discard from the aggregation the $\text{DID}_{g,\ell}$s such that at $F_g - 1 + \ell$, g has experienced both a strictly lower and a strictly larger treatment than its period-one treatment: for such (g, ℓ)s, $\delta_{g,\ell}$ can be written as a linear combination, with negative weights, of the effects of increasing different treatment lags. Thus, $\delta_{g,\ell}$ does not satisfy the following no-sign reversal property: one may have that $(d_1, ..., d_t) \mapsto Y_{g,t}(d_1, ..., d_t)$ is increasing in each of its arguments, but $\delta_{g,\ell}$ is negative.[3] Among the remaining (g, ℓ)s, either

$$D_{g,F_g} > D_{g,1}, D_{g,F_g+1} \geq D_{g,1}, ..., D_{g,F_g-1+\ell} \geq D_{g,1},$$
$$\text{or } D_{g,F_g} < D_{g,1}, D_{g,F_g+1} \leq D_{g,1}, ..., D_{g,F_g-1+\ell} \leq D_{g,1}.$$

In the former case, $\delta_{g,\ell}$ is the effect of having been exposed to a weakly higher treatment for ℓ periods, but in the latter case $\delta_{g,\ell}$ is the effect of having been exposed to a weakly lower treatment for ℓ periods. Accordingly, when we aggregate the $\text{DID}_{g,\ell}$ estimators, we multiply by minus one the $\text{DID}_{g,\ell}$ for groups such that $D_{g,F_g} < D_{g,1}$. This finally yields a DID_ℓ estimator of the effect of having been exposed to a weakly higher treatment dose for ℓ periods.

Beyond this "reduced-form" interpretation, lending a more structural interpretation to the DID_ℓ estimators might be difficult in the potentially complicated designs we consider. For instance, with three periods and three groups such that $(D_{1,1} = 0, D_{1,2} = 4, D_{1,3} = 0)$, $(D_{2,1} = 0, D_{2,2} =$

[3]The no-sign reversal property was introduced in a static model where the outcome is only affected by the current treatment (Small and Tan, 2007). When the outcome can be affected by the current and lagged treatments, this property is well-defined only under the assumption that all those treatments affect the outcome in the same direction. Otherwise, an estimand can never be of the "wrong" sign. The assumption that the current and lagged treatments all affect the outcome in the same direction is of course very strong. Thus, satisfying our extension of no-sign reversal is an even more minimal requirement than in a static model.

$2, D_{2,3} = 3)$, and $(D_{3,1} = 0, D_{3,2} = 0, D_{3,3} = 0)$, DID_2 estimates the average of $E(Y_{1,3}(0,4,0) - Y_{1,3}(0,0,0))$ and $E(Y_{2,3}(0,2,3) - Y_{2,3}(0,0,0))$, so DID_2 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 three strategies. First, when the number of $F_g - 1$-to-$F_g - 1 + \ell$ treatment trajectories is low relative to the number of groups, one may estimate trajectory-specific versions of DID_ℓ, which may be more easily interpretable. Second, we propose to normalize our DID_ℓ estimators by the average of $\sum_{k=1}^{\ell} |D_{g,F_g-1+k} - D_{g,1}|$, the total treatment increments from F_g to $F_g - 1 + \ell$ that generate $\delta_{g,\ell}$. We show that DID^n_ℓ, the normalized DID_ℓ, is unbiased for a weighted average of the effect of the current treatment and of its $\ell - 1$ first lags on the outcome. We also show that DID^n_ℓ can be used to test a null hypothesis which is often of great interest, namely whether the current and lagged treatments all have the same effect on the outcome. Third, we show that a weighted sum across ℓ of the DID_ℓs is unbiased for a parameter with a clear economic interpretation, which may be used to conduct a cost-benefit analysis assessing if the treatment changes that took place over the panel led to a better situation than the one that would have prevailed if no 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 effects of a treatment increment, at the time when it takes place and at later periods.

In our aforementioned census of highly-cited TWFE papers published in the AER, 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 a panel-data version of the local-projection method proposed by Jordà (2005), while other researchers have estimated TWFE regressions of the outcome on the treatment and its first K lags, the so-called distributed-lag regression. We show that under the parallel-trends assumption underlying our estimators, those three regressions may not estimate a convex combination of the $\delta_{g,\ell}$s. Thus, they may not satisfy the no-sign reversal property. The local-projection

and distributed-lag regressions also suffer from a contamination problem: coefficients supposed to estimate the effect of a given length of exposure to treatment or of a given treatment lag are actually contaminated by effects of other exposure lengths or other lags. Strikingly, the local-projection regression may yield biased estimators, possibly of the wrong sign, even if the treatment effect is constant across groups and over time. Therefore, despite their limitations, our estimators improve on current practice.

Finally, we use our estimators to revisit Favara and Imbs (2015). These authors use the differential timing and intensity of banking deregulations in US states during the 1990s to estimate deregulations' effects on credit supply. Our estimators suggest that banking deregulations have very persistent effects on the growth of credit supply and house prices. These findings differ sharply from those obtained by Favara and Imbs (2015). Using the panel-data version of the local-projection method, the authors find significant short-run effects, which quickly vanish.

The paper is organized as follows. Section 2 introduces the set up and our assumptions. Section 3 defines our parameters of interest and presents our estimators. Section 4 studies the properties of commonly used TWFE regressions to estimate current and lagged treatment effects in general designs. Section 5 uses our estimators to revisit Favara and Imbs (2015). Finally, the Web Appendix details several extensions and gathers all the proofs.

Related literature

In binary and staggered designs, our DID_ℓ estimators reduce to the estimators of the event-study parameters $\theta_{es}(\ell-1)$ of Callaway and Sant'Anna (2021), using not-yet-treated as controls. DID_1 also reduces to the DID_M estimator of de Chaisemartin and D'Haultfœuille (2020).

In non-binary and/or non-staggered designs, when all groups are untreated at period one, DID_ℓ is equivalent to the estimator obtained by:

1. redefining the treatment as an indicator equal to one if group g has ever received a non-zero treatment at t, as Deryugina (2017) or Krolikowski (2018), among others, have done in their empirical analysis;

2. computing the estimator of $\theta_{es}(\ell-1)$ of Callaway and Sant'Anna (2021) with this binarized

and staggerized treatment.

Then, our contribution is to propose three strategies to overcome DID_ℓ's lack of "structural" interpretation (treatment-trajectory-specific versions of DID_ℓ, normalized DID_ℓ estimators, and a linear combination of the DID_ℓs with an economic interpretation).

In non-binary and/or non-staggered designs, when groups do not all have the same period-one treatment, DID_ℓ is not equivalent to the estimator obtained by:

1. redefining the treatment as an indicator equal to one if group g's treatment has ever changed at t, as Callaway and Sant'Anna (2021) or East et al. (2023), among others, have done in their empirical analysis;

2. computing the estimator of $\theta_{es}(\ell-1)$ of Callaway and Sant'Anna (2021) with this binarized and staggerized treatment.

Those two strategies are not equivalent, because DID_ℓ only compares switchers and not-yet-switchers with the same period-one treatment, whereas the Callaway and Sant'Anna (2021) estimator applied to the binarized and staggerized treatment compares switchers and not-yet-switchers with different period-one treatments. Accordingly, we show that this second estimator relies on a stronger parallel-trends assumption than our DID_ℓ estimator: when combined with a parallel-trends assumption on groups' never-treated outcome, it rules out both effects of lagged treatments on the outcome and time-varying treatment effects.

In work posterior to ours, Callaway et al. (2021) consider designs where groups are all untreated at period one, and then get treated at heterogeneous dates, with heterogeneous intensities. Our estimators can also be used in those designs. Some of the estimands they propose compare the outcome evolution of treated and not-yet-treated, like ours.

An important literature has proposed estimators of the effects of current and lagged treatments, when the treatment 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). Instead, we consider a parallel-trends assumption, and consider alternative estimands and estimators under that assumption.

Finally, our decompositions of the regression coefficients used by applied researchers to estimate 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; Goodman-Bacon, 2021; Borusyak and Jaravel, 2017), and in particular to Sun and Abraham (2021) and de Chaisemartin and D'Haultfœuille (2023), who respectively study event-study regressions with a binary and staggered treatment, and regressions with several treatments. Our Propositions 1 and 2 below are novel results in this literature. The fact that TWFE local-projection regressions are misspecified and could exhibit sign reversal even under constant effects is unique to that particular class of TWFE regressions. On the other hand, our Proposition 3 is an application of Corollary 1 in de Chaisemartin and D'Haultfœuille (2023).

2 Setup, design, and identifying assumptions

2.1 Setup

Group-level panel data. We seek to estimate the effect of a treatment on an outcome. For that purpose, we use a panel of G groups observed at T periods, respectively indexed by g and t. Typically, groups are geographical entities, like states, counties or municipalities, but a group could also just be a single individual or firm. The group-level panel data may be constructed by aggregating an individual-level panel or repeated cross-section data set at the (g,t) level, defining groups, say, as individuals' county of birth. The group-level panel data may also be constructed from a single cross-section dataset, with cohort of birth playing the role of the time variable. The estimators we propose below are not weighted by $N_{g,t}$, the population of cell (g,t). This is just to reduce notational complexity: proposing weighted estimators is a mechanical extension.[4]

[4]The `did_multiplegt_dyn` command that implements the estimators proposed in this paper can be used with data at a more disaggregated level than the (g,t) level. Then, it aggregates the data at the (g,t) level internally and automatically weights (g,t) cells by their number of observations in the data. To use different weights, or to weight the estimation when working directly with a group-level panel data, one can use the `weight` option.

Treatment. When the treatment is assigned at the (g,t) level, let $D_{g,t}$ denote the treatment of group g at period t. When the treatment varies within (g,t) cells, a so-called fuzzy design (de Chaisemartin and D'Haultfœuille, 2018) that may arise when groups are geographical entities and individuals or firms within the same cell may not all have the same treatment, let $D_{g,t}$ denote the average treatment in cell (g,t). Throughout the paper, we assume that the treatment is non-negative: $D_{g,t} \geq 0$, as is most often the case in applications.[5] Let $\boldsymbol{D}_g = (D_{g,1}, ..., D_{g,T})$ be a vector stacking g's treatments 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. We refer to $\boldsymbol{D}$ as the study's design. Finally, let $\mathcal{D}$ be the set of values $\boldsymbol{D}_g$ can take (i.e.: its support), which is assumed to be the same for every g.

Potential outcomes. For all $(d_1, ..., d_T) \in \mathcal{D}$, let $Y_{g,t}(d_1, ..., d_T)$ denote the potential outcome of group g at t if $(D_{g,1}, ..., D_{g,T}) = (d_1, ..., d_T)$, and let $Y_{g,t} = Y_{g,t}(\boldsymbol{D}_g)$ denote the observed outcome of g at t. This dynamic potential outcome framework follows Robins (1986). It explicitly allows groups' outcome at t to depend on their past and future treatments.

Initial conditions. Some observations may have already been treated prior to period 1, and those treatments may still affect some of their period-1-to-T outcomes, the so-called initial-conditions problem. However, we cannot estimate those effects, as treatments and outcomes are not observed before period 1, so we do not account for this potential dependency in our notation. We discuss strategies to account for this dependency in Section 1.8 of the Web Appendix.

Fuzzy designs. In fuzzy designs, our potential outcome notation assumes that the outcome of g at t can only depend on the average treatments of units in g from period 1 to T, not on the identities of the treated units. We show in Web Appendix Section 1.7 that our estimators can still be used in fuzzy designs, even if potential outcomes depend on treated units' identities.

Conditioning on the design. When defining our target parameters, we will take the perspective of a social planner, seeking to conduct a cost-benefit analysis comparing groups' actual treatments $\boldsymbol{D}$ to the counterfactual "status-quo" scenario where every group would have kept the same treatment as in period 1 all the time. In other words, the planner wants to know if the

[5] When $D_{g,t} \geq \underline{d} < 0$, one can redefine the treatment as $D_{g,t} - \underline{d}$, a non-negative treatment.

treatment/policy changes that took place over the duration of the panel led to a better situation than the one that would have prevailed without any policy change, an arguably natural policy question. Then, all our analysis is conditional on $\boldsymbol{D}$, the study's design. This implies that our parameters of interest are dictated by the design, rather than chosen by the researcher. They thus share some commonalities with the local average treatment effect parameter proposed by Imbens and Angrist (1994). As such, they are very far from exhausting the set of all possible treatment effect parameters one may be interested in, and may only very partially uncover the structural function mapping the current, lagged, and future treatments to the average outcome. Conditional on the design, only groups' potential outcomes are random. We take a model-based approach to uncertainty: potential outcomes are random because of random shocks. Probabilistic statements below are with respect to potential outcomes' joint probability distribution.

2.2 Design

Condition that the design has to satisfy for our estimators to be applicable. The date at which a group's treatment changes for the first time plays a key role in our analysis.

Definition 1 *(First treatment change) For all g, let $F_g = \min\{t : t \geq 2, D_{g,t} \neq D_{g,t-1}\}$.*

We adopt the convention that $F_g = T + 1$ if g's treatment never changes. Our estimators are applicable to any design that satisfies Assumption 1 below.

Assumption 1 *(Restriction on the design) $\exists (g, g')$ such that: (i) $D_{g,1} = D_{g',1}$, (ii) $F_g \neq F_{g'}$.*

(i) requires that there exist groups with the same period-one treatment. This is essentially a restriction on the support of the period-one treatment. If groups' period-one treatments are i.i.d. draws from a continuous distribution, $D_{g,1} \neq D_{g',1}$ for all (g, g'), so (i) fails.[6] On the other hand, if $D_{g,1}$ can only take a finite number of values K, (i) automatically holds as long as $K < G$. (ii) requires that there is at least one set of groups sharing the same period-one treatment within which there is heterogeneity in the date at which groups change treatment for the first time. (ii) fails if groups' treatment is extremely non-persistent, so that $D_{g,1} \neq D_{g,2}$ and $F_g = 2$ for all g. (ii) also fails if all groups keep their period-one treatment throughout the panel: $F_g = T + 1$ for

[6] Estimators for this type of designs are proposed in de Chaisemartin et al. (2022).

all g, or if groups all change treatment for the first time at the same date, for instance due to a universal policy affecting them all: $F_g = t_0$ for all g. Overall, Assumption 1 mainly rules out designs with a continuous period-one treatment, designs where groups' treatment is extremely non-persistent, and designs with no or a universal treatment change. Assumption 1 is a minimal condition ensuring our estimators can be computed. For our estimators to be consistent and asymptotically normal, the number of pairs (g, g') meeting conditions (i) and (ii) essentially needs to go to infinity when $G \to +\infty$ (see Section 3.4).

Only untreated groups at period one? In Assumption 2 below, we single out designs where all groups are untreated at period one. Importantly, our estimators can be used when Assumption 2 fails, but designs where this assumption holds are frequent, and easier to analyze.[7]

Assumption 2 *(Zero treatment at baseline)* $\forall\ g,\ D_{g,1} = 0$.

Four commonly-found designs. We now define four commonly-found designs where Assumption 1 often holds, illustrating them with examples from our census of highly-cited TWFE papers published in the AER. The first three designs meet Assumption 2, the last one does not. Our estimators can be used in those four designs, provided Assumption 1 holds.

Design 1 *(Binary and staggered treatment)* $\forall\ (g,t),\ D_{g,t} = 1\{t \geq F_g\}$, *with* $F_g \geq 2$.

Design 1 has been studied by Callaway and Sant'Anna (2021), Goodman-Bacon (2021), Sun and Abraham (2021), and Borusyak et al. (2021), among others. Assuming that all groups are initially untreated ($F_g \geq 2$) is without loss of generality: in Design 1, the treatment effects of always-treated groups cannot be estimated under parallel-trends assumptions.

Design 2 *(Binary treatment, groups can join and then leave treatment)* $\forall\ (g,t),\ D_{g,t} = 1\{E_g \geq t \geq F_g\}$, *with* $2 \leq F_g \leq E_g$.

In Design 2, groups may get treated and leave treatment. For instance, the design of Burgess et al. (2015) is very close to Design 2.[8] An important special case of Design 2 is when treated

[7] When some but not all groups are initially untreated, Assumption 2 holds among initially untreated groups.

[8] Specifically, for the co-ethnicity × democracy treatment in their Equation (2), districts can enter and leave treatment at most twice over their entire study period.

groups are treated for only one period ($E_g = F_g$), leading to

$$D_{g,t} = 1\{t = F_g\}, \tag{1}$$

which we hereafter refer to as a "one-shot-treatment design". There are many applications where it is unclear whether the treatment should be defined as binary-and-staggered or one-shot. In Section 3.3, we propose a cost-benefit framework to help practitioners make that decision.

Design 3 *(Staggered design with group-specific intensities)* $\forall\ (g,t)$, $D_{g,t} = I_g 1\{t \geq F_g\}$, $F_g \geq 2$.

In Design 3, there is variation across groups in treatment timing and/or intensity. For instance, the design of Pierce and Schott (2016) corresponds to Design 3.

Design 4 *(Discrete treatment at baseline)* $\forall\ g$, $D_{g,1} \in \{0, 1, ..., K\}$.

In Design 4, groups' baseline treatment is discrete, and treatment paths are unrestricted. For instance, the design of Fuest et al. (2018) corresponds to Design 4. A special case of Design 4 is when groups' baseline treatment is binary.

2.3 Identifying assumptions

No-anticipation. We maintain throughout the paper the following condition.

Assumption 3 *(No Anticipation)* $\forall g$, $\forall (d_1, ..., d_T) \in \mathcal{D}$, $Y_{g,t}(d_1, ..., d_T) = Y_{g,t}(d_1, ..., d_t)$.

Assumption 3 requires that a group's current outcome does not depend on its future treatments, the so-called no-anticipation hypothesis. This assumption appears in Robins (1986), who introduced the dynamic potential outcomes we use in this paper. Abbring and Van den Berg (2003) have used it in duration models, and Malani and Reif (2015), Botosaru and Gutierrez (2018), Callaway and Sant'Anna (2021), and Sun and Abraham (2021) have used it in DID models.

Parallel trends. Let $\mathcal{D}_1^{\text{r}} = \{d : \exists (g, g') \in \{1, ..., G\}^2 : D_{g,1} = D_{g',1} = d, F_g \neq F_{g'}\}$ be the set of period-one-treatment values such that two groups with different values of F_g have that period-one treatment. For any d in $\mathcal{D}_1^{\text{r}}$ and any t, let $\boldsymbol{d}_t$ denote a $1 \times t$ vector of ds. For any integer k, let $\boldsymbol{D}_{g,1,k}$ be a $1 \times k$ vector whose coordinates are all equal to $D_{g,1}$. $Y_{g,t}(\boldsymbol{D}_{g,1,t})$ is g's

period-t outcome in a counterfactual where it keeps its period-one treatment till period t. We refer to it as its "status-quo" potential outcome.

Assumption 4 *(Parallel trends for the status-quo outcome, conditional on the period-one treatment)* $\forall(g,g')$*, if* $D_{g,1}=D_{g',1}\in\mathcal{D}_1^r$*, then* $\forall t\geq 2$*,*

$$E[Y_{g,t}(\boldsymbol{D}_{g,1,t})-Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}]=E[Y_{g',t}(\boldsymbol{D}_{g',1,t})-Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}].$$

Assumption 4 requires that if two groups have the same period-one treatment, then they have the same expected evolution of their status-quo outcome. Assumption 4 is a generalization of the standard parallel-trends assumption in DID models (see, e.g., Abadie, 2005) to our set-up that allows for dynamic effects and for potentially complicated designs where groups may not all be untreated at period one. Under Assumptions 1-2, $\mathcal{D}_1^{\text{r}}=\{0\}$, so Assumption 4 only restricts the never-treated potential outcome $Y_{g,t}(\mathbf{0}_t)$, and is similar to the identifying assumption considered by Callaway and Sant'Anna (2021) and Sun and Abraham (2021) for binary and staggered designs. Note that Assumption 4 restricts only one potential outcome per group, so even when $\mathcal{D}_1^{\text{r}}$ contains more than one value, Assumption 4 alone does not place any restriction on groups' treatment effects. There may be instances where due to the way selection into treatment takes place, Assumption 4 is implausible for some values in $\mathcal{D}_1^{\text{r}}$. In such instances, the estimators we propose can still be used, restricting the sample to groups with a period-one-treatment-value for which Assumption 4 is plausible. Assumption 4 has testable implications that are easy to assess, using placebo estimators mimicking the actual estimators we propose (see Section 3.5).

3 Target parameters and estimators

3.1 Non-normalized actual-versus-status-quo and event-study effects

Definition of the non-normalized actual-versus-status-quo effects. 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 period-one treatment as g 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 $\ell \in \{1, ..., T_g - F_g + 1\}$, let

$$\delta_{g,\ell} = E\left[Y_{g,F_g-1+\ell} - Y_{g,F_g-1+\ell}(D_{g,1}, ..., D_{g,1}) | \boldsymbol{D}\right] \tag{2}$$

be the expected difference between group g's actual outcome at $F_g - 1 + \ell$ and its counterfactual "status quo" outcome if its treatment had remained equal to its period-one value from period one to $F_g - 1 + \ell$. We refer to $\delta_{g,\ell}$ as the actual-versus-status-quo (AVSQ) effect of g at $F_g - 1 + \ell$.

Estimation of the non-normalized actual-versus-status-quo effects. For any finite set A, let $\#A$ be the number of elements of A. For all (g, t), let

$$N_t^g = \#\{g' : D_{g',1} = D_{g,1}, F_{g'} > t\}$$

be the number of groups g' with the same period-one treatment as g, and that have kept the same treatment from period 1 to t. For every g such that $F_g \leq T_g$, and every $\ell \in \{1, ..., T_g - F_g + 1\}$, $N^g_{F_g-1+\ell} > 0$. To estimate $\delta_{g,\ell}$, we use

$$\text{DID}_{g,\ell} = Y_{g,F_g-1+\ell} - Y_{g,F_g-1} - \frac{1}{N^g_{F_g-1+\ell}} \sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell} (Y_{g',F_g-1+\ell} - Y_{g',F_g-1}), \tag{3}$$

a DID estimator comparing the $F_g - 1$-to-$F_g - 1 + \ell$ outcome evolution of g to that of groups with the same baseline treatment, and that have kept that treatment from period 1 to $F_g - 1 + \ell$.[9]

Lemma 1 *If Assumptions 3 and 4 hold, then for every (g, ℓ) such that $1 \leq \ell \leq T_g - F_g + 1$, $E\left[DID_{g,\ell} | \boldsymbol{D}\right] = \delta_{g,\ell}$.*

No-crossing condition. Assume that g_0 is such that $D_{g_0,1} = 1, D_{g_0,2} = 2, D_{g_0,3} = 0$. Then,

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

[9] $\text{DID}_{g,\ell}$ uses groups' $F_g - 1$ outcome as the baseline. Instead, one could average their period-1-to-$F_g - 1$ outcomes. This would give rise to another unbiased estimator of $\delta_{g,\ell}$. Which estimator is more precise depends on the outcome's serial correlation (see Borusyak et al., 2021; Harmon, 2022). Moreover, this second estimator could be more biased than $\text{DID}_{g,\ell}$ if Assumption 4 does not exactly hold and the discrepancy between groups' trends gets larger over longer horizons. See de Chaisemartin and D'Haultfœuille (2022) for further discussion.

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 $\delta_{g_0,2}$ is negative. Therefore, $\delta_{g_0,2}$ does not satisfy the no-sign reversal property: $(d_1, d_2, d_3) \mapsto Y_{g,t}(d_1, d_2, d_3)$ may be increasing in each of its arguments but $\delta_{g_0,2}$ is negative. Beyond this example, 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 different treatment lags. Throughout this section, we assume away the existence of such (g, ℓ)s.

Assumption 5 *(No-crossing condition)* $\forall g \in \{1, ..., G\}$, *either* $D_{g,t} \geq D_{g,1}$ *for every* t, *or* $D_{g,t} \leq D_{g,1}$ *for every* t.

Assumption 5 is implied by Assumption 2. It also holds automatically when the treatment is binary, or when groups' treatment can only change once. When Assumption 5 fails, one can just discard 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.[10] This yields an unbalanced panel of groups where Assumption 5 holds by construction, on which our estimators can be applied. The only reason we impose Assumption 5 is to avoid the notational burden of defining our estimators on an unbalanced panel of groups.[11]

Definition of the non-normalized event-study effects. Let $L = \max_g(T_g - F_g + 1)$ denote the largest ℓ such that $\delta_{g,\ell}$ can be estimated for at least one g. Under Assumption 1, $L \geq 1$. For every $\ell \in \{1, ..., L\}$, let

$$N_\ell = \#\{g : F_g - 1 + \ell \leq T_g\}$$

be the number of groups for which $\delta_{g,\ell}$ can be estimated. For all g such that $F_g \leq T$, let

$$S_g = 1\{D_{g,F_g} > D_{g,1}\} - 1\{D_{g,F_g} < D_{g,1}\}$$

be equal to 1 (resp. -1) for groups whose treatment increases (resp. decreases) at F_g. Then, let

$$\delta_\ell = \frac{1}{N_\ell} \sum_{g: F_g - 1 + \ell \leq T_g} S_g \delta_{g,\ell}, \tag{4}$$

[10] `did_multiplegt_dyn` automatically drops those cells if the `drop_larger_lower` option is specified.

[11] `did_multiplegt_dyn` can be used with an unbalanced panel of groups, see the help file.

be the average of $S_g\delta_{g,\ell}$. Under Assumption 5, for groups with $S_g = 1$, $D_{g,t} \geq D_{g,1}$ for all t, so $\delta_{g,\ell}$ is the effect of having been exposed to a weakly higher treatment dose for ℓ periods. Conversely, for groups with $S_g = -1$, $D_{g,t} \leq D_{g,1}$ for all t, so $\delta_{g,\ell}$ is the effect of having been exposed to a weakly lower dose for ℓ periods. Taking the negative of $\delta_{g,\ell}$ for those groups ensures that δ_ℓ is an average effect of having been exposed to a weakly larger dose for ℓ periods.

Estimation of the non-normalized event-study effects. For every $\ell \in \{1, ..., L\}$, let

$$\text{DID}_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} S_g \text{DID}_{g,\ell}. \tag{5}$$

Lemma 1 implies that DID_ℓ is conditionally unbiased for δ_ℓ under Assumptions 1, 3, and 4.

Interpreting non-normalized event-study effects. In Design 1,

$$\delta_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} E\left[Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_\ell) - Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1+\ell}) | \boldsymbol{D}\right].$$

Thus, δ_ℓ is just the average effect of having been treated rather than untreated for ℓ periods, across all groups reaching ℓ treatment periods at or before the last period where there is still at least one untreated group. Outside of Design 1, the interpretation of δ_ℓ is more complicated. For instance, even in Design 2 ($D_{g,t} = 1\{E_g \geq t \geq F_g\}$), for all g such that $F_g - 1 + \ell > E_g$, group g has exited the treatment at $F_g - 1 + \ell$, and

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

is the effect of having been treated for $E_g - F_g + 1$ periods, $F_g - 1 + \ell - E_g$ periods before the outcome is measured. Thus, the number and the recency of the treatment periods that generate $\delta_{g,\ell}$ vary across groups, complicating the interpretation of δ_ℓ. Similarly, with three periods and three groups such that $(D_{1,1} = 0, D_{1,2} = 4, D_{1,3} = 0)$, $(D_{2,1} = 0, D_{2,2} = 2, D_{2,3} = 3)$, and $(D_{2,1} = 0, D_{2,2} = 0, D_{2,3} = 0)$, DID_2 estimates the average of $E(Y_{1,3}(0, 4, 0) - Y_{1,3}(0, 0, 0))$ and $E(Y_{2,3}(0, 2, 3) - Y_{2,3}(0, 0, 0))$. Thus, the magnitude and timing of the treatment increments generating $\delta_{g,\ell}$ varies across groups, which again complicates the interpretation of δ_ℓ. Still, in any design satisfying Assumption 5, δ_ℓ can be interpreted as an average effect of having been exposed to a weakly higher treatment for ℓ periods.

More disaggregated and more interpretable effects? If the design is such that $P_\ell := \#\{(D_{g,1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell}) : F_g - 1 + \ell \leq T_g\}$ is very low relative to G, the number of period-1-to-$F_g - 1 + \ell$ treatment paths among the groups entering in δ_ℓ may be sufficiently low to precisely estimate the average of $\delta_{g,\ell}$ separately across all groups with the same path, thus yielding estimates of the average effects of specific treatment paths.[12] For instance, in Design 2 the number of treatment paths may often be low enough for this solution to be practical. But in more complicated designs where treatment paths are very group-specific, P_ℓ may often be too large for this solution to be practical, especially as ℓ increases. In such instances, we still recommend that practitioners report the values in $\{(D_{g,1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell}) : F_g - 1 + \ell \leq T_g\}$, and their distribution: this information may be helpful to interpret DID_ℓ.

Comparison with existing estimators, in binary and staggered designs. In Design 1, δ_ℓ reduces to the parameter $\theta_{es}(\ell - 1)$ in Callaway and Sant'Anna (2021), and DID_ℓ reduces to their estimator of $\theta_{es}(\ell - 1)$ using not-yet-treated as the control group and without covariates. DID_1 also reduces to the DID_M estimator in de Chaisemartin and D'Haultfœuille (2020).

Comparison with existing estimators, in non-binary and/or non-staggered designs. When groups are all initially untreated, as in Assumption 2, DID_ℓ is numerically equivalent to the estimator obtained by redefining the treatment as an indicator equal to one if group g has ever received a non-zero treatment at t, and computing the estimators of $\theta_{es}(\ell - 1)$ of Callaway and Sant'Anna (2021) with this binarized and staggered treatment. When Assumption 2 fails, redefining the treatment as an indicator equal to one if group g's treatment has ever changed at t and computing the estimators of $\theta_{es}(\ell - 1)$ of Callaway and Sant'Anna (2021) with this binarized and staggerized treatment, as was for instance done in Callaway and Sant'Anna (2021) and East et al. (2023), does not yield estimators numerically equivalent to our DID_ℓ estimators. This is because our estimators only compare switchers and non-switchers with the same period-one treatment, whereas the Callaway and Sant'Anna (2021) estimator applied to the binarized and staggerized treatment compares switchers and non-switchers with different

[12] For any treatment path $(D_{g,1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell}) = \boldsymbol{d}$, a treatment-path-specific version of δ_ℓ can easily be estimated using the `did_multiplegt_dyn` command, restricting the sample to (g,t)s such that $t < F_g$, or $F_g \leq t \leq F_g - 1 + \ell$ and $(D_{g,1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell}) = \boldsymbol{d}$.

period-one treatments. Accordingly, the latter estimators rely on Assumption 6 below.

Assumption 6 *(Unconditional parallel trends for the status-quo outcome)* $\forall(g,g')$, $\forall t \geq 2$,

$$E[Y_{g,t}(\boldsymbol{D}_{g,1,t}) - Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}] = E[Y_{g',t}(\boldsymbol{D}_{g',1,t}) - Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}].$$

Assumption 6 is stronger than Assumption 4: the former imposes that all groups experience the same evolution of their status-quo outcome, whereas the latter only imposes that groups with the same period-one treatment experience the same evolution of their status-quo outcome. To simplify the remainder of the discussion, let us momentarily assume that the treatment is binary. Then, when combined with the standard parallel-trends assumption on groups' never-treated outcome, Assumption 6 implies that for all groups such that $D_{g,1} = 1$,

$$E\left[Y_{g,t}(\mathbf{1}_t) - Y_{g,t-1}(\mathbf{1}_{t-1})|\boldsymbol{D}\right] = E\left[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1})|\boldsymbol{D}\right], \tag{6}$$

which is equivalent to

$$E\left[Y_{g,t}(\mathbf{1}_t) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right] = E\left[Y_{g,t-1}(\mathbf{1}_{t-1}) - Y_{g,t-1}(\mathbf{0}_{t-1})|\boldsymbol{D}\right]. \tag{7}$$

In initially treated groups, the effect of being treated for t periods should be the same as the effect of being treated for $t-1$ periods. (7) may hold if lagged treatments do not affect the current outcome and the effect of the current treatment does not change over time, but it is likely to fail whenever lagged treatments affect the outcome and/or treatment effects are time-varying. By contrast, when combined with the standard parallel-trends assumption on groups' never-treated outcome, Assumption 4 implies that for all groups such that $D_{g,1} = 1$,

$$E(Y_{g,t}(\mathbf{1}_t) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}) - (E(Y_{g,t-1}(\mathbf{1}_{t-1}) - Y_{g,t-1}(\mathbf{0}_{t-1})|\boldsymbol{D})) \text{ does not vary across } g. \tag{8}$$

Thus, the incremental effect of one treatment period should be the same in every initially treated group. This is of course a strong restriction, but unlike (7), it does not rule out effects of lagged treatments on the outcome, and it allows for time-varying effects. It also allows $E[Y_{g,t}(\mathbf{1}_t) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}]$ to vary across groups, because it does not place any restriction on $E[Y_{g,1}(1) - Y_{g,1}(0)|\boldsymbol{D}]$. de Chaisemartin and D'Haultfœuille (2018) and de Chaisemartin and D'Haultfœuille (2020) had already noted that in a model assuming away effects of lagged treatments on the outcome, parallel trends across groups with different baseline treatments essentially rules out time-varying effects. In models allowing for effects of lagged treatments, that assumption imposes even more implausible restrictions.

3.2 Normalized actual-versus-status-quo and event-study effects

Definition and estimation of the normalized actual-versus-status-quo effects. For any g such that $F_g \leq T_g$ and any $\ell \in \{1, ..., T_g - F_g + 1\}$, let

$$\delta^D_{g,\ell} = \sum_{k=0}^{\ell-1} (D_{g,F_g+k} - D_{g,1}) \tag{9}$$

be the difference between the total treatment dose received by group g from F_g to $F_g - 1 + \ell$, and the total treatment dose it would have received in the status-quo counterfactual. In binary and staggered designs, $\delta^D_{g,\ell} = \ell$ for all g. But in more complicated designs, $\delta^D_{g,\ell}$ may vary across g. For instance, in Design 2 ($D_{g,t} = 1\{E_g \geq t \geq F_g\}$), $\delta^D_{g,\ell} = \min(\ell, E_g - F_g + 1)$. Then, let

$$\delta^n_{g,\ell} = \frac{\delta_{g,\ell}}{\delta^D_{g,\ell}}. \tag{10}$$

We refer to $\delta^n_{g,\ell}$ as the normalized AVSQ effect of group g at $F_g - 1 + \ell$. It follows directly from Lemma 1 that $\text{DID}_{g,\ell}/\delta^D_{g,\ell}$ is conditionally unbiased for $\delta^n_{g,\ell}$.

$\delta^n_{g,\ell}$ is a weighted average of the effects the current and $\ell - 1$ first treatment lags. We adopt the convention that for $k = 0$ $\boldsymbol{D}_{g,1,k}$ stands for the empty vector, and accordingly when $k = \ell - 1$, $(D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1})$ also stands for the empty vector. We sometimes refer to a cell's current treatment as its 0th treatment lag, and use the convention 0/0=0. For $k \in \{0, ..., \ell - 1\}$, let

$$\begin{aligned} s_{g,\ell,k} =& E\Big[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, \underline{D_{g,F_g-1+\ell-k}}, \boldsymbol{D}_{g,1,k}) \\ & - Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, \underline{D_{g,1}}, \boldsymbol{D}_{g,1,k}) | \boldsymbol{D}\Big] / (D_{g,F_g-1+\ell-k} - D_{g,1}) \end{aligned}$$

be the slope of the expected potential outcome function of group g at $F_g - 1 + \ell$ with respect to its kth treatment lag (the underlined term), when that lag is switched from its status-quo counterfactual value $D_{g,1}$ to its actual value $D_{g,F_g-1+\ell-k}$, whereas all its previous treatments are held at their actual values, and all its subsequent treatments are held at their status-quo value. For any $k \in \{0, ..., \ell - 1\}$, let

$$w_{g,\ell,k} = \frac{D_{g,F_g-1+\ell-k} - D_{g,1}}{\delta^D_{g,\ell}}.$$

Under Assumption 5, $w_{g,\ell,k} \geq 0$ for all (g, ℓ, k). Moreover, $\sum_{k=0}^{\ell-1} w_{g,\ell,k} = 1$ for all (g, ℓ).

Lemma 2 *For every* (g,ℓ) *such that* $1 \leq \ell \leq T_g - F_g + 1$, $\delta^n_{g,\ell} = \sum_{k=0}^{\ell-1} w_{g,\ell,k} s_{g,\ell,k}$.

Lemma 2 shows that $\delta^n_{g,\ell}$ is a weighted average of the slopes of g's potential outcome at $F_g - 1 + \ell$ with respect to its $\ell - 1$ first treatment lags, where for $k \in \{0, ..., \ell - 1\}$, the effect of the kth lag receives a weight proportional to the absolute value of the difference between g's kth treatment lag and its status-quo treatment. While $\delta^n_{g,\ell}$ can be unbiasedly estimated under Assumption 3-4, in general one cannot estimate the slopes $s_{g,\ell,k}$ under those assumptions. Doing so would require unbiasedly estimating the counterfactual outcomes

$$Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, D_{g,1}, \boldsymbol{D}_{g,1,k})$$

for every $k \in \{0, ..., \ell - 1\}$. But Assumptions 3-4 only allow us to estimate the status-quo counterfactual outcome, which corresponds to the outcome in the previous display for $k = \ell - 1$.

Lemma 2 in some specific designs. For concreteness, we rewrite the result in Lemma 2 in three specific designs. First, in binary and staggered designs, Lemma 2 reduces to

$$\delta^n_{g,\ell} = \frac{1}{\ell} \sum_{k=0}^{\ell-1} E\left[Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_{\ell-k-1}, 1, \mathbf{0}_k) - Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, \mathbf{1}_{\ell-k-1}, 0, \mathbf{0}_k) | \boldsymbol{D}\right].$$

Then, $\delta^n_{g,\ell}$ is the average, across k ranging from 0 to $\ell - 1$, of the effect of switching the kth treatment lag from 0 to 1, holding previous treatments at 1 and subsequent treatments at 0. Second, in the one-shot-treatment designs defined in (1), Lemma 2 reduces to

$$\delta^n_{g,\ell} = E\left[Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, 1, \mathbf{0}_{\ell-1}) - Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, 0, \mathbf{0}_{\ell-1}) | \boldsymbol{D}\right], \tag{11}$$

so $\delta^n_{g,\ell}$ is just the effect of switching the $\ell - 1$th treatment lag from 0 to 1, holding all other treatments at 0. Third, in Design 3 ($D_{g,t} = I_g 1\{t \geq F_g\}$), Lemma 2 reduces to

$$\delta^n_{g,\ell} = \frac{1}{\ell} \sum_{k=0}^{\ell-1} \frac{E\left[Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, \boldsymbol{I}_{g,\ell-k-1}, I_g, \mathbf{0}_k) - Y_{g,F_g-1+\ell}(\mathbf{0}_{F_g-1}, \boldsymbol{I}_{g,\ell-k-1}, 0, \mathbf{0}_k) | \boldsymbol{D}\right]}{I_g}. \tag{12}$$

Thus, in staggered designs with group-specific treatment intensities, $\delta^n_{g,\ell}$ is the average, across k ranging from 0 to $\ell - 1$, of the effect of switching the kth lag from 0 to I_g, normalized by I_g.

Definition and estimation of the normalized event-study effects. Let $\delta^D_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} |\delta^D_{g,\ell}|$. For $\ell \in \{1, ..., L\}$, let

$$\delta^n_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|\delta^D_{g,\ell}|}{\delta^D_\ell} \delta^n_{g,\ell}. \tag{13}$$

Note the following relation between the non-normalized and normalized event-study effects:

$$\delta_\ell^n = \frac{\delta_\ell}{\delta_\ell^D}. \tag{14}$$

δ_ℓ^n is a weighted average of all the $\delta_{g,\ell}^n$ parameters that can be estimated, with weights proportional to $|\delta_{g,\ell}^D|$. In designs where some gs are such that $|\delta_{g,\ell}^D|$ is close to zero, the estimator of the unweighted average of the $\delta_{g,\ell}^n$s may be very noisy. This is what leads us to consider the weighted average above. It follows directly from Lemma 1 that

$$\text{DID}_\ell^n := \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|\delta_{g,\ell}^D|}{\delta_\ell^D} \frac{\text{DID}_{g,\ell}}{\delta_{g,\ell}^D} \tag{15}$$

is conditionally unbiased for δ_ℓ^n. Similarly to (14), $\text{DID}_\ell^n = \text{DID}_\ell / \delta_\ell^D$.

Interpretation of δ_ℓ^n. It follows from Lemma 2 that δ_ℓ^n is a weighted average of the effects of groups' current and $\ell - 1$ first treatment lags on their outcome. The total weight assigned by δ_ℓ^n to the effect of the kth-lag (for $0 \leq k \leq \ell - 1$) is equal to

$$w_{\ell,k} = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|D_{g,F_g-1+\ell-k} - D_{g,1}|}{\delta_\ell^D}.$$

In the one-shot treatment designs defined in (1), $w_{\ell,\ell-1} = 1$, so δ_ℓ^n only measures the effect of the $\ell - 1$th treatment lag. In more complicated designs, δ_ℓ^n averages together the effect of different treatment lags. When groups' treatment can only change once, $w_{\ell,k} = 1/\ell$, so $\ell \mapsto w_{\ell,k}$ is decreasing for $\ell \geq k + 1$: δ_ℓ^n assigns less weight to the effect of recent treatments when ℓ increases. Finally, one always has $w_{1,0} = 1$ and $w_{\ell,0} \leq 1$ for $\ell \geq 2$: δ_1^n only averages effects of groups' current treatment, whereas δ_ℓ^n also averages effects of treatment lags for $\ell \geq 2$. We recommend reporting $k \mapsto w_{\ell,k}$, to document which lags contribute the most to δ_ℓ^n.[13]

Using normalized event-study effects to test the null that the current and lagged treatments have the same effects, and that those effects do not change over time. For every $\ell' \in \{1, ..., L\}$ and every $\ell \in \{1, ..., \ell'\}$, let

$$\delta_{\ell,\ell'}^{n,bal} = \frac{1}{N_{\ell'}} \sum_{g:F_g-1+\ell' \leq T_g} \frac{|\delta_{g,\ell}^D|}{\frac{1}{N_{\ell'}} \sum_{g':F_{g'}-1+\ell' \leq T_{g'}} |\delta_{g',\ell}^D|} \delta_{g,\ell}^n \tag{16}$$

[13] One may use $\ell \mapsto \delta_{g,\ell}^n$ to separately estimate the effect of each treatment lag, if one assumes that lags' effects are additively separable, linear, and constant over time, as we had shown in a previous version of this paper (see Section 4 of de Chaisemartin and D'Haultfœuille, 2020).

denote a version of δ^n_ℓ, defined on the subsample of groups such that $g : F_g - 1 + \ell' \leq T_g$. Thus, for every $\ell \in \{1, ..., \ell'\}$, $\delta^{n,bal}_{\ell,\ell'}$ applies to the same set of groups.

Lemma 3 *Suppose that groups' treatment only changes once: $D_{g,t} = D_{g,F_g}$ for all $t \geq F_g$. Suppose also that for all $\ell \in \{1, ..., \ell'\}$ and all $k \in \{0, \ldots, \ell - 1\}$, $s_{g,\ell,k} = s_{g,k}$ for some $s_{g,k}$. Then:*

1. *If $s_{g,0} = \cdots = s_{g,\ell-1}$, $\delta^{n,bal}_{1,\ell'} = \cdots = \delta^{n,bal}_{\ell',\ell'}$.*
2. *If $s_{g,0} \leq \cdots \leq s_{g,\ell-1}$, $\delta^{n,bal}_{1,\ell'} \leq \cdots \leq \delta^{n,bal}_{\ell',\ell'}$.*

Point 1 of Lemma 3 assumes that $s_{g,\ell,k} = s_{g,k} = s_g$, meaning that the slope of the expected potential outcome function of group g at $F_g - 1 + \ell$ with respect to its kth treatment lag does not depend on ℓ or k: the current and lagged treatments have the same effects on the outcome, and those effects do not change over time. Testing whether the current and lagged treatments have the same effect on the outcome is often of great economic interest, for instance to assess if treatment effects tend to disappear as time since exposure increases. Point 1 of Lemma 3 provides a joint test of this null hypothesis, and of the null hypothesis that treatment effects do not change over time: together, these two hypotheses imply that $\ell \mapsto \delta^{n,bal}_{\ell,\ell'}$ is constant. Similarly, Point 2 of Lemma 3 provides a joint test of the null hypothesis that the effect of lagged treatments decreases with the number of lags, and of the null hypothesis that treatment effects do not change over time: together, these two hypotheses imply that $\ell \mapsto \delta^{n,bal}_{\ell,\ell'}$ is decreasing. We can test both testable implications using estimators $\text{DID}^{n,bal}_{\ell',\ell'}$ of $\delta^{n,bal}_{\ell,\ell'}$, through a joint equality test for the first testable implication, and a multivariate inequality test for the second one (see, e.g., Cox and Shi, 2023, for a recent and simple proposal). Lemma 3 assumes that groups' treatment changes only once. A similar result holds in more general designs, if $\ell \mapsto |D_{g,F_g-1+\ell} - D_{g,1}|$ is decreasing in ℓ for $\ell \geq 1$, meaning that groups' treatment can change multiple times, but tends to revert to its baseline value after the initial treatment change, as is for instance the case in Design 2. Then, the testable implications, instead of involving "balanced" versions of δ^n_ℓ, involve "balanced" versions of unweighted averages of the $\delta^n_{g,\ell}$ parameters.

3.3 Cost-benefit analysis

Assumption on the design. In this section, we strengthen Assumption 5.

Assumption 7 *(Lowest treatment at period one)* $\forall g \in \{1, ..., G\}$, $D_{g,t} \geq D_{g,1}$ *for every* t.

We impose Assumption 7 to reduce the notational burden. When it fails, one can just conduct the cost-benefit analysis separately for groups with $S_g = 1$ and for groups with $S_g = -1$, as was done in a previous version of this paper (see de Chaisemartin and D'Haultfœuille, 2020).

Counterfactuals compared in the cost-benefit analysis. We take the perspective of a planner, seeking to conduct a cost-benefit analysis comparing groups' actual treatments $\boldsymbol{D}$ to the counterfactual "status-quo" scenario where they would have kept throughout their period-1 treatment. 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.

Treatment benefits and costs. We assume that the outcome can be converted into a monetary equivalent, or at least that treatment effects per USD spent can be used to determine if the treatment is worthwhile. For instance, test scores may not be readily converted into a monetary equivalent, but a test score increase, in percent of a standard deviation, per USD spent, can be compared to the same metric for other treatments. Then, $\delta_{g,\ell}$ is the expected benefit or loss, in group g and at period $F_g - 1 + \ell$, 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's cost is linear in the dose administered, and we let $c_{g,\ell} \geq 0$ denote the cost of administering one treatment unit in group g at period $F_g - 1 + \ell$. 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. The treatment cost may have to be estimated, by comparing the actual costs of the organization delivering the treatment to its counterfactual costs had it not delivered the treatment. In that case, c defined below can be estimated under a parallel-trends assumption on costs, using the estimator we propose for

the normalized treatment benefit, but defining the outcome as the costs of the organization delivering the treatment in group g at period t.

Cost-benefit analysis. Assuming that the planner's discount factor is equal to 1, groups' actual treatments are beneficial in monetary terms relative to the status quo, up to period T_g, if and only if

$$\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}\delta_{g,\ell}-\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}c_{g,\ell}(D_{g,F_g-1+\ell}-D_{g,1})>0. \tag{17}$$

Let

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

denote the average treatment cost, across all the incremental treatment doses received with respect to the status-quo counterfactual. Assumptions 1 and 7 ensure that c's denominator is positive. Dividing the left- and right-hand side of (17) by that denominator, we get that the actual treatments are beneficial if and only if

$$\delta:=\frac{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}\delta_{g,\ell}}{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}(D_{g,F_g-1+\ell}-D_{g,1})}>c.$$

Estimating δ. It follows directly from Lemma 1 that

$$\widehat{\delta}:=\frac{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}\text{DID}_{g,\ell}}{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}(D_{g,F_g-1+\ell}-D_{g,1})}$$

is conditionally unbiased for δ.

Formulas for δ in some specific designs. In binary and staggered designs, one has

$$\delta=E\left[\frac{1}{\sum_{(g,t)}D_{g,t}}\sum_{(g,t):D_{g,t}=1}\left(Y_{g,t}\left(\mathbf{0}_{F_g-1},\mathbf{1}_{t-F_g+1}\right)-Y_{g,t}\left(\mathbf{0}_t\right)\right)\middle|\boldsymbol{D}\right], \tag{18}$$

if there is at least one never-treated group. Then, δ is the average effect of having been treated for $t-F_g+1$ periods across all treated (g,t) cells, a parameter that generalizes the average treatment effect on the treated to our setting with dynamic effects. In one-shot-treatment designs, defined in (1), if at least one group never gets treated,

$$\delta=E\left[\frac{1}{\#\{g:F_g\leq T\}}\sum_{g:F_g\leq T}\sum_{\ell=1}^{T-(F_g-1)}\left(Y_{g,F_g-1+\ell}\left(\mathbf{0}_{F_g-1},1,\mathbf{0}_{\ell-1}\right)-Y_{g,F_g-1+\ell}\left(\mathbf{0}_{F_g-1+\ell}\right)\right)\middle|\boldsymbol{D}\right], \tag{19}$$

the average, across all groups that ever get treated, of the total effect of having been treated at period F_g on groups' outcomes from period F_g to T.

One-shot or binary-and-staggered treatment? Assume one is interested in the effect of being hit by a hurricane. To simplify, assume that locations can only be hit once over the study period (the discussion below still applies when that is not the case). Then, one may define the treatment as being hit by a hurricane at t, in which case treatment is a one-shot treatment, as in (1). Alternatively, one may define the treatment as having ever been hit by a hurricane at t, in which case the treatment is binary and staggered. In our cost-benefit framework, the treatment should only differ from 0 if (g,t) receives a treatment that is either costly to administer, or costly to prevent when the treatment is undesirable. Under this logic, one should thus define the treatment as being one-shot. With that definition, (19) shows that δ averages the total damages caused by hurricanes, at the time when they happen but also at later periods. δ may then be compared, say, to the cost of reducing CO2 emissions by a sufficient amount to lower the expected number of hurricanes by one. As another example, assume one is interested in a teacher training, that only generates costs in the year of implementation, but may affect students for many years thereafter. Then, in our cost-benefit framework, $D_{g,t}$ should be equal to 1 the year when the teachers of location g get treated, and to 0 otherwise.

Interpretation of δ as an average total effect per unit of treatment. Let us first consider a simple example with two groups and four periods, such that $D_{1,1}=0$, $D_{1,2}=1$, $D_{1,3}=1$, and $D_{1,4}=0$, while group 2 is never treated. Then,

$$\begin{aligned}\delta &= \frac{E\left[Y_{1,2}(0,1)-Y_{1,2}(0,0)+Y_{1,3}(0,1,1)-Y_{1,3}(0,0,0)+Y_{1,4}(0,1,1,0)-Y_{1,4}(0,0,0,0)|\boldsymbol{D}\right]}{1+1+0}\\ &= \frac{1}{2}E\left[Y_{1,2}(0,1)-Y_{1,2}(0,0)+Y_{1,3}(0,1,0)-Y_{1,3}(0,0,0)+Y_{1,4}(0,1,0,0)-Y_{1,4}(0,0,0,0)|\boldsymbol{D}\right]\\ &+ \frac{1}{2}E\left[Y_{1,3}(0,1,1)-Y_{1,3}(0,1,0)+Y_{1,4}(0,1,1,0)-Y_{1,4}(0,1,0,0)|\boldsymbol{D}\right]. \qquad (20)\end{aligned}$$

The first expectation in (20) 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 (20) 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. For $k\in\{0,...,\ell-1\}$, let $\delta_{g,\ell,k}$ be the numerator of the slope $s_{g,\ell,k}$ defined above. $\delta_{g,\ell,k}$ is the effect, on the expected

potential outcome of group g at $F_g-1+\ell$, of switching its kth treatment lag from its status-quo to its actual value, whereas all its previous treatments are held at their actual values, and all its subsequent treatments are held at their status-quo value. As $\delta_{g,\ell}=\sum_{k=0}^{\ell-1}\delta_{g,\ell,k}$,

$$\delta=\frac{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}\sum_{k=0}^{\ell-1}\delta_{g,\ell,k}}{\sum_{g:F_g\leq T_g}\sum_{\ell=1}^{T_g-F_g+1}(D_{g,F_g-1+\ell}-D_{g,1})}=\frac{\sum_{g:F_g\leq T_g}\sum_{k=0}^{T_g-F_g}\sum_{\ell=k+1}^{T_g-F_g+1}\delta_{g,\ell,k}}{\sum_{g:D_{g,1}=0,F_g\leq T_g}\sum_{k=0}^{T_g-F_g}(D_{g,F_g+k}-D_{g,1})}.$$

$\sum_{\ell=k+1}^{T_g-F_g+1}\delta_{g,\ell,k}$ is the total effect, from period F_g+k to T_g, of switching g's period F_g+k treatment from $D_{g,1}$ to D_{g,F_g+k}. The sum of total effects in the numerator of δ is scaled by the sum of all the incremental treatments $D_{g,F_g+k}-D_{g,1}$ that generate those total effects. Accordingly, δ may be interpreted as an average total effect per unit of treatment. The horizons over which the total effects are cumulated are conditional on the design, and vary across (g,k). To help interpret δ, one may compute the average number of cumulated effects T_g-F_g-k+1 across all (g,k).

Connection between δ and the event-study effects δ_ℓ. For $\ell\in\{1,...,L\}$, let $w_\ell=N_\ell/\sum_{g:F_g\leq T_g}\sum_{\ell'=1}^{T_g-F_g+1}(D_{g,F_g-1+\ell'}-D_{g,1})$.

Lemma 4 *If Assumptions 1 and 7 hold,*

$$\delta=\sum_{\ell=1}^{L}w_\ell\delta_\ell. \tag{21}$$

Lemma 4 shows that δ is equal to a linear combination, with non-negative factors, of the non-normalized event-study effects $(\delta_\ell)_{1\leq\ell\leq L}$.[14]

3.4 Inference

Our asymptotic results are based on the three assumptions below. Hereafter, let $\Sigma_g=V(\boldsymbol{Y}_g|\boldsymbol{D})$. For any symmetric matrix Σ, $\underline{\rho}(\Sigma)$ denotes its smallest eigenvalue. Let

$$\mathcal{L}=\{\ell:\lim_{G\to\infty}N_\ell=\infty \text{ almost surely}\}$$

[14] Following Lemma 4, one may define cost-benefit parameters combining event-study effects over a shorter horizon than δ. For values of ℓ close to L, the estimators of δ_ℓ may sometimes be very noisy: there may be few groups observed many periods after F_g, and at that point there may be few groups left that have not changed treatment and can be used as controls. Then, trimming those last effects may be warranted to increase statistical 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 related proposal.

be the set of values of ℓ such that the number of groups for which $\delta_{g,\ell}$ can be estimated goes to infinity almost surely (a.s.) as G tends to infinity.

Assumption 8 *(Independent groups) Conditional on $(\boldsymbol{D}_g)_{g\geq 1}$, the vectors $(\boldsymbol{Y}_g)_{g\geq 1}$ are mutually independent.*

Assumption 9 *(Asymptotic restriction on the design)*

1. $\limsup_G \#\mathcal{D}_1^r < \infty$ *almost surely (a.s.).*
2. $\mathcal{L} \neq \emptyset$.
3. *For $\ell \in \mathcal{L}$, let $v_{d,s,\ell}^G := \#\{g \leq G : D_{g,1} = d, S_g = s, F_g - 1 + \ell \leq T_g\}$. For all (d,s) such that $\lim_G v_{d,s,\ell}^G = \infty$ a.s., we have*

$$\liminf_G \frac{\#\{g \leq G : D_{g,1} = d, F_g = \max_{g':D_{g',1}=d} F_{g'}\}}{v_{d,s,\ell}^G} > 0 \quad a.s.$$

Assumption 10 *(Regularity conditions) We have, for some $\delta > 0$ and a.s.,*

$$\sup_{g\geq 1, t\in\{1,\ldots,T\}} E[|Y_{g,t}|^{2+\delta}|\boldsymbol{D}] < \infty \text{ and } \inf_{g\geq 1} \underline{\rho}\left(\Sigma_g\right) > 0.$$

Assumption 8 allows for serial correlation of the treatments and outcomes within each group, which are important features to account for in DID studies (Bertrand et al., 2004). Assumption 8 is also weaker than the common assumption that $(\boldsymbol{D}_g, \boldsymbol{Y}_g)_{g\geq 1}$ are i.i.d. or even simply independent. In particular, it allows for any form of dependence between groups' treatments $(\boldsymbol{D}_g)_{g\geq 1}$. Assumption 9 may be seen as a reinforcement and an asymptotic version of Assumption 1. Point 1 requires that the number of values of the baseline treatment remains finite when the number of groups goes to infinity. Point 2 requires that there is at least one value of ℓ for which the number of groups such that $\delta_{g,\ell}$ can be estimated goes to ∞ a.s., which is necessary to have that DID_ℓ is consistent. With i.i.d. groups, N_ℓ/G converges almost surely to a positive constant for all $\ell \in \mathcal{L}$. Instead, we do not impose any restriction on the rate at which N_ℓ goes to infinity. Point 3 requires that if the number of groups g such that (i) $\delta_{g,\ell}$ can be estimated and (ii) $(D_{g,1}, S_g) = (d,s)$ goes to infinity, then the number of groups g' with $D_{g',1} = d$ that switch the latest (or that never switch) also tends to infinity at the same rate. Otherwise, some control groups intervening in DID_ℓ may have a weight tending to infinity, which could make

asymptotic normality of DID_ℓ fail. Noteworthy, Point 3 of Assumption 9 automatically holds if groups' treatment paths $(\boldsymbol{D}_g)_{g\geq 1}$ are i.i.d. Finally, Assumption 10 ensures in particular that we 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 some non-trivial linear combinations of the $(Y_{g,1},...,Y_{g,T})$ in DID_ℓ would actually be constant.

We now establish the asymptotic properties of DID_ℓ when the number of groups tends to infinity. We also propose a confidence interval for δ_ℓ and show its asymptotic validity. To define this confidence interval, let $N^g_{t,\ell} = \sum_{g'\leq G: D_{g',1}=D_{g,1}} S_{g'} \mathbb{1}\left\{F_{g'} = t-\ell+1\right\}$ and

$$\begin{aligned}\lambda^G_{g,\ell,t} =& S_g \mathbb{1}\left\{F_g \leq T_g - \ell + 1\right\}\left(\mathbb{1}\left\{F_g = t-\ell+1\right\} - \mathbb{1}\left\{F_g = t+1\right\}\right) - \frac{N^g_{t,\ell}}{N^g_t}\mathbb{1}\left\{F_g > t\right\} \\ &+ \frac{N^g_{t+\ell,\ell}}{N^g_{t+\ell}}\mathbb{1}\left\{F_g > t+\ell\right\}.\end{aligned}$$

where, again, we use the convention that $0/0 = 0$. Then, define $\boldsymbol{\lambda}^G_{g,\ell} = (\lambda^G_{g,\ell,1},\cdots,\lambda^G_{g,\ell,T})$, $\boldsymbol{Y}_g = (Y_{g,1},\cdots,Y_{g,T})'$ and $U^G_{g,\ell} = \boldsymbol{\lambda}^G_{g,\ell}\boldsymbol{Y}_g$. Some algebra shows that

$$\text{DID}_\ell = \frac{1}{N_\ell}\sum_{g=1}^G U^G_{g,\ell}. \tag{22}$$

As a result, under Assumption 8,

$$V(N_\ell^{1/2}\text{DID}_\ell|\boldsymbol{D}) = \frac{1}{N_\ell}\sum_{g=1}^G E\left[(U^G_{g,\ell} - E(U^G_{g,\ell}|\boldsymbol{D}))^2|\boldsymbol{D}\right]. \tag{23}$$

We then consider an estimator of $V(N_\ell^{1/2}\text{DID}_\ell|\boldsymbol{D})$ of the form

$$\frac{1}{N_\ell}\sum_{g=1}^G \left(U^G_{g,\ell} - \widehat{\theta}_g\right)^2,$$

where $\widehat{\theta}_g$ is an estimator of $E(U^G_{g,\ell}|\boldsymbol{D})$. A difficulty is that in our i.n.i.d set-up, we cannot estimate consistently $E(U^G_{g,\ell}|\boldsymbol{D})$, as this expectation varies across groups. As a result, inference is in general conservative, whichever estimators $(\widehat{\theta}_g)_{g\geq 1}$ we consider. However, some estimators $(\widehat{\theta}_g)_{g\geq 1}$ lead to non-conservative inference if the $(\boldsymbol{D}_g,\boldsymbol{Y}_g)_{g\geq 1}$ are i.i.d. To define these $(\widehat{\theta}_g)_{g\geq 1}$, let us remark that under Assumption 9, the number of distinct values of $(D_{g,1},F_g,S_g)_{g:D_{g,1}\in\mathcal{D}_1^{\text{r}}}$ is finite;[15] we can thus write this set as $\{(d_k,f_k,s_k):\ k=1,...,K\}$. Then, let $\mathcal{C}_k = \{g\geq 1 : D_{g,1} =$

[15]Groups g such that $D_{g,1}\notin\mathcal{D}_1^{\text{r}}$ do not affect DID_ℓ (since $\boldsymbol{\lambda}^G_{g,\ell} = \boldsymbol{0}_T$) and thus can be ignored.

$d_k, F_g = f_k, S_g = s_k\}$ and $\mathcal{C}_k^G = \mathcal{C}_k \cap \{1, ..., G\}$. The $(\mathcal{C}_k^G)_{k=1,...,K}$ are called cohorts hereafter. If $g \in \mathcal{C}_k^G$, we define $\widehat{E}(U_{g,\ell}^G|\boldsymbol{D}) = (1/\#\mathcal{C}_k^G) \sum_{g' \in \mathcal{C}_k^G} U_{g',\ell}^G$, and we let

$$\widehat{\sigma}_\ell^2 = \frac{1}{N_\ell} \sum_{g=1}^{G} \left(U_{g,\ell}^G - \widehat{E}(U_{g,\ell}^G|\boldsymbol{D}) \right)^2.$$

The confidence interval of nominal level $1-\alpha$ on δ_ℓ that we consider is then

$$\text{CI}_{1-\alpha} = \left[\text{DID}_\ell \pm z_{1-\alpha/2} \frac{\widehat{\sigma}_\ell}{N_\ell^{1/2}} \right],$$

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

Theorem 1 *Suppose that Assumptions 3-4 and 8-10 hold. Then, for all $\ell \in \mathcal{L}$, conditional on $(\boldsymbol{D}_g)_{g\geq 1}$ and almost surely,*

$$DID_\ell - \delta_\ell \xrightarrow{P} 0,$$
$$\sqrt{N_\ell} \frac{DID_\ell - \delta_\ell}{\left(\frac{1}{N_\ell} \sum_{g=1}^{G} V(U_{g,\ell}^G|\boldsymbol{D}) \right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),$$
$$\liminf_{G\to\infty} \Pr\left[\delta_\ell \in CI_{1-\alpha} | \boldsymbol{D} \right] \geq 1-\alpha,$$

with an equality if we assume, instead of Assumption 8, that $(\boldsymbol{D}_g, \boldsymbol{Y}_g)_{g\geq 1}$ are i.i.d.

Theorem 1 shows that for almost all realizations of $(\boldsymbol{D}_g)_{g\geq 1}$, DID_ℓ is asymptotically normal, and the confidence interval of δ_ℓ is asymptotically conservative. If groups are identically distributed, the inequality becomes an equality and the confidence interval reaches its nominal level asymptotically. The results of Theorem 1 easily extend to the normalized parameter δ_ℓ^n: recall that $\delta_\ell^n = \delta_\ell / \delta_\ell^D$ and $\text{DID}_\ell^n = \text{DID}_\ell / \delta_\ell^D$, where δ_ℓ^D is a function of $\boldsymbol{D}$ only. Similar asymptotic results could also be obtained for the cost-benefit parameter δ.

3.5 Extensions

In this section we briefly review some extensions, discussed in details in our Web Appendix.

Placebo estimators In Section 1.1 of the Web Appendix, we propose placebo estimators one can use to test Assumptions 3 and 4. For any $g : 3 \leq F_g \leq T_g$ and $\ell \in \{1, ..., \min(T_g - F_g +$

$1, F_g-2)\}$, let

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

$\text{DID}^{\text{pl}}_{g,\ell}$ mimicks $\text{DID}_{g,\ell}$. Like $\text{DID}_{g,\ell}$, it compares the outcome evolution of g to that of groups with the same baseline treatment as g, and that have kept that treatment from period 1 to $F_g-1+\ell$. But unlike $\text{DID}_{g,\ell}$, it compares those groups' outcome evolutions from period $F_g-1-\ell$ to period F_g-1, namely before group g's treatment changes for the first time. Accordingly, $\text{DID}^{\text{pl}}_{g,\ell}$ assesses if g and groups not yet treated at $F_g-1+\ell$ experience the same evolution of their status-quo outcome over ℓ periods, the number of periods over which parallel trends has to hold for $\text{DID}_{g,\ell}$ to be unbiased for $\delta_{g,\ell}$. One can show that under Assumptions 3 and 4, $E(\text{DID}^{\text{pl}}_{g,\ell}|\boldsymbol{D}) = 0$. Whereas $\text{DID}_{g,\ell}$ goes from the past (period F_g-1) to the future (period $F_g-1+\ell$), $\text{DID}^{\text{pl}}_{g,\ell}$ goes from the future (period F_g-1) to the past (period $F_g-1-\ell$). This follows the standard practice in event-study regressions, where the reference period is the one before the event. In Section 1.1 of the Web Appendix, we define placebo estimators $\text{DID}^{\text{pl}}_{\ell}$ mimicking DID_{ℓ}. Essentially, the placebos replace $\text{DID}_{g,\ell}$ by $\text{DID}^{\text{pl}}_{g,\ell}$ in the estimators' definitions.[16]

Controlling for covariates. With covariates, we replace the equality in Assumption 4 by

$$\begin{aligned} &E[Y_{g,t}(\boldsymbol{D}_{g,1,t}) - Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1}) - (X_{g,t} - X_{g,t-1})'\theta_{D_{g,1}}|\boldsymbol{D},\boldsymbol{X}] \\ =&E[Y_{g',t}(\boldsymbol{D}_{g',1,t}) - Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1}) - (X_{g',t} - X_{g',t-1})'\theta_{D_{g',1}}|\boldsymbol{D},\boldsymbol{X}], \end{aligned} \tag{24}$$

where $\boldsymbol{X}$ stacks the covariates of all groups at all periods. (24) means that groups can experience differential trends, provided those differential trends are fully explained by changes in their covariates. Then, we estimate θ_d by regressing $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ in the sample of (g,t)s such that $D_{g,1} = d$ and $F_g > t$. Finally, we define similar estimators of δ_ℓ, δ^n_ℓ and δ as above, replacing $Y_{g,t} - Y_{g,t-1}$ by $Y_{g,t} - Y_{g,t-1} - (X_{g,t} - X_{g,t-1})'\widehat{\theta}_{D_{g,1}}$.[17] This approach can also be used to control for time-invariant covariates, defining $X_{g,t} = X_g \times t$. Then, $(X_{g,t} - X_{g,t-1})'\theta_{D_{g,1}} = X_g'\theta_{D_{g,1}}$, so (24) reduces to a conditional parallel-trends assumption, with a linear time-invariant functional form for the conditional counterfactual trend. Alternatively, one may define $X_{g,t}$ as

[16] `did_multiplegt_dyn` computes placebo estimators if one specifies the `placebo` option.

[17] `did_multiplegt_dyn` controls for covariates if one specifies the `controls` option.

the interaction of X_g and period fixed effects, in which case (24) reduces to a conditional parallel-trends assumption, with a period-specific linear functional form for the conditional trend.

Other extensions First, we show how to allow for group-specific linear trends. Second, we show how to allow for different trends across sets of groups defined by their value of a time-invariant covariate X_g. For instance, one may want to allow for state-specific trends in a county-level analysis.[18] Third, we show how to estimate heterogeneous treatment effects. Fourth, in Design 2, we propose an extension of our estimators, to separately estimate the effects of joining and leaving treatment. Fifth, we discuss how our estimators can be used in fuzzy designs. Sixth, ruling out the effect of past treatments beyond k lags allows us to propose a solution to the initial-conditions problem.

4 Current Practice

We conducted a census 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 reviewed 100 papers, and found 26 that have estimated at least one TWFE regression. The list of papers can be found in Table 1 in the Web Appendix. Of those 26 papers, two have a binary treatment with no variation in treatment timing,[19] 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

[18] `did_multiplegt_dyn` allows for different trends across sets of groups if one specifies the `trends_nonparam` option.

[19] 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'Haultfœuille (2022).

estimate dynamic effects, using one of the three estimations methods described below.

4.1 TWFE regressions with treatment intensity interacted with period FEs.

Designs with variation in treatment intensity and no variation in timing. Among the nine papers in our census that do not have a binary and staggered design and estimate dynamic effects, some are such that $D_{g,t} = I_g 1\{t \geq F\}$, for $2 \leq F \leq T$, a special case of Design 3 where groups all start getting treated at the same date F with group-specific intensities I_g, where I_g may be equal to zero if there are control groups that remain completely untreated. In such designs, researchers often estimate the following TWFE regression.

Regression 1 *(TWFE regression with treatment intensity interacted with period FEs) For every* $\ell \in \{-F+2, ..., T-F+1\}, \ell \neq 0$, *let* $\widehat{\beta}_{fe,\ell}$ *be the coefficient on* $I_g 1\{t = F-1+\ell\}$, *in a regression of* $Y_{g,t}$ *on group and period FEs and* $(I_g 1\{t = F-1+\ell\})_{\ell \in \{-F+2,...,T-F+1\}, \ell \neq 0}$.

Intuitively, for $\ell \in \{1, ..., T-F+1\}$, researchers hope that $\widehat{\beta}_{fe,\ell}$ estimates some average effect of increasing the treatment by one unit for ℓ periods. For $\ell \in \{-F+2, ..., -1\}$, researchers hope that $\widehat{\beta}_{fe,\ell}$ can be used as a placebo estimator, to test Assumptions 3 and 4.[20] Instead of Regression 1, researchers have also estimated regressions of $Y_{g,F-1+\ell} - Y_{g,F-1}$ on I_g. One can show that the coefficient on I_g in that regression is equal to $\widehat{\beta}_{fe,\ell}$.

A decomposition of the coefficients in TWFE regressions with treatment intensity interacted with period FEs. We now study what the coefficients $\widehat{\beta}_{fe,\ell}$ identify. First, note that when $D_{g,t} = I_g 1\{t \geq F\}$, $\delta_{g,\ell} = E(Y_{g,F-1+\ell}(\mathbf{0}_{F-1}, \boldsymbol{I}_{g,\ell}) - Y_{g,F-1+\ell}(\mathbf{0}_{F-1+\ell})|\boldsymbol{D})$, the effect of having received I_g rather than 0 units of treatment for ℓ periods. Let $\bar{I} = (1/G)\sum_{g=1}^{G} I_g$.

Proposition 1 *If Assumptions 3-4 hold,* $D_{g,t} = I_g 1\{t \geq F\}$ *for* $2 \leq F \leq T$ *and* I_g *is not constant across groups, then:*

1. $\forall\ \ell \in \{1, ..., T-F+1\}$,

$$E\left(\widehat{\beta}_{fe,\ell}|\boldsymbol{D}\right) = \sum_{g:I_g \neq 0} w_g^{fe} \frac{\delta_{g,\ell}}{I_g}, \tag{25}$$

[20] Assumption 4 reduces to a parallel-trends assumption on the never-treated outcome in those designs.

where

$$w_g^{fe} = \frac{I_g(I_g - \bar{I})}{\sum_{g':I_{g'}\neq 0} I_{g'}(I_{g'} - \bar{I})}.$$

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

$$E\left(\widehat{\beta}_{fe,\ell}|\boldsymbol{D}\right) = 0. \tag{26}$$

Interpretation of, and some remarks on, Proposition 1. Point 1 of Proposition 1 shows that for $\ell \in \{1, ..., T-F+1\}$, $\widehat{\beta}_{fe,\ell}$ estimates a weighted sum across groups of $\delta_{g,\ell}$, group g's effect of having received I_g rather than 0 units of treatment for ℓ periods, normalized by I_g. The weights w_g^{fe} sum to one but are not equal to $1/\#\{g : I_g \neq 0\}$, so $\widehat{\beta}_{fe,\ell}$ may be biased for the average of $\delta_{g,\ell}/I_g$ across the treated groups. Perhaps more worryingly, some of the weights may be negative: treatment effects of groups with a positive treatment intensity smaller than $\bar{I}$ are weighted negatively by $\widehat{\beta}_{fe,\ell}$. Then, $E\left(\widehat{\beta}_{fe,\ell}|\boldsymbol{D}\right)$ does not satisfy the no-sign-reversal property: it could be, say, negative, even if for all (g,t), $(d_1, ..., d_t) \mapsto Y_{g,t}(d_1, ..., d_t)$ is increasing in each of its arguments. By contrast, one can show that when $D_{g,t} = I_g 1\{t \geq F\}$,

$$\delta_\ell^n = \sum_{g:I_g\neq 0} \frac{I_g}{\sum_{g':I_{g'}\neq 0} I_{g'}} \frac{\delta_{g,\ell}}{I_g \ell},$$

a weighted average of group g's effect of having received I_g rather than 0 units of treatment for ℓ periods, normalized by $I_g\ell$, with only non-negative weights. Therefore, δ_ℓ^n satisfies the no-sign reversal property. The normalization of $\delta_{g,\ell}$ by $I_g\ell$ ensures that δ_ℓ^n can be interpreted as a weighted average of the slopes of the potential outcome function with respect to the current treatment and its first $\ell - 1$ lags. By contrast, the implicit normalization by I_g in $E\left(\widehat{\beta}_{fe,\ell}|\boldsymbol{D}\right)$ is hard to justify: at $F_g - 1 + \ell$, g has received $I_g\ell$ more treatment doses than in the status-quo counterfactual. Point 2 of Proposition 1 shows that for $\ell \in \{1, ..., F-2\}$, $E\left(\widehat{\beta}_{fe,\ell}|\boldsymbol{D}\right) = 0$ under Assumptions 3 and 4. Accordingly, testing $\beta_{fe,\ell} = 0$ is a valid test of Assumptions 3 and 4.

4.2 Local-projection panel regressions.

Local-projection panel regressions. Among the nine papers in our census that do not have a binary and staggered design and estimate dynamic effects, the second estimation method we

found are 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 2 *(Local-projection panel regressions) For every* $\ell \in \{1, ..., T-1\}$ *such that the regression that follows is well-defined, let* $\widehat{\beta}_{lp,\ell}$ *denote the coefficient on* $D_{g,t}$ *in a regression of* $Y_{g,t-1+\ell}$ *on group and period FEs and* $D_{g,t}$*, in the subsample such that* $1 \leq t \leq T-\ell+1$*.*

Intuitively, researchers hope that $\widehat{\beta}_{lp,\ell}$ estimates some average effect of increasing groups' period-t treatment by one unit on their period $t-1+\ell$ outcome. Researchers may estimate Regression 2 in first-difference, without group fixed effects. A result similar to that in Proposition 2 below also applies to that specification.

A decomposition of the coefficients in local-projection panel regressions. We decompose local-projection panel regressions under Assumption 2: this assumption is met in the papers of our census that have used those regressions. For all $\ell \in \{1, ..., T-1\}$, let $\widehat{\varepsilon}^{\ell}_{g,t}$ denote the residual from a regression of $D_{g,t}$ on group and time fixed effects, in the subsample such that $1 \leq t \leq T-\ell+1$. For all g such that $F_g \leq T$ and all $k \in \{1, ..., T-F_g+1\}$, let

$$\bar{D}_{g,k} = \frac{1}{k} \sum_{t=F_g}^{F_g-1+k} D_{g,t}$$

denote g's average treatment from F_g to F_g-1+k. Finally, let $\underline{F} = \min_g F_g$.

Proposition 2 *Suppose that Assumptions 2-4 hold.*

1. $\forall \ell \in \{1, ..., T-1\}$ *such that* $\widehat{\beta}_{lp,\ell}$ *is well-defined,*

$$E[\widehat{\beta}_{lp,\ell}|\boldsymbol{D}] = \sum_{k=1}^{T-\underline{F}+1} \sum_{g:\ell-k+1 \leq F_g \leq T-k+1} w^{lp,\ell}_{g,k} \frac{\delta_{g,k}}{\bar{D}_{g,k}},$$

where

$$w^{lp,\ell}_{g,k} = \frac{\bar{D}_{g,k}\widehat{\varepsilon}^{\ell}_{g,F_g-\ell+k}}{\sum_{k'=\ell}^{T-\underline{F}+1} \sum_{g':F_{g'} \leq T-k'+1} D_{g',F_{g'}-\ell+k'}\widehat{\varepsilon}^{\ell}_{g',F_{g'}-\ell+k'}}.$$

2. *In Design 1,* $\min_{g,k} w^{lp,\ell}_{g,k} < 0$ *for all* $\ell \geq 2$ *and* $\sum_{k=1}^{T-\underline{F}+1} \sum_{g:F_g-1+k \leq T} w^{lp,\ell}_{g,k} < 1$ *for all* $\ell \in \{2, ..., \underline{F}\}$*.*

3. *In Design 3,* $\sum_{k=1}^{T-\underline{F}+1} \sum_{g:F_g-1+k \leq T} w^{lp,1}_{g,k} = 1$*.*

Interpretation of, and some remarks on, Proposition 2. Proposition 2 shows that under Assumption 2, $\widehat{\beta}_{lp,\ell}$ estimates a weighted sum of $\delta_{g,k}$, group g's effect of having been exposed to the treatments $(D_{g,F_g},...,D_{g,F_g-1+k})$ rather than $\mathbf{0}_k$ for k periods, normalized by $\bar{D}_{g,k}$, g's average treatment from F_g to F_g-1+k. The weighted sum is across groups, and across the number k of periods of exposure to higher treatment doses. Accordingly, $\widehat{\beta}_{lp,\ell}$ does not estimate an average across groups of the effect of ℓ periods of exposure to higher treatment doses: $\widehat{\beta}_{lp,\ell}$ is contaminated by effects of other lengths of exposure. It is easy to understand where this contamination issue stems from in binary and staggered designs. Then, some groups with $D_{g,t}=1$ may have started receiving the treatment before period t, so the local-projection regression of $Y_{g,t-1+\ell}$ on $D_{g,t}$ captures an effect of more than ℓ periods of exposure for those groups. Similarly, some groups with $D_{g,t}=0$, which are supposed to be control groups, may have started receiving treatment between periods $t+1$ and $t-1+\ell$, and the local-projection regression of $Y_{g,t-1+\ell}$ on $D_{g,t}$ captures an effect of less than ℓ periods of exposure for those groups. Beyond this contamination phenomenon, a further issue is that some of the weights may be negative, and we actually show that for $\ell \geq 2$, some weights are always negative in Design 1. A last and perhaps even more concerning issue is that in Design 1, for $\ell \in \{2,...,\underline{F}\}$, the weights $w_{g,t}^{lp,\ell}$ sum to strictly less than one. This implies that even if there is a, say, positive real number θ such that $\delta_{g,k}/\bar{D}_{g,k}=\theta$, meaning that the treatment effect does not vary with length of exposure k or across groups g, $E\left(\widehat{\beta}_{lp,\ell}\right)<\theta$: the local-projection regression is downward biased. This is because the regression is misspecified: it considers groups with $D_{g,t}=0$ as untreated, whereas some of them may actually be treated at $t-1+\ell$. Though we were only able to prove formally that weights sum to less than one in binary and staggered designs, we also find weights summing to less than one in our empirical application, where the design is not binary or staggered. In fact, we even find that for some values of ℓ, weights sum to less than zero, meaning that even if $\delta_{g,k}/\bar{D}_{g,k}=\theta$, the expectation of the local-projection coefficient is of a different sign than θ. The issues with local-projection regressions we highlight in Proposition 2 are specific to Regression 2, an extension of local projection to panel data commonly found in applied work. Those issues are absent with time-series data where the unit receives a single shock, the design considered by Jordà (2005). Those issues are also absent in the careful extension of local projection regressions to panel data with binary and staggered treatments recently proposed by Dube et al. (2023), in

work posterior to ours.

4.3 Distributed-lag regressions.

Distributed-lag regressions. Among the nine papers in our census that do not have a binary and staggered design and estimate dynamic effects, the third estimation method we found are distributed-lag regressions.

Regression 3 *(Distributed-lag regression) For $l \in \{0, ..., K\}$, let $\widehat{\beta}_{dl,l}$ denote the coefficient on $D_{g,t-l}$ in a regression of $Y_{g,t}$ on group and period FEs and $(D_{g,t-l})_{l\in\{0,...,K\}}$, in the subsample such that $t \geq K+1$.*

Regression 3 is discussed in Angrist and Pischke (2008), see Equation (5.2.6) therein. In practice, researchers may slightly augment or modify Regression 3. 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 3 below apply to all those variations on Regression 3.

A decomposition of the coefficients in distributed-lag regressions.

Proposition 3 *(Application of Corollary 1 in de Chaisemartin and D'Haultfœuille (2023) to Regression 3) If Regression 3 is well-defined, Assumptions 3-4 hold and for all g and $t \geq K+1$ there exists real numbers $\left(\gamma^l_{g,t}\right)_{l\in\{0,...,K\}}$ such that for all $\boldsymbol{d} \in \{0,1\}^t$, $Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(\mathbf{0}_t) + \sum_{l=0}^{K} \gamma^l_{g,t} d_{t-l}$, then for all $l \in \{0, ..., K\}$,*

$$E\left[\widehat{\beta}_{dl,l}|\boldsymbol{D}\right] = \sum_{\substack{(g,t):D_{g,t-l}\neq 0,\\ t\geq K+1}} w^{dl,l}_{g,t}\gamma^l_{g,t} + \sum_{\substack{l'=0\\ l'\neq l}}^{K} \sum_{\substack{(g,t):D_{g,t-l'}\neq 0,\\ t\geq K+1}} w^{dl,l}_{g,t}\gamma^{l'}_{g,t}, \tag{27}$$

where $w^{dl,l}_{g,t}$ are weights proportional to the residuals in a regression of $D_{g,t-l}$ on period and group fixed effects and $(D_{g,t-l'})_{l'\in\{0,...,K\},l'\neq l}$, and such that

$$\sum_{\substack{(g,t):D_{g,t-l}\neq 0,\\ t\geq K+1}} w^{dl,l}_{g,t} = 1, \qquad \sum_{\substack{(g,t):D_{g,t-l'}\neq 0,\\ t\geq K+1}} w^{dl,l}_{g,t} = 0 \quad \forall l' \neq l.$$

Interpretation of, and some remarks on, Proposition 3. Proposition 3 is an application of Corollary 1 in de Chaisemartin and D'Haultfœuille (2023), a more general result applicable to any TWFE regression with several treatments, to distributed-lag regressions. Corollary 1 in de Chaisemartin and D'Haultfœuille (2023) was only proven for binary treatments, but extending it to non-binary treatments is straightforward. Proposition 3 is also related to the results of Sun and Abraham (2021) for event-study regressions in binary-and-staggered designs. To obtain this decomposition, we assume that

$$Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(\mathbf{0}_t) + \sum_{l=0}^{K} \gamma_{g,t}^{l} d_{t-l},$$

meaning that the functional form of 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, Proposition 3 shows that $\widehat{\beta}_{dl,l}$, the coefficient on the lth treatment lag, may not estimate a well-defined causal effect. Specifically, Proposition 3 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 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 summing to zero. If the effects of the other lags vary across (g,t) cells, those terms may differ from zero and may contaminate $\widehat{\beta}_{dl,l}$. Unlike Propositions 1 and 2, Proposition 3 does not decompose the regression coefficient under consideration as a weighted sum of the actual-versus-status-quo effects $\delta_{g,\ell}$, the building blocks of our analysis in Section 3, but as a weighted sum of the effects of the current and lagged treatments $\gamma_{g,t}^{l}$. However, under the assumptions of Proposition 3, $\delta_{g,\ell} = \sum_{l=0}^{K} \gamma_{g,t}^{l}(D_{g,F_g-1+\ell-l} - D_{g,1})$, so AVSQ effects are linear combinations of the effects of the current and lagged treatments.[21]

[21] The `twowayfeweights` command can be used to compute the weights attached to distributed-lag regressions, using the `other_treatments` option.

5 Application to banking deregulations and the housing market

5.1 Setting and research question

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 each state and year from 1994 to 2005. Favara and Imbs (2015) reverse their index, so their treatment is the number of restrictions lifted, ranging from 0 to 4. Several papers in the finance literature have aggregated these qualitatively different deregulations into a scalar variable. As the treatment is equal to 0 in every state in 1994, our event-study parameters are just average effects of having experienced some deregulations versus none, and do not rely on any homogeneity assumption on the effects of different policies.[22] Favara and Imbs (2015) use 1994-to-2005 county-level data to estimate the effect of the number of regulations lifted on the growth of mortgages originated by banks, and on the growth of houses prices.

5.2 Local-projection-regressions

Regression model. Favara and Imbs (2015) use a panel data version of the local-projection regressions proposed by Jordà (2005). Let Δ denote the first-difference operator. To estimate the treatment effect on, say, the growth rate of loan volume, they regress, for every $\ell \in \{1, ..., 9\}$, $\Delta \ln(L_{g,t-1+\ell})$, the log growth rate of loans in county g in year $t-1+\ell$, on county and year FEs, $D_{g,t-1}$, the number of deregulations in county g and year $t-1$, and some controls. Proposition 2 does not apply to their specification, because the authors define their treatment variable as $D_{g,t-1}$, the lagged treatment, and because the regression has some controls. Thus, we reestimate their local-projection regression, with $D_{g,t}$ instead of $D_{g,t-1}$ and without controls. The top-left

[22] Since no state had conducted deregulations prior to 1994, the initial-conditions problem is also not an issue.

panel of Figure 1 below shows that doing so, we find significant and positive coefficients $\widehat{\beta}_{lp,\ell}$ till $\ell = 5$, insignificant coefficients at $\ell = 6$, 7, and 9, and a significantly negative coefficient at $\ell = 8$. This is in line with the results the authors obtained with their regression, except that with their specification the coefficient remains positive and significant till $\ell = 4$ and is not significantly negative at $\ell = 8$ (see their Figure 1, Panel B). These results lead the authors to conclude that the growth effects of deregulation on credit supply are temporary. The bottom-left panel of Figure 1 shows that results are similar for the log growth rate of houses prices: coefficients are marginally significant and positive till $\ell = 4$, and insignificant after.

Decomposition of the local-projection coefficients. We use Proposition 2 to decompose the local-projection coefficients shown on the top-left panel of Figure 1.[23] $\widehat{\beta}_{lp,1}$ is a weighted sum of 7,626 effects $\delta_{g,k}/\bar{D}_{g,k}$, where 4,670 effects are weighted positively and 2,956 effects are weighted negatively, and where positive and negative weights respectively sum to 1.067 and -0.125. Weights do not sum to one, because this application does not meet the conditions in Design 3, but their sum is still not very far from one. $\widehat{\beta}_{lp,1}$, which is supposed to measure the effect of one year of exposure to treatment, is contaminated by effects of other lengths of exposure: weights on effects of one year of exposure sum to 0.294, and weights on effects of other lengths of exposure sum to 0.648. $\widehat{\beta}_{lp,2}$ is a weighted sum of 7,626 effects, where 4,424 effects are weighted positively and 3,202 effects are weighted negatively, and where positive and negative weights sum to 1.085 and -0.584. As in Point 3 of Proposition 2, and even though the design is not binary and staggered, the weights sum to strictly less than one, and they actually sum to much less than one. Then, even if $\delta_{g,k}/\bar{D}_{g,k}$ did not vary across g or k, $\widehat{\beta}_{lp,2}$ would be severely biased towards zero. $\widehat{\beta}_{lp,2}$ is contaminated by effects of other lengths of exposure than two years. Most of those effects are weighted positively, except for the effects of one year of exposure, whose weights sum to -0.472. Intuitively, this is due to the fact that groups with $D_{g,t} = 0, D_{g,t+1} > 0$ are used as "control groups" by $\widehat{\beta}_{lp,2}$, whereas they are treated at $t + 1$. Results are similar for $\widehat{\beta}_{lp,3}$, except that weights now only sum to 0.069, and effects of one and two years of exposure are weighted negatively, with weights summing to -0.429 and -0.478, respectively. For $\widehat{\beta}_{lp,4}$, weights sum to -0.018. This implies that even if $\delta_{g,k}/\bar{D}_{g,k}$ did not vary across g or k, $E\left[\widehat{\beta}_{lp,4}\right]$

[23] Results are very similar though not completely identical for the local-projection coefficients shown on the bottom-left panel, because counties with missing values are not exactly the same for the two outcomes.

would be of a different sign than the treatment effect. Results for $\widehat{\beta}_{lp,5}$ to $\widehat{\beta}_{lp,9}$ are similar to the results for $\widehat{\beta}_{lp,4}$. In particular, the sum of weights is negative for those coefficients.

5.3 DID_ℓ and DID_ℓ^n estimators.

Details on the design. Eight states never deregulate. 33 states deregulate only once, eight states deregulate twice, and one state deregulates three times. Of these 42 states that deregulate at least once, 38 do so for the first time in 1995, 1996, 1997, or 1998. Accordingly, up to $\ell = 2005-1998+1 = 8$, the δ_ℓ parameters apply to similar sets of counties: δ_1 and δ_8 respectively apply to 905 and 773 counties. δ_9, δ_{10}, and δ_{11} on the other hand only apply to a smaller subsample of counties (357, 238, and 1 county, respectively). Hence, we do not report estimates of those parameters.[24] Similarly, five placebo estimators can be computed, but only three apply to more than 50% of the 905 counties whose treatment changes at least once. The four and fifth placebos only apply to 128 and 120 counties respectively, so we do not report them.[25]

DID_ℓ estimators, for the loan-volume outcome. The top-center panel of Figure 1 plots the DID_ℓ estimators for the loan-volume outcome to the right of zero, and the corresponding placebo estimators to the left. $\text{DID}_1 = 0.043$ (s.e.=0.035): after one year of exposure to deregulation, the loan-volume growth rate increases by 4.3 percentage points 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 three years of exposure ($\text{DID}_3 = 0.081$, s.e.=0.049), and at the 5% level after five years of exposure ($\text{DID}_5 = 0.148$, s.e.=0.064). To the left of zero, placebo estimates are shown. Placebos are jointly insignificant (F-test p-value=0.400). But the placebos are relatively large, and imprecisely estimated. For instance, several DID_ℓ estimates are below the line connecting the lower bound of the confidence interval of DID_1^{pl} and $(0, 0)$. Moreover, we can only test the parallel-trends assumption over three years, whereas that assumption has to hold for eight years for DID_8 to be unbiased. Overall, even if the placebos do not clearly indicate a violation of the parallel-trends assumption, they may fail to detect violations of parallel trends

[24] For the two outcomes we consider, the estimates of δ_9 and δ_{10} are even larger than those of δ_8.

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

large enough to substantially bias the DID_ℓ estimates (see Roth, 2022).

DID^n_ℓ estimators, for the loan-volume outcome. The top-right panel of Figure 1 plots the DID^n_ℓ estimators for the loan-volume outcome. In order to test for Point 1 of Lemma 3, we restrict the sample of switchers to the 773 counties for which $\text{DID}_{g,8}$ can be computed, and further discard 141 switchers in states that experienced more than one treatment change (results are not very different if one does not restrict the sample). DID^n_ℓ are estimators of the effect of the current treatment and of its $\ell-1$ first lags on the outcome. In this restricted sample, each lag receives a weight equal to $1/\ell$. With the exception of DID^n_1, which is much larger than the other estimates but also imprecisely estimated, all estimates are relatively close to each other, suggesting that treatment effects are relatively stable over time, and that the effects of the current and lagged treatments are not very different. To formally test this, we follow Point 1 of Lemma 3 and test whether $\ell \mapsto \text{DID}^n_\ell$ is constant. The test is not rejected (p-value = 0.328), which goes against the authors' conclusion that deregulations only have short-lived effects on mortage volume. Our decompositions of their local projection estimators suggest that their coefficients may become small and insignificant because the sum of the weights attached to $\widehat{\beta}_{lp,\ell}$ decreases as ℓ increases.

DID_ℓ and DID^n_ℓ estimators, for the houses-prices outcome. The bottom-center panel of Figure 1 plots the DID_ℓ estimators for the houses-prices outcome. $\text{DID}_1 = 0.003$ (s.e.=0.004): after one year of exposure to deregulation, the houses-price growth rate increases by 0.3 percentage points 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 four years of exposure ($\text{DID}_4 = 0.016$, s.e.=0.009), and at the 5% level after five years of exposure ($\text{DID}_5 = 0.026$, s.e.=0.010). Placebos are jointly insignificant (F-test p-value=0.703). For this outcome, the placebos are small, fairly precisely estimated, and suggest that if anything, differential pre-trends would downward-bias the estimated effects. The bottom-right panel of Figure 1 plots the DID^n_ℓ estimators, restricting the sample of switchers to counties for which $\text{DID}_{g,8}$ can be computed and in states that experienced only one treatment change. Estimates increase with ℓ, thus suggesting that older treatments have a larger impact than more recent ones: deregulations seem to also have long-lasting effects on houses prices.

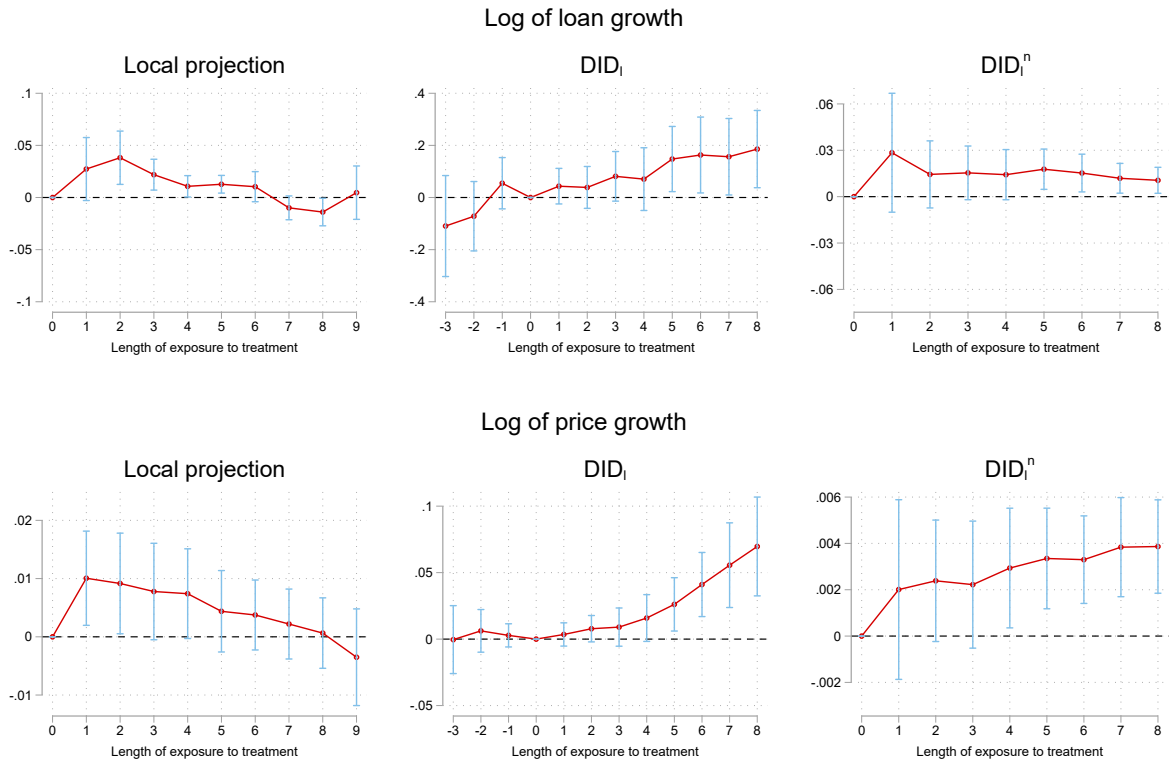

Figure 1: Effect of banking deregulations on loan volume and houses prices.

Notes. The top-left panel shows estimated effects of banking deregulations on loan volume, according to local-projection regressions of $\Delta \ln(L_{g,t-1+\ell})$, the log growth rate of loans in county g in year $t-1+\ell$, on county and year FEs, and $D_{g,t}$, the number of deregulations in county g and year t, for $\ell \in \{1, ..., 9\}$. To the right of zero, the top-centre panel shows DID_ℓ estimates of the effect of having been exposed to banking deregulations for ℓ periods on the log growth rate of loans, for $\ell \in \{1, ..., 8\}$. To the left of zero, the top-center panel shows $\text{DID}_\ell^{\text{pl}}$ placebo estimates, for $\ell \in \{1, ..., 3\}$. The top-right panel shows the DID_ℓ^n estimates, for $\ell \in \{1, ..., 8\}$. The bottom-left, bottom-centre, and bottom-right panels are similar to the top-left, top-centre, and top-right panels, with houses prices instead of loans as the outcome. Estimates in the bottom panels are weighted by the inverse of the number of counties per state: Favara and Imbs (2015) weight their regressions by that variable when they study the effect of deregulations on houses prices. 95% confidence intervals relying on a normal approximation and standard errors clustered at the state level are shown in red on all panels. All estimates are computed using the 1994-2005 county-level panel data set constructed by Favara and Imbs (2015).

References

Abadie, A. (2005, 01). Semiparametric difference-in-differences estimators. *Review of Economic Studies 72*(1), 1–19.

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

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

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

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

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

Borusyak, K. and X. Jaravel (2017). Revisiting event study designs. Working Paper.

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

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

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

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

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

Cox, G. and X. Shi (2023). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *The Review of Economic Studies 90*(1), 201–228.

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

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

de Chaisemartin, C. and X. D'Haultfœuille (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. and X. D'Haultfœuille (2023). Two-way fixed effects and differences-in-differences estimators with several treatments. arXiv preprint arXiv:2012.10077.

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

de Chaisemartin, C. and X. D'Haultfœuille (2020). Difference-in-differences estimators of intertemporal treatment effects. arXiv preprint 2007.04267v8.

Deryugina, T. (2017). The fiscal cost of hurricanes: Disaster aid versus social insurance. *American Economic Journal: Economic Policy 9*(3), 168–98.

Dube, A., D. Girardi, O. Jorda, and A. M. Taylor (2023). A local projections approach to difference-in-differences event studies. National Bureau of Economic Research.

East, C. N., S. Miller, M. Page, and L. R. Wherry (2023). Multigenerational impacts of childhood access to the safety net: Early life exposure to medicaid and the next generation's health. *American Economic Review 113*(1), 98–135.

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

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

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

Harmon, N. A. (2022). Difference-in-differences and efficient estimation of treatment effects. working paper.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

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

Krolikowski, P. (2018). Choosing a control group for displaced workers. *ILR Review 71*(5), 1232–1254.

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

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

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

Rice, T. and P. E. Strahan (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. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305–22.

Small, D. and Z. Tan (2007). A stochastic monotonicity assumption for the instrumental variables method. Working paper.

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

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

Clément de Chaisemartin* Xavier D'Haultfœuille†

Abstract

In the first section of this web appendix, we consider various extensions of our baseline estimators: we introduce placebo estimators; we propose estimators controlling for covariates; we propose estimators allowing for group-specific linear trends; we propose estimators allowing for different trends across sets of groups; in Design 2, we propose an extension of our estimators, to separately estimate the effects of joining and leaving the treatment; we discuss how our estimators can be used in fuzzy designs; we outline some of the benefits of ruling out the effect of past treatments beyond k lags. In the second section, we list the papers in our census of highly cited TWFE papers published by the AER. In the last section, we prove the results in the paper and in the web appendix.

1 Extensions

1.1 Placebo estimators

We now propose placebo estimators of Assumptions 3 and 4. First, for any g such that $3 \leq F_g \leq T_g$ and for any $\ell \in \{1, ..., \min(T_g - F_g + 1, F_g - 2)\}$, let

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

$\text{DID}^{\text{pl}}_{g,\ell}$ 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 with the same baseline treatment, and that have kept that treatment from period 1 to $F_g - 1 + \ell$. But unlike $\text{DID}_{g,\ell}$, it compares those groups' outcome evolutions from period $F_g - \ell - 1$ to period $F_g - 1$, namely before group g's treatment changes for the

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

first time. Accordingly, $\text{DID}^{\text{pl}}_{g,\ell}$ assesses if g and its control groups experience parallel evolutions of their status-quo potential outcomes, for ℓ 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 $F_g-1+\ell$), $\text{DID}^{\text{pl}}_{g,\ell}$ goes from the future (period F_g-1) to the past (period $F_g-1-\ell$). This follows the standard practice in event-study regressions, where the reference period is the one before the event.

Then, let $L^{\text{pl}}=\max_g\{\min(T_g-F_g+1,F_g-2)\}$. L^{pl} is the largest ℓ such that there is at least one group for which $\text{DID}_{g,\ell}$ can be estimated ($F_g-1+\ell\leq T_g$) and for which $\text{DID}^{\text{pl}}_{g,\ell}$ can also be estimated ($F_g-1-\ell\geq 1\Leftrightarrow\ell\leq F_g-2$). One may have $L^{\text{pl}}=-1$, if all groups that switch treatment do so for the first time at period 2. Then, none of the placebos defined below can be computed, as one does not observe the outcome evolution of any switching group before it switches. Outside of that special case, $L^{\text{pl}}\geq 0$, so one can at least compute one of the $\text{DID}^{\text{pl}}_{\ell}$ estimators below.

For all $\ell\in\{1,...,L^{\text{pl}}\}$, let $N^{\text{pl}}_{\ell}=\#\{g:1\leq F_g-1-\ell,F_g-1+\ell\leq T_g\}$ be the number of groups for which both $\text{DID}_{g,\ell}$ and $\text{DID}^{\text{pl}}_{g,\ell}$ can be computed. Then, let

$$\text{DID}^{\text{pl}}_{\ell}=\frac{1}{N^{\text{pl}}_{\ell}}\sum_{g:1\leq F_g-1-\ell,F_g-1+\ell\leq T_g}S_g\text{DID}^{\text{pl}}_{g,\ell}.$$

Lemma 5 *If Assumptions 1, 3 and 4 hold and $L^{pl}\geq 1$, then $\forall\ell\in\{1,...,L^{pl}\}$, $E\left[DID^{pl}_{\ell}|\boldsymbol{D}\right]=0$.*

Lemma 5 shows that $E\left[\text{DID}^{\text{pl}}_{\ell}|\boldsymbol{D}\right]=0$ is a testable implication of Assumptions 3 and 4.

The placebo estimators $\text{DID}^{\text{pl}}_{\ell}$ mimick our actual estimators, but accordingly they do not exhaust all our assumptions' testable implications. 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 DID_{ℓ}'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 $F_{g'}>F_g-1+\ell$ and $g':D_{g',1}=D_{g,1}$, the sign of

$$E\left[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,t})-Y_{g,F_g-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}\right]-E\left[Y_{g',F_g-1+\ell}(\boldsymbol{D}_{g',1,t})-Y_{g',F_g-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}\right]$$

does not depend on (g,g'). In words, the sign of the difference in trends between a switcher and one of its control group has to be the same, for every pair of switcher and control. Then, the sign of the bias of DID_{ℓ} 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 controls differential trends after switchers' treatment changes. This requirement is more plausible when placebos closely mimick the actual estimators.

Computation. Our placebo estimators are computed by the `did_multiplegt_dyn` Stata package, when the `placebo` option is specified.

1.2 Controlling for covariates

Identifying assumption. To control for covariates, we replace Assumption 4 by the following assumption. Let $\boldsymbol{X}_g = (X'_{g,1}, ..., X'_{g,T})$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_G)$.

Assumption 11 *(Parallel trends with covariates) There are vectors $(\theta_d)_{d\in\mathcal{D}_1^r}$ of same dimension as $X_{g,t}$ such that $\forall(g,g')$ such that $D_{g,1} = D_{g',1} \in \mathcal{D}_1^r$, then $\forall t \geq 2$,*

$$\begin{aligned}&E[Y_{g,t}(\boldsymbol{D}_{g,1,t}) - Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1}) - (X_{g,t} - X_{g,t-1})'\theta_{D_{g,1}}|\boldsymbol{D}, \boldsymbol{X}]\\ =&E[Y_{g',t}(\boldsymbol{D}_{g',1,t}) - Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1}) - (X_{g',t} - X_{g',t-1})'\theta_{D_{g',1}}|\boldsymbol{D}, \boldsymbol{X}],\end{aligned}$$

Assumption 11 is similar to Assumption 4, except that now groups can experience differential trends, provided those differential trends are fully explained by changes in their covariates. Note that the effect of these covariates is allowed to vary with groups' baseline-treatment value. To control for time-invariant covariates, one can for instance let $X_{g,t} = X_g \times t$. Then, $(X_{g,t} - X_{g,t-1})'\theta_{D_{g,1}} = X'_g\theta_{D_{g,1}}$, so Assumption 11 amounts to a conditional parallel-trends assumption, with a linear time-invariant functional form for the conditional counterfactual trend. Alternatively, one can let $X_{g,t}$ be a vector of interactions between X_g and time fixed effects. Then, Assumption 11 amounts to a conditional parallel-trends assumption, with a linear period-specific functional form for the conditional counterfactual trend.

Estimation. For all $d \in \mathcal{D}_1^{\mathrm{r}}$, let $\widehat{\theta}_d$ 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 $D_{g,1} = d$ and $F_g > t$. For every g such that $F_g \leq T_g$ and $\widehat{\theta}_{D_{g,1}}$ can be computed, and for every $\ell \in \{1, ..., T_g - F_g + 1\}$, let

$$\begin{aligned}\mathrm{DID}^X_{g,\ell} =&Y_{g,F_g-1+\ell} - Y_{g,F_g-1} - (X_{g,F_g-1+\ell} - X_{g,F_g-1})'\widehat{\theta}_{D_{g,1}}\\ &-\frac{1}{N^g_{F_g-1+\ell}}\sum_{g':D_{g',1}=d,F_{g'}>F_g+\ell}\left[Y_{g',F_g-1+\ell} - Y_{g',F_g-1} - (X_{g',F_g-1+\ell} - X_{g',F_g-1})'\widehat{\theta}_{D_{g,1}}\right].\end{aligned}$$

$\mathrm{DID}^X_{g,\ell}$ is similar to $\mathrm{DID}_{g,\ell}$, except that instead of comparing groups' outcome evolution, it compares the part of that evolution that is not explained by a change in their covariates. Following the same steps as those used to prove Lemma 1, one can show that if Assumptions 1-3 and 11 hold, then for every g such that $F_g \leq T_g$ and $\widehat{\theta}_{D_{g,1}}$ can be computed, and for every $\ell \in \{1, ..., T_g - F_g + 1\}$, we have

$$E\left[\mathrm{DID}^X_{g,\ell}\middle|\boldsymbol{D}, \boldsymbol{X}\right] = \delta_{g,\ell}.$$

Then, conditionally unbiased estimators of $(\delta_\ell)_{1\leq\ell\leq L}$ controlling for the covariates can be proposed, by replacing the $\delta_{g,\ell}$ parameters by their estimators $\text{DID}^X_{g,\ell}$ in Equation (4). Similarly, conditionally unbiased estimators of the $(\delta^n_\ell)_{1\leq\ell\leq L}$ and δ can be proposed. One can also follow the exact same steps to propose placebo estimators to test Assumptions 3 and 11.

Computation. Our estimators controlling for covariates are computed by the `did_multiplegt_dyn` Stata package, when the `controls` option is specified.

1.3 Allowing for group-specific linear trends

Identifying assumption. To allow for group-specific linear trends, we replace Assumption 4 by the following assumption.

Assumption 12 *(Parallel trends for the first-differenced status-quo outcome, conditional on the period-one treatment)* $\forall(g,g')$*, if* $D_{g,1}=D_{g',1}\in\mathcal{D}^r_1$*, then* $\forall t\geq 3$*,*

$$\begin{aligned}&E[Y_{g,t}(\boldsymbol{D}_{g,1,t})-Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})-(Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})-Y_{g,t-2}(\boldsymbol{D}_{g,1,t-2})|\boldsymbol{D}]\\ =&E[Y_{g',t}(\boldsymbol{D}_{g',1,t})-Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})-(Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})-Y_{g',t-2}(\boldsymbol{D}_{g',1,t-2})|\boldsymbol{D}].\end{aligned}$$

Assumption 12 is a parallel trends assumption on groups' first-differenced status-quo outcomes, under which groups may have different first-differenced outcome, provided groups with the same baseline treatment all have the same expected second-differenced outcome. Mora and Reggio (2019) already considered Assumption 12 in designs with a binary treatment and no variation in timing. Assumption 12 extends their assumption to the designs we consider here.

Estimation. For every g such that $3\leq F_g\leq T_g$, and every $\ell\in\{1,...,T_g-F_g+1\}$, let

$$\begin{aligned}\text{DID}^{fd}_{g,\ell}=&Y_{g,F_g-1+\ell}-Y_{g,F_g-1+\ell-1}-(Y_{g,F_g-1}-Y_{g,F_g-2})\\ &-\frac{1}{N^g_{F_g-1+\ell}}\sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell}(Y_{g',F_g-1+\ell}-Y_{g',F_g-1+\ell-1}-(Y_{g',F_g-1}-Y_{g',F_g-2})),\end{aligned}$$

a DID estimator comparing the F_g-1-to-$F_g-1+\ell$ evolution of g's first-differenced outcome to that of groups with the same baseline treatment, and that have kept that treatment from period 1 to $F_g-1+\ell$.

Lemma 6 *If Assumptions 3 and 12 hold, then for every* (g,ℓ) *such that* $F_g\geq 3$ *and* $1\leq\ell\leq T_g-F_g+1$*,* $E\left[DID^{fd}_{g,\ell}\middle|\boldsymbol{D}\right]=\delta_{g,\ell}-\delta_{g,\ell-1}$*, with the convention that* $\delta_{g,0}=0$*.*

Let $L^{fd}=\max_{g:F_g\geq 3}(T_g-F_g+1)$. For every $\ell\in\{1,...,L^{fd}\}$ and every $\ell'\in\{1,...,\ell\}$, let $N^{fd}_\ell=\#\{g:F_g\geq 3,F_g-1+\ell\leq T_g\}$, and

$$\text{DID}^{fd}_{\ell,\ell'}=\frac{1}{N^{fd}_\ell}\sum_{g:F_g\geq 3,F_g-1+\ell\leq T_g}S_g\text{DID}^{fd}_{g,\ell'}.$$

It directly follows from Lemma 6 that

$$E\left[\text{DID}^{fd}_{\ell,\ell'}\middle|\boldsymbol{D}\right]=\frac{1}{N^{fd}_{\ell}}\sum_{g:F_g\geq 3,F_g-1+\ell\leq T_g}S_g(\delta_{g,\ell'}-\delta_{g,\ell'-1}).$$

Therefore,

$$\sum_{\ell'=1}^{\ell}\text{DID}^{fd}_{\ell,\ell'}=\frac{1}{N^{fd}_{\ell}}\sum_{g:F_g\geq 3,F_g-1+\ell\leq T_g}S_g\delta_{g,\ell}\equiv\delta^{fd}_{\ell},$$

where δ^{fd}_{ℓ} generalizes the non-normalized event-study effect δ_ℓ to a model allowing for group-specific linear trends. The normalized event-study effects and cost-benefit ratio can be generalized and estimated similarly.

1.4 Allowing for different trends across sets of groups

Identifying assumption. 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 $s(g)\in\{1,...,S\}$ denote the set of groups g belongs to. In this set-up, we modify our assumptions as follows.

Assumption 13 *(Parallel trends within sets of groups)* $\forall(g,g')$*, if* $s(g)=s(g')$*,* $D_{g,1}=D_{g',1}\in\mathcal{D}^r_1$*, then* $\forall t\geq 2$*,*

$$E[Y_{g,t}(\boldsymbol{D}_{g,1,t})-Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}]=E[Y_{g',t}(\boldsymbol{D}_{g',1,t})-Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}].$$

Assumption 13 is a weakening of Assumption 4, as it only requires that the status-quo potential outcome of groups in the same set of groups follow the same evolution over time.

Assumption 14 *(Restriction on the design, with sets of groups)* $\exists(g,g'):s(g)=s(g')$*,* $D_{g,1}=D_{g',1}$ *and* $F_g<F_{g'}$*.*

Assumption 14 is a strengthening of Assumption 1: we should now observe some variation in the timing of the first switch not only among groups with the same baseline treatment, but also within the same set of groups.

Estimation. We redefine T_g and N^g_t as follows. First, we let $T^s_g=\max_{g':s(g')=s(g),D_{g',1}=D_{g,1}}F_{g'}-1$. Second, we let $N^{g,s}_t=\#\{g':s(g')=s(g),D_{g',1}=D_{g,1},F_{g'}>t\}$. Then, for every g such that $F_g\leq T^s_g$, and every $\ell\in\{1,...,T^s_g-F_g+1\}$, let

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

$\text{DID}^s_{g,\ell}$ is similar to $\text{DID}_{g,\ell}$, except that it only compares the outcome evolution of groups in the same set of groups. For instance, if groups are US counties and sets of groups are US states, $\text{DID}^s_{g,\ell}$ compares the outcome evolution of counties in the same state. Following the same steps as those used to prove Lemma 1, one can show that if Assumptions 3 and 13-14 hold, then for every g such that $F_g \leq T^s_g$ and $\ell \in \{1, ..., T^s_g - F_g + 1\}$, we have $E\left[\text{DID}^s_{g,\ell} \middle| \boldsymbol{D}\right] = \delta_{g,\ell}$. Then, conditionally unbiased estimators of $(\delta_\ell)_{1\leq\ell\leq L^s}$ (with $L^s := \max_g(T^s_g - F_g + 1)$) 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 (4). Similarly, conditionally unbiased estimators of the $(\delta^n_\ell)_{1\leq\ell\leq L^s}$ and δ can be proposed. One can also follow the exact same steps to propose placebo estimators to test Assumptions 3 and 13.

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

1.5 Estimating and testing the heterogeneity of treatment effects.

Let β^{het}_ℓ denote the coefficients on X_g in an (unfeasible) OLS regression of $S_g\delta_{g,\ell}$, the effect of having been exposed to a weakly higher treatment for ℓ periods in group g on X_g and indicators for all possible values of $F_g \times D_{g,1} \times S_g$, in the sample of groups such that $F_g - 1 + \ell \leq T_g$. β^{het}_ℓ is a relevant quantity to test for heterogeneous treatment effects. If X_g is a scalar binary variable, it follows from Angrist (1998) that β^{het}_ℓ is a variance-weighted average of comparisons of $S_g\delta_{g,\ell}$ across groups with $X_g = 1$ and $X_g = 0$ belonging to the same $F_g \times D_{g,1} \times S_g$ cell. It follows from Angrist and Krueger (1999) that a similar interpretation holds whenever X_g is a scalar ordered variable: for instance, if X_g is continuous, β^{het}_ℓ is a variance-weighted average of the derivative of the conditional mean of $S_g\delta_{g,\ell}$ with respect to X_g, within $F_g \times D_{g,1} \times S_g$ cells. When X_g is a vector, β^{het}_ℓ may no longer be interpreted as an average of average differences of $S_g\delta_{g,\ell}$ across groups with different values of X_g and the same value of $F_g \times D_{g,1} \times S_g$. Still, it can be used to test the null that $S_g\delta_{g,\ell}$ does not vary across g, and to identify covariates that predict $S_g\delta_{g,\ell}$.

To estimate β^{het}_ℓ, we consider $\widehat{\beta}^{het}_\ell$, the coefficients on X_g in OLS regressions of $S_g(Y_{g,F_g-1+\ell} - Y_{g,F_g-1})$ on X_g and indicators for all possible values of $F_g \times D_{g,1} \times S_g$, in the sample of groups such that $F_g - 1 + \ell \leq T_g$. We thus use $S_g(Y_{g,F_g-1+\ell} - Y_{g,F_g-1})$ rather than $S_g\text{DID}_{g,\ell}$ as the regression outcome. The advantage is that inference does not need to account for the fact the $\text{DID}_{g,\ell}$s are estimated. Lemma 7 shows that $\widehat{\beta}^{het}_\ell$ is unbiased for β^{het}_ℓ, under the following parallel trends assumption.

Assumption 15 *(Parallel trends to test for heterogeneous effects)* $\forall(g, g')$ *such that* $D_{g,1} =$

$D_{g',1} \in \mathcal{D}_1^r$, $\forall t \geq 2$,

$$E[Y_{g,t}(\boldsymbol{D}_{g,1,t}) - Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}, \boldsymbol{X}] = E[Y_{g',t}(\boldsymbol{D}_{g',1,t}) - Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}, \boldsymbol{X}].$$

Lemma 7 *If Assumptions 1, 3 and 15 hold, then* $\forall \ell \in \{1, ..., L\}$, $E\left[\widehat{\beta}_\ell^{het} \middle| \boldsymbol{D}, \boldsymbol{X}\right] = \beta_\ell^{het}$.

β_ℓ^{het} does not make use of control groups. That such groups are not necessary to estimate heterogeneous treatment effects in DID studies was already noted before this paper by Shahn (2023) and de Chaisemartin and D'Haultfœuille (2023), in simple designs without variation in treatment intensity or timing. Lemma 7 shows that their results extend to more general designs, provided one controls for indicators for all possible values of $F_g \times D_{g,1} \times S_g$. Our procedure cannot be used to estimate if $\delta_{g,\ell}^n$ varies across values of X_g, but one can control for $\delta_{g,\ell}^D$ in the heterogeneity analysis, to assess if X_g predicts $S_g\delta_{g,\ell}$ even controlling for $\delta_{g,\ell}^D$.

1.6 Separately identifying switching-in and switching-out effects in Design 2

Consider Design 2, in which $D_{g,t} = 1\{E_g \geq t \geq F_g\}$, with the convention that $E_g = T + 1$ if $F_g \leq T$ and g never leaves treatment. Then, we show that one can separately estimate the effects of switching in and switching out of the treatment. First, let $N_\ell^+ = \#\{g : E_g - F_g \geq \ell - 1\}$ and let

$$\delta_\ell^+ = \frac{1}{N_\ell^+} \sum_{g:E_g - F_g \geq \ell - 1} \delta_{g,\ell}. \tag{28}$$

This parameter is the average effect of having been treated for ℓ periods, across all groups that have not switched out of the treatment ℓ periods after the first date where they go treated. If there exist (g, g') such that $\ell \leq E_g - F_g + 1$ and $F_{g'} \geq E_g - F_g + 1$, we can unbiasedly estimate δ_ℓ^+ under Assumptions 3-4, by replacing $\delta_{g,\ell}$ by $\text{DID}_{g,\ell}$ in (28). In practice, this estimator can be computed using the `did_multiplegt_dyn` command, on the subsample of (g, t)s such that g has not switched out of the treatment at or before t.

Next, let us define

$$\delta_{g,\ell}^- = E[Y_{g,E_g+\ell}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{E_g-F_g+1}, \boldsymbol{0}_\ell) - Y_{g,E_g+\ell}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{E_g-F_g+1+\ell})|\boldsymbol{D}].$$

$\delta_{g,\ell}^-$ is the effect of having been untreated rather than treated for ℓ periods, in group g and at period $E_g + \ell$. To unbiasedly estimate $\delta_{g,\ell}^-$, we consider the following modified version of Assumption 4:

Assumption 16 *(Parallel trends if groups treated at the same date never leave treatment)* $\forall (g, g')$, *if* $F_g = F_{g'}$, *then* $\forall t \geq F_g$,

$$\begin{aligned} &E[Y_{g,t}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{t-(F_g-1)}) - Y_{g,t-1}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{t-1-(F_g-1)})|\boldsymbol{D}] \\ =&E[Y_{g',t}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{t-(F_g-1)}) - Y_{g',t-1}(\boldsymbol{0}_{F_g-1}, \boldsymbol{1}_{t-1-(F_g-1)})|\boldsymbol{D}]. \end{aligned}$$

Now, let $N_\ell^{g,-} = \#\{E_g + \ell \leq T, \exists g' : F_{g'} = F_g, E_{g'} \geq E_g + \ell\}$ and assume that $N_\ell^{g,-} > 0$. Then, under Assumptions 3 and 16,

$$\text{DID}_{g,\ell}^- = Y_{g,E_g+\ell} - Y_{g,E_g} - \frac{1}{N_\ell^{g,-}} \sum_{g':F_{g'}=F_g, E_{g'}\geq E_g+\ell} Y_{g',E_g+\ell} - Y_{g',E_g}$$

is an unbiased estimator of $\delta_{g,\ell}^-$.

Finally, let $N_\ell^- = \#\{g : N_\ell^{g,-} > 0\}$ and define the aggregated effect

$$\delta_\ell^- = \frac{1}{N_\ell^-} \sum_{g:N_\ell^{g,-}>0} \delta_{g,\ell}^-.$$

We can unbiasedly estimate δ_ℓ^- by replacing $\delta_{g,\ell}^-$ by $\text{DID}_{g,\ell}^-$ in the previous display. In practice, we can obtain this estimator by using the `did_multiplegt_dyn` command on the subsample of (g,t)s such that group g has switched from 0 to 1 at or before t, including the date of that switch in the `trends_nonparam` option, and defining the treatment as an indicator for switching from 1 to 0. Assumptions 3 and 16 can be tested by adding the `placebo` option to that estimation.

1.7 Fuzzy designs

In this subsection, we briefly discuss fuzzy designs. These correspond to cases with $N_{g,t} > 1$ units in group g at t, and where treatment varies across units within (g,t) cells. The estimators we propose in this paper are readily applicable if all groups are fully untreated at period 1. Then, one just needs to redefine the treatment, the counterfactual outcome without any treatment, and the observed treatment of cell (g,t) as follows. Let $D_{i,g,t}$ denote the treatment of unit i in cell (g,t), and let $D_{g,t} = 1/N_{g,t} \sum_{i=1}^{N_{g,t}} D_{i,g,t}$ denote the average treatment in cell (g,t). Also, let $Y_{i,g,t}(\mathbf{0}_t)$ denote the potential outcome of unit i in cell (g,t) without any treatment from period 1 to t, and let $Y_{i,g,t}$ denote her observed outcome. Then, we let $Y_{g,t}(\mathbf{0}_t) = 1/N_{g,t} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}(\mathbf{0}_t)$ and $Y_{g,t} = 1/N_{g,t} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}$. With this notation, we can show that the estimators defined in Section 3 are conditionally unbiased for the causal effects defined therein, under Assumptions 3 and 4. For instance, the $(\text{DID}_\ell)_{1\leq\ell\leq L}$ are conditionally unbiased for the $(\delta_\ell)_{1\leq\ell\leq L}$. Thus, our estimators can still be used, even if groups' potential outcomes depend on the identities of treated units.

In groups partly treated at period 1, the estimators proposed in this paper are not applicable. To simplify the discussion, let us momentarily assume that the individual-level treatment is binary. Then, 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, which 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).

1.8 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 17-k *(No effect of past treatments beyond k lags)*

For all (g,t) and all $(d_1,...,d_t)$ in the support of $(D_1,...,D_t)$, $Y_{g,t}(d_1,...,d_t) = Y_{g,t}(d_{t-k},...,d_t)$.

Assumption 17-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 17-k 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 Designs 1-3, our main results still hold if potential outcomes depend on $(D_{g,t})_{t\leq 0}$. In groups 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 17-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 $k+1$.

2 Details on the literature review

Table 1: 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.

3 Proofs

3.1 Proof of Lemma 1

Let g be such that $F_g \leq T_g$. For any $\ell \in \{1, ..., T_g - F_g + 1\}$,

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

The first equality follows from the definitions of F_g and $\text{DID}_{g,\ell}$ and from Assumption 3. The second equality follows from adding and subtracting $Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell})$ and from the definition of $\delta_{g,\ell}$. The third equality follows from Assumption 4 and the definition of $N^g_{F_g-1+\ell}$.

3.2 Proof of Lemma 2

For any $\ell \in \{1, ..., T_g - F_g + 1\}$,

$$\begin{aligned}
\delta^n_{g,\ell} =& \frac{E\left[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell}) - Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell})|\boldsymbol{D}\right]}{\sum_{k=0}^{\ell-1}(D_{g,F_g-1+\ell-k} - D_{g,1})} \\
=& \sum_{k=0}^{\ell-1} E\Big[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, D_{g,F_g-1+\ell-k}, \boldsymbol{D}_{g,1,k}) \\
&- Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, D_{g,1}, \boldsymbol{D}_{g,1,k})|\boldsymbol{D}\Big] \Big/ \left(\sum_{k=0}^{\ell-1}(D_{g,F_g-1+\ell-k} - D_{g,1})\right) \\
=& \sum_{k=0}^{\ell-1} w_{g,\ell,k} s_{g,\ell,k}.
\end{aligned}$$

3.3 Proof of Lemma 3

As groups' treatment only changes once, $|\delta^D_{g,\ell}| = \ell \times (D_{g,F_g} - D_{g,1})$ and $\sum_{g':F_{g'}-1+\ell'\leq T_{g'}} |\delta^D_{g',\ell}| = \ell \times \sum_{g':F_{g'}-1+\ell'\leq T_{g'}}(D_{g',F_{g'}} - D_{g',1})$. Hence,

$$\frac{|\delta^D_{g,\ell}|}{\sum_{g':F_{g'}-1+\ell'\leq T_{g'}} |\delta^D_{g',\ell}|} = \frac{D_{g,F_g} - D_{g,1}}{\sum_{g':F_{g'}-1+\ell'\leq T_{g'}}(D_{g',F_{g'}} - D_{g',1})},$$

which does not vary with $\ell \in \{1, \ldots, \ell'\}$. We call this weight W_g below, so that $\delta^{n,bal}_{\ell,\ell'} = \sum_{g':F_{g'}-1+\ell' \leq T_{g'}} W_g \delta^n_{g,\ell}$.

Point 1. By Lemma 2, $\delta^n_{g,\ell} = \sum_{k=0}^{\ell-1} w_{g,\ell,k} s_{g,k}$. Since $s_{g,k} = s_g$ for some s_g, $\delta^n_{g,\ell} = s_g$. Hence, $\delta^{n,bal}_{\ell,\ell'} = \sum_{g':F_{g'}-1+\ell' \leq T_{g'}} W_g s_g$ does not depend on ℓ. The result follows.

Point 2. Using again $\delta^n_{g,\ell} = \sum_{k=0}^{\ell-1} w_{g,\ell,k} s_{g,k}$, we obtain

$$\delta^n_{g,\ell+1} - \delta^n_{g,\ell} = w_{g,\ell+1,\ell} s_{g,\ell} + \sum_{k=0}^{\ell-1} (w_{g,\ell+1,k} - w_{g,\ell,k}) s_{g,k}.$$

Because groups' treatment only changes once, $w_{g,\ell+1,k} - w_{g,\ell,k} = [1/(\ell+1) - 1/\ell] < 0$. Since $s_{g,k} \geq s_{g,\ell}$ for all $k \leq \ell - 1$ by assumption, we obtain

$$\begin{aligned} \delta^n_{g,\ell+1} - \delta^n_{g,\ell} &\leq s_{g,\ell} \left[w_{g,\ell+1,\ell} + \sum_{k=0}^{\ell-1} (w_{g,\ell+1,k} - w_{g,\ell,k}) \right] \\ &\leq 0, \end{aligned}$$

where the second inequality holds since $\sum_{k=0}^{\ell} w_{g,\ell+1,k} = \sum_{k=0}^{\ell-1} w_{g,\ell,k} = 1$. The result follows using $\delta^{n,bal}_{\ell+1,\ell'} - \delta^{n,bal}_{\ell,\ell'} = \sum_{g':F_{g'}-1+\ell' \leq T_{g'}} W_g (\delta^n_{g,\ell+1} - \delta^n_{g,\ell})$.

3.4 Proof of Lemma 4

We have:

$$\sum_{\ell=1}^{L} w_\ell \delta_\ell = \frac{\sum_{\ell=1}^{L} \sum_{g:F_g-1+\ell \leq T_g} \delta_{g,\ell}}{\sum_{g:F_g \leq T_g} \sum_{\ell=1}^{T_g-F_g+1} (D_{g,F_g-1+\ell} - D_{g,1})} = \frac{\sum_{g:F_g \leq T_g} \sum_{\ell=1}^{T_g-F_g+1} \delta_{g,\ell}}{\sum_{g:F_g \leq T_g} \sum_{\ell=1}^{T_g-F_g+1} (D_{g,F_g-1+\ell} - D_{g,1})} = \delta.$$

The first equality follows from the definitions of w_ℓ, and δ_ℓ, and from Assumption 7. The second equality follows from the fact that by the definitions of F_g, T_g, and L,

$$\{(g,\ell) : 1 \leq \ell \leq L, F_g + \ell - 1 \leq T_g\} = \{(g,\ell) : 1 \leq \ell \leq T_g - F_g + 1, F_g \leq T_g\}.$$

3.5 Proof of Proposition 1

First, by Frisch-Waugh's theorem,

$$\widehat{\beta}_{fe,\ell} = \frac{\sum_{g=1}^{G} \sum_{t=1}^{T} \widehat{\varepsilon}^\ell_{g,t} Y_{g,t}}{\sum_{g=1}^{G} \sum_{t=1}^{T} \widehat{\varepsilon}^\ell_{g,t} I_g \mathbb{1}\{t = F - 1 + \ell\}}, \tag{29}$$

where $\widehat{\varepsilon}^\ell_{g,t}$ is the residual of the regression of $I_g \mathbb{1}\{t = F - 1 + \ell\}$ on group and time fixed effects and $(I_g 1\{t = t'\})_{t' \notin \{F-1, F-1+\ell\}}$. Now, reasoning as in Lemma 1 in de Chaisemartin and

D'Haultfœuille (2020) and using Assumptions 3-4, we obtain that for all g, g' and $t \geq F > t'$,

$$E[Y_{g,t} - Y_{g,t'} - (Y_{g',t} - Y_{g',t'}) \,|\boldsymbol{D}] = I_g \frac{\delta_{g,t-(F-1)}}{I_g} - I_{g'} \frac{\delta_{g',t-(F-1)}}{I_{g'}},$$

with the convention that $0/0 = 0$. Then, by the same reasoning as in the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020), we obtain

$$E\left[\sum_{g=1}^{G}\sum_{t=1}^{T} \widehat{\varepsilon}^{\ell}_{g,t} Y_{g,t} \middle| \boldsymbol{D}\right] = \sum_{g:I_g \neq 0} \sum_{t=F}^{T} \widehat{\varepsilon}^{\ell}_{g,t} I_g \frac{\delta_{g,t-(F-1)}}{I_g}. \tag{30}$$

Now, we compute $\widehat{\varepsilon}^{\ell}_{g,t}$. Let $(\widehat{\alpha}_g)_{g=1...G}$, $(\widehat{\gamma}_t)_{t=1...T}$ and $(\widehat{\delta}_t)_{t \notin \{F-1, F-1+\ell\}}$ denote the coefficients of the regression of $I_g \mathbb{1}\{t = F-1+\ell\}$ on group and time fixed effects and the variables $(I_g \mathbb{1}\{t = t'\})_{t' \notin \{F-1, F-1+\ell\}}$. Then

$$\widehat{\varepsilon}^{\ell}_{g,t} = I_g \mathbb{1}\{t = F-1+\ell\} - \widehat{\alpha}_g - \widehat{\gamma}_t - I_g \mathbb{1}\{t \notin \{F-1, F-1+\ell\}\} \widehat{\delta}_t.$$

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

$$\widehat{\gamma}_t = \bar{I}\left(\mathbb{1}\{t = F-1+\ell\} - \widehat{\delta}_t \mathbb{1}\{t \notin \{F-1, F-1+\ell\}\}\right).$$

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

$$\widehat{\alpha}_g = \frac{I_g - \bar{I}}{T}\left[1 - \sum_{t \notin \{F-1, F-1+\ell\}} \widehat{\delta}_t\right].$$

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

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

with $V_e(I) = (1/G)\sum_{g=1}^{G}(I_g - \bar{I})^2$. This last condition implies that $\widehat{\delta}_t = -1/2$. Hence, $\widehat{\alpha}_g = (I_g - \bar{I})/2$ and

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

As a result, for all $t \neq F-1+\ell$, we obtain $\widehat{\varepsilon}^{\ell}_{g,t} = 0$. Point 2 of the proposition follows by (29) and (30). Moreover, for any $\ell \geq 1$, by what precedes, $\widehat{\varepsilon}^{\ell}_{g,F-1+\ell} = (I_g - \bar{I})/2$. By, again, (29), (30), we get

$$E\left[\widehat{\beta}_{fe,\ell} | \boldsymbol{D}\right] = \frac{\sum_{g:I_g \neq 0} I_g (I_g - \bar{I})(\delta_{g,\ell}/I_g)}{\sum_{g:I_g \neq 0} I_g (I_g - \bar{I})}.$$

Point 1 follows.

3.6 Proof of Proposition 2

Point 1 First, by Frisch-Waugh's theorem,

$$\widehat{\beta}_{lp,\ell} = \frac{\sum_{g=1}^{G}\sum_{t=1}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}Y_{g,t-1+\ell}}{\sum_{g=1}^{G}\sum_{t=1}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}D_{g,t}}. \tag{31}$$

Reasoning as in Lemma 1 in de Chaisemartin and D'Haultfœuille (2020), and using Assumptions 2-4, we obtain, for any $g \neq g'$,

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

Reasoning as in the proof of Theorem 1 in de Chaisemartin and D'Haultfœuille (2020) and using the change of variable $k = t + \ell - F_g$ yields

$$\begin{aligned}&E\left[\sum_{g=1}^{G}\sum_{t=1}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}Y_{g,t-1+\ell}\Big|\boldsymbol{D}\right] \\ &= \sum_{g:F_g\leq T}\sum_{t=1}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}\bar{D}_{g,t+\ell-F_g}\mathbb{1}\left\{t-1+\ell \geq F_g\right\}\frac{\delta_{g,t+\ell-F_g}}{\bar{D}_{g,t+\ell-F_g}} \\ &= \sum_{g:F_g\leq T}\sum_{k=\max(1,1+\ell-F_g)}^{T-F_g+1}\bar{D}_{g,k}\widehat{\varepsilon}^{\ell}_{g,F_g-\ell+k}\frac{\delta_{g,k}}{\bar{D}_{g,k}}.\end{aligned} \tag{32}$$

Under Assumption 2, the denominator of $\widehat{\beta}_{lp,\ell}$ satisfies

$$\sum_{g=1}^{G}\sum_{t=1}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}D_{g,t} = \sum_{g:F_g\leq T}\sum_{t=F_g}^{T-\ell+1}\widehat{\varepsilon}^{\ell}_{g,t}D_{g,t} = \sum_{g:F_g\leq T}\sum_{k=\ell}^{T-F_g+1}D_{g,F_g-\ell+k}\widehat{\varepsilon}^{\ell}_{g,F_g-\ell+k}.$$

Thus, we obtain

$$E[\widehat{\beta}_{lp,\ell}|\boldsymbol{D}] = \sum_{g:F_g\leq T}\sum_{k=1}^{T-F_g+1}\mathbb{1}\left\{\ell-k+1 \leq F_g \leq T-k+1\right\}w^{lp,\ell}_{g,k}\frac{\delta_{g,k}}{\bar{D}_{g,k}},$$

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

$$w^{lp,\ell}_{g,k} = \frac{\bar{D}_{g,k}\widehat{\varepsilon}^{\ell}_{g,F_g-\ell+k}}{\sum_{g':F_{g'}\leq T}\sum_{k'=\ell}^{T-F_{g'}+1}D_{g',F_{g'}-\ell+k'}\widehat{\varepsilon}^{\ell}_{g',F_{g'}-\ell+k'}}.$$

The result follows, after permuting the two summations.

Point 2 First note that $\widehat{\varepsilon}^{\ell}_{g,t} = D_{g,t} - D^{\ell}_{g,\cdot} - D_{\cdot,t} + D^{\ell}_{\cdot,\cdot}$, with $D^{\ell}_{g,\cdot} = 1/(T-\ell+1)\sum_{t=1}^{T-\ell+1} D_{g,t}$, $D_{\cdot,t} = 1/G\sum_{g=1}^{G} D_{g,t}$, and $D^{\ell}_{\cdot,\cdot} = 1/(T-\ell+1)\sum_{t=1}^{T-\ell+1} D_{\cdot,t} = 1/G\sum_{g=1}^{G} D^{\ell}_{g,\cdot}$.

Assume we are in Design 1. Let $g^* = \arg\min_g F_g$ and suppose that $\ell > T - \underline{F} + 1$. Then $D_{g,t} = 0$ for all the cells in the sample on which the regression is estimated, so $\widehat{\beta}_{lp,\ell}$ is not well-defined, a contradiction. Thus, $\ell \leq T - \underline{F} + 1$. Notice also that the denominator of the weights is positive. Then, for $\ell \geq 2$, the sign of $w^{lp,\ell}_{g^*,\ell-1}$ is the same as that of

$$-D^{\ell}_{g^*,\cdot} - D_{\cdot,\underline{F}-1} + D^{\ell}_{\cdot,\cdot} = -D^{\ell}_{g^*,\cdot} + D^{\ell}_{\cdot,\cdot}.$$

Moreover,

$$\begin{aligned} D^{\ell}_{g^*,\cdot} &= \frac{T-\ell-\underline{F}+2}{T-\ell+1}, \\ D^{\ell}_{\cdot,\cdot} &= \frac{\sum_{g=1}^{G}\sum_{t=1}^{T-\ell+1} D_{g,t}}{G(T-\ell+1)} = \frac{\sum_{g=1}^{G}\max(0, T-\ell-F_g+2)}{G(T-\ell+1)}. \end{aligned}$$

By construction, $T-\ell-\underline{F}+2 \geq T-\ell-F_g+2$ for all g. Moreover, $T-\ell-\underline{F}+2 > 0$ since $\ell \leq T-\underline{F}+1$. Because $\widehat{\beta}_{lp,\ell}$ is well-defined, F_g is not constant. Thus, $T-\ell-\underline{F}+2 > T-\ell-F_g+2$ for at least one g. As a result,

$$\sum_{g=1}^{G}\max(0, T-\ell-F_g+2) < \sum_{g=1}^{G}(T-\ell-\underline{F}+2),$$

which implies that $D^{\ell}_{\cdot,\cdot} < D^{\ell}_{g^*,\cdot}$. Hence,

$$\min_{g,k} w^{lp,\ell}_{g,k} \leq w^{lp,\ell}_{g^*,\ell-1} < 0.$$

Next, we show that for all $\ell \in \{2, ..., \underline{F}\}$,

$$\sum_{k=1}^{T-\underline{F}+1} \sum_{g:2-k+1\leq F_g \leq T-k+1} w^{lp,\ell}_{g,k} < 1. \tag{33}$$

In Design 1, $\bar{D}_{g,k} = 1$ for all g and $k \geq 1$, and $D_{g,F_g-\ell+k} = 1$ for all g and $k \geq \ell$. Moreover, the condition $\ell - k + 1 \leq F_g$ is not binding since $\ell \in \{2, ..., \underline{F}\}$. Thus, (33) follows if

$$\sum_{k=1}^{\ell-1} \sum_{g:F_g \leq T-k+1} (D_{g,F_g-\ell+k} - D^{\ell}_{g,\cdot} - D_{\cdot,F_g-\ell+k} + D^{\ell}_{\cdot,\cdot}) < 0. \tag{34}$$

First, note that for $k \in \{1, ..., \ell-1\}$, $D_{g,F_g-\ell+k} = 0$ in Design 1. Then, letting $G^k = \#\{g : F_g \leq T-k+1\}$, we have

$$\begin{aligned} \sum_{g:\ell-k+1\leq F_g\leq T-k+1} (-D^{\ell}_{g,\cdot} + D^{\ell}_{\cdot,\cdot}) &= -\sum_{g:F_g\leq T-k+1} D^{\ell}_{g,\cdot} + \frac{G^k}{G}\sum_{g=1}^{G} D^{\ell}_{g,\cdot} \\ &= \sum_{g:F_g\leq T-k+1} D^{\ell}_{g,\cdot}\left(-1 + \frac{G^k}{G}\right) \leq 0, \end{aligned} \tag{35}$$

where the second equality follows from the fact $D^{\ell}_{g,\cdot} = 0$ for all g such that $F_g - 1 + k > T$, as $T - \ell + 1 < F_g$ for those gs. Moreover, the last inequality is strict if there are never takers, since then $G^1 < G$. Then, (34) and in turn (33) hold.

Now, assume that there are no never takers and let $\overline{F} = \max_{g:F_g \leq T} F_g$. Then,

$$\begin{aligned}\sum_{k=1}^{\ell-1} \sum_{g:F_g \leq T-k+1} -D_{\cdot,F_g-\ell+k} &= \sum_{g:F_g \leq T} \sum_{k=1}^{\min(\ell-1,T-F_g+1)} -D_{\cdot,F_g-\ell+k} \\ &\leq \sum_{k=1}^{\min(\ell-1,T-\overline{F}+1)} -D_{\cdot,\overline{F}-\ell+k} \\ &\leq -D_{\cdot,\overline{F}-\ell+\min(\ell-1,T-\overline{F}+1)} < 0. \end{aligned} \tag{36}$$

The last inequality follows from the fact that if $\min(\ell-1, T-\overline{F}+1) = \ell-1$, $-D_{\cdot,\overline{F}-1} < 0$ as $\underline{F} < \overline{F}$ under Assumption 1 and because there are no never takers. If $\min(\ell-1, T-\overline{F}+1) = T-\overline{F}+1$, $-D_{\cdot,T-\ell+1} < 0$, as we have shown above that $\underline{F} \leq T - \ell + 1$. (34) follows from (35) and (36).

Point 3 In Design 3, $\bar{D}_{g,k} = I_g$ for all g and $k \geq 1$, and $D_{g,F_g-\ell+k} = I_g$ for all g and $k \geq \ell$. Therefore,

$$w^{lp,\ell}_{g,k} = \frac{I_g(D_{g,F_g-\ell+k} - D^{\ell}_{g,\cdot} - D_{\cdot,F_g-\ell+k} + D^{\ell}_{\cdot,\cdot})}{\sum_{k'=\underline{F}}^{T-\underline{F}+1} \sum_{g':F_{g'} \leq T-k'+1} I_{g'}(D_{g',F_{g'}-\ell+k'} - D^{\ell}_{g',\cdot} - D_{\cdot,F_{g'}-\ell+k'} + D^{\ell}_{\cdot,\cdot})}.$$

Remark that the constraint $2 - k + 1 \leq F_g$ does not bind. Then, we obtain

$$\sum_{k=1}^{T-\underline{F}+1} \sum_{g:2-k+1 \leq F_g \leq T-k+1} w^{lp,1}_{g,k} = 1.$$

3.7 Proof of Theorem 1

3.7.1 Consistency of DID$_\ell$

We prove the stronger result that $E[(\mathrm{DID}_\ell - \delta_\ell)^2 | \boldsymbol{D}] \to 0$ a.s. First, $E[\mathrm{DID}_\ell | \boldsymbol{D}] = \delta_\ell$. Second, letting $\Delta Y_g = Y_{g,F_g-1+\ell} - Y_{g,F_g-1}$, we have, for some $C_1 > 0$,

$$\begin{aligned} V(\Delta Y_g | \boldsymbol{D}) &\leq 2\left[E(Y^2_{g,F_g-1+\ell} | \boldsymbol{D}) + E(Y^2_{g,F_g-1} | \boldsymbol{D})\right] \\ &\leq 4 \sup_{g,t} E\left[Y^{2+\delta}_{g,t} | \boldsymbol{D}\right]^{1/(1+\delta/2)} \\ &< C_1. \end{aligned}$$

The first inequality uses $V(A) \leq E[A^2]$ and convexity of $x \mapsto x^2$, the second follows by Jensen's inequality and the last holds by Assumption 10. Now, let

$$\overline{\Delta Y}_{c,g} := \frac{1}{N^g_{F_g-1+\ell}} \sum_{g':F_{g'} > F_g-1+\ell} Y_{g',F_g-1+\ell} - Y_{g',F_g-1}.$$

By the same reasoning as above, $V(\overline{\Delta Y}_{c,g}|\boldsymbol{D}) \leq C_1/N^g_{F_g-1+\ell}$. Also, remark that for all $g \in \mathcal{C}^G_k$,

$$N^g_{F_g-1+\ell} \geq \underline{N}_k := \#\left\{g' \leq G : D_{g',1} = d_k, F_{g'} = \max_{g'':D_{g'',1}=d_k} F_{g''}\right\}.$$

Finally, note that for all (g, g'), $\text{cov}(\Delta Y_g, \overline{\Delta Y}_{c,g'}|\boldsymbol{D}) \geq 0$. Then,

$$\begin{aligned}
V(\text{DID}_\ell|\boldsymbol{D}) &= \frac{1}{N_\ell^2}\Bigg[\sum_{g:F_g\leq T_g-\ell+1} V(\Delta Y_g|\boldsymbol{D}) - 2\sum_{\substack{(g,g'):\\ F_g\leq T_g-\ell+1\\ F_{g'}\leq T_{g'}-\ell+1}} \text{cov}(\Delta Y_g, \overline{\Delta Y}_{c,g'}|\boldsymbol{D}) \\
&\qquad + \sum_{\substack{(g,g'):\\ F_g\leq T_g-\ell+1\\ F_{g'}\leq T_{g'}-\ell+1}} \text{cov}(\overline{\Delta Y}_{c,g}, \overline{\Delta Y}_{c,g'}|\boldsymbol{D})\Bigg] \\
&\leq \frac{1}{N_\ell^2}\left[C_1 N_\ell + \left(\sum_{g:F_g\leq T_g-\ell+1} V(\overline{\Delta Y}_{c,g}|\boldsymbol{D})^{1/2}\right)^2\right] \\
&\leq \frac{C_1}{N_\ell^2}\left[N_\ell + \left(\sum_{k\in\mathcal{K}_\ell} \frac{\#\mathcal{C}^G_k}{\underline{N}_k^{1/2}}\right)^2\right], \qquad (37)
\end{aligned}$$

$$(38)$$

where $\mathcal{K}_\ell = \{k = 1, ..., K : f_k \leq T_k - \ell + 1\}$ and with a slight abuse of notation, we let $T_k = \max_{g:D_{g,1}=d_k} F_g - 1$. Note that the first inequality above follows by the Cauchy-Schwarz inequality. Now, observe that $\{\limsup_G \#\mathcal{C}^G_k = \infty\}$ is a tail event. Thus, by Kolmogorov's zero-one law, we either have $\limsup_G \#\mathcal{C}^G_k < \infty$ a.s., or $\limsup_G \#\mathcal{C}^G_k = \infty$ a.s. In the second case, by Assumption 9, $\#\mathcal{C}^G_k \leq C_2\underline{N}_k$ for some $C_2 > 0$. Thus, whether or not $\limsup_G \#\mathcal{C}^G_k = \infty$, there exists $C_3 > 0$ such that

$$\#\mathcal{C}^{G1/2}_k \leq C_3\underline{N}_k^{1/2}.$$

Combined with (37), this yields

$$\begin{aligned}
V(\text{DID}_\ell|\boldsymbol{D}) &\leq \frac{C_1}{N_\ell^2}\left[N_\ell + C_3^2\left(\sum_{k\in\mathcal{K}_\ell} \#\mathcal{C}^{G1/2}_k\right)^2\right] \\
&\leq \frac{C_1}{N_\ell^2}\left[N_\ell + C_3^2\#\mathcal{K}_\ell\left(\sum_{k\in\mathcal{K}_\ell} \#\mathcal{C}^G_k\right)\right] \\
&\leq \frac{C_1 + C_3^2\#\mathcal{K}_\ell}{N_\ell},
\end{aligned}$$

where the second inequality follows by convexity of $x \mapsto x^2$. Thus, $V(\text{DID}_\ell|\boldsymbol{D}) \to 0$ a.s. The result follows.

3.7.2 Asymptotic normality of DID_ℓ

Let us define

$$\mathcal{T}^G = \sqrt{N_\ell} \frac{\mathrm{DID}_\ell - \delta_\ell}{\left(\frac{1}{N_\ell}\sum_{g=1}^G V(U_{g,\ell}^G|\boldsymbol{D})\right)^{1/2}}.$$

The proof is in four steps. In the first step, we write $\mathcal{T}^G$ as a weighted sum of normalized averages over cohorts. In a second step, we show that if the size of a cohort tends to infinity, the corresponding average satisfies a central limit theorem. In a third step, we show that we can neglect other cohorts. The last step concludes.

First step: $\mathcal{T}^G$ is a weighted sum of normalized averages over cohorts

First, some algebra shows that

$$\mathrm{DID}_\ell - \delta_\ell = \frac{1}{N_\ell}\sum_{g=1}^G W_{g,\ell}^G,$$

where $W_{g,\ell}^G = U_{g,\ell}^G - E[U_{g,\ell}^G|\boldsymbol{D}]$. For all $k \leq K$, let us also define $\mathcal{T}_k^G$ as

$$\mathcal{T}_k^G = \frac{\sum_{g\in\mathcal{C}_k^G} W_{g,\ell}^G}{\left(\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G|\boldsymbol{D})\right)^{1/2}}.$$

Recall that as per our convention, $\mathcal{T}_k^G = 0$ if $\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G|\boldsymbol{D}) = 0$. Then, by construction,

$$\mathcal{T}^G = \sum_{k=1}^K \omega_k^G \mathcal{T}_k^G, \tag{39}$$

where $\omega_k^G = \left[\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G|\boldsymbol{D}) / \sum_{g=1}^G V(W_{g,\ell}^G|\boldsymbol{D})\right]^{1/2}$.

Second step: asymptotic normality of $\mathcal{T}_k^G$ if $\#\mathcal{C}_k = \infty$ a.s.

Fix k such that $\#\mathcal{C}_k = \infty$ almost surely. We show below that

$$\lim_{G\to\infty} \frac{\sum_{g\in\mathcal{C}_k^G} E\left[|W_{g,\ell}^G|^{2+\delta}| \, |\boldsymbol{D}\right]}{\left(\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G \, |\boldsymbol{D})\right)^{1+\delta/2}} = 0 \ \text{ a.s.} \tag{40}$$

By Lyapunov's central limit theorem for triangular arrays, this implies that conditional on $(\boldsymbol{D}_g)_{g\geq 1}$ $\mathcal{T}_k^G \xrightarrow{d} \mathcal{N}(0,1)$ a.s. By definition of $\mathcal{C}_k^G$, $\boldsymbol{\lambda}_{g,\ell}^G$ does not depend on $g \in \mathcal{C}_k^G$; we denote it by $\boldsymbol{\lambda}_{k,\ell}^G$ for all $g \in \mathcal{C}_k^G$. Then, by Cauchy-Schwarz inequality and convexity of $x \mapsto x^{2+\delta}$, for all $g \in \mathcal{C}_k^G$,

$$E\left[|W_{g,\ell}^G|^{2+\delta}| \, |\boldsymbol{D}\right] \leq \|\boldsymbol{\lambda}_{k,\ell}^G\|^{2+\delta} 2^{1+\delta} \sup_{g\in\mathcal{C}_k^G} E[\|\boldsymbol{Y}_g\|^{2+\delta}|\boldsymbol{D}]. \tag{41}$$

Moreover,

$$\begin{aligned}\left(\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G\,|\boldsymbol{D})\right)^{1+\delta/2} &= \left[\sum_{g\in\mathcal{C}_k^G} \boldsymbol{\lambda}_{k,\ell}^G \Sigma_g \boldsymbol{\lambda}_{k,\ell}^{G\,\prime}\right]^{1+\delta/2} \\ &\geq \left[\left(\sum_{g\in\mathcal{C}_k^G} \underline{\rho}(\Sigma_g)\right)\|\boldsymbol{\lambda}_{k,\ell}^G\|^2\right]^{1+\delta/2} \\ &\geq \left[\#\mathcal{C}_k^G\,\underline{\rho}\right]^{1+\delta/2}\|\boldsymbol{\lambda}_{k,\ell}^G\|^{2+\delta}. \end{aligned} \tag{42}$$

for some $\underline{\rho} > 0$, by Assumption 10. Hence,

$$\frac{\sum_{g\in\mathcal{C}_k^G} E\left[|W_{g,\ell}^G|^{2+\delta}|\,|\boldsymbol{D}\right]}{\left(\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G\,|\boldsymbol{D})\right)^{1+\delta/2}} \leq \frac{2^{1+\delta}\sup_{g\in\mathcal{C}_k^G} E[\|\boldsymbol{Y}_g\|^{2+\delta}|\boldsymbol{D}]}{(\#\mathcal{C}_k^G)^{\delta/2}\underline{\rho}^{1+\delta/2}}.$$

Equation (40) follows from Assumption 10 and $\#\mathcal{C}_k^G \to \#\mathcal{C}_k = \infty$ almost surely.

Third step: $\mathcal{T}_k^G = O_p(1)$ and $\omega_k^G \to 0$ a.s. if $\#\mathcal{C}_k$ is a.s. bounded.

We now consider k such that $\#\mathcal{C}_k < \infty$ almost surely. We can exclude the case $\sum_{g\in\mathcal{C}_k^G} V(W_{g,\ell}^G|\boldsymbol{D}) = 0$ because if so, $\mathcal{T}_k^G = 0$ (by convention) and $\omega_k^G = 0$. Otherwise, $\mathcal{T}_k^G = O_p(1)$ follows directly by the Bienaymé–Chebyshev inequality since $E(\mathcal{T}_k^G) = 0$ and $V(\mathcal{T}_k^G) = 1$. Next, we prove that $\omega_k^G \to 0$ a.s. First, by Cauchy-Schwarz inequality and Jensen's inequality, we have, for all $g \in \mathcal{C}_k$,

$$\begin{aligned} V(W_{g,\ell}^G|\boldsymbol{D}) &\leq E\left[(\boldsymbol{\lambda}_{k,\ell}^G \boldsymbol{Y}_g)^2\,|\boldsymbol{D}\right] \\ &\leq \|\boldsymbol{\lambda}_{k,\ell}^G\|^2 E\left[\|\boldsymbol{Y}_g\|^2\,|\boldsymbol{D}\right] \\ &\leq T\|\boldsymbol{\lambda}_{k,\ell}^G\|^2 \sup_{g,t} E[|Y_{g,t}|^{2+\delta}|\boldsymbol{D}]^{1/(1+\delta/2)}. \end{aligned}$$

Hence, by (42) and Assumption 10,

$$\omega_k^G \leq C_4\left(\frac{\#\mathcal{C}_k^G \times \|\boldsymbol{\lambda}_{k,\ell}^G\|^2}{\sum_{k'=1}^K \#\mathcal{C}_{k'}^G \times \|\boldsymbol{\lambda}_{k',\ell}^G\|^2}\right)^{1/2}, \tag{43}$$

for some constant $C_4 > 0$. Next, remark that if $F_g \leq T_g - \ell + 1$, then $\lambda_{g,\ell,F_g-1+\ell}^G = S_g$ and thus $\|\boldsymbol{\lambda}_{g,\ell}^G\|^2 \geq 1$. As a result,

$$\begin{aligned} \sum_{k'=1}^K \#\mathcal{C}_{k'}^G \times \|\boldsymbol{\lambda}_{k',\ell}^G\|^2 &\geq \sum_{g:F_g\leq T_g-\ell+1} \|\boldsymbol{\lambda}_{g,\ell}^G\|^2 \\ &\geq N_\ell \qquad (44) \\ &\to \infty \ \text{ a.s.}, \qquad (45) \end{aligned}$$

where the convergence holds by Assumption 9. Next, by definition of $\lambda^G_{g,\ell,t}$,

$$\|\boldsymbol{\lambda}^G_{k,\ell}\|^2 \leq T\left(1+\max_{t=1,...,T_k}\frac{|N^k_{t,\ell}|}{N^k_t}\right), \tag{46}$$

where, with a slight abuse of notation, we let $N^k_t = |\{g \leq G : D_{g,1} = d_k, F_g > t\}|$ and $N^k_{t,\ell} = \sum_{g'\leq G: D_{g',1}=d_k} S_{g'}\mathbb{1}\{F_{g'} = t-\ell+1\}$.

Now, if $\limsup_G |N^k_{t,\ell}| < \infty$ a.s. for all t, we also have $\limsup_G \max_{t=1,...,T}|N^k_{t,\ell}| < \infty$ a.s., and then, by (46), $\limsup_G \|\boldsymbol{\lambda}^G_{k,\ell}\|^2 < \infty$ a.s. In view of (43) and (45), this implies $\omega^G_k \to 0$ a.s.

Otherwise, fix t such that $P(\limsup_G |N^k_{t,\ell}| = \infty) > 0$. Since $\{\limsup_G |N^k_{t,\ell}| = \infty\}$ is a tail event, by Kolmogorov's zero-one law, $\limsup_G |N^k_{t,\ell}| = \infty$ a.s. By the same reasoning, we actually either have $\limsup_G N^k_{t,\ell} = \infty$ a.s. or $\limsup_G N^k_{t,\ell} = -\infty$ a.s. Consider the first case; the reasoning is the same in the second case. We have

$$\begin{aligned} N^k_{t,\ell} \leq& \#\{g \leq G : D_{g,1} = d_k, S_g = 1, F_g = t-\ell+1\} \\ \leq& \#\{g \leq G : D_{g,1} = d_k, S_g = 1, F_g \leq T_g - \ell + 1\} \\ =& v^G_{d_k,1,\ell}. \end{aligned}$$

The second inequality follows since $t \leq T_k$ and $T_k = T_g$ for all g such that $D_{g,1} = d_k$. The equality stems from the definition of $v^G_{d_k,1,\ell}$ in Assumption 9. Hence, $v^G_{d_k,1,\ell} \to \infty$ a.s. Moreover,

$$N^k_t \geq \#\{g \leq G : D_{g,1} = d_k, F_g = \max_{g':D_{g',1}=d_k} F_{g'}\}.$$

Then, in view of Assumption 9,

$$\limsup_G |N^k_{t,\ell}|/N^k_t < \infty \quad \text{a.s.} \tag{47}$$

This holds for all t such that $P(\limsup_G |N^k_{t,\ell}| = \infty) > 0$. Equation (47) obviously also holds if $P(\limsup_G |N^k_{t,\ell}| = \infty) = 0$. Thus, we obtain $\limsup_G \max_{t=1,...,T}|N^k_{t,\ell}|/N^k_t < \infty$ a.s. Then, again by (43)-(46), $\omega^G_k \to 0$ a.s.

Fourth step: conclusion.

Let $\mathcal{K}_b = \{k \in \{1,...,K\} : \#\mathcal{C}_k < \infty \text{ a.s.}\}$ and $\mathcal{K}_u = \{k \in \{1,...,K\} : \#\mathcal{C}_k = \infty \text{ a.s.}\}$. Because $\{\#\mathcal{C}_k < \infty\}$ is a tail event, by Kolmogorov's zero-one law, with probability one, either $\#\mathcal{C}_k < \infty$ or $\#\mathcal{C}_k = \infty$. Hence, $\{1,...,K\} = \mathcal{K}_b \cup \mathcal{K}_u$ a.s. Now, let $(Z_k)_{k\in\mathcal{K}_b}$ be i.i.d. normal variables, independent of the $(\mathcal{T}^G_k)_{k\in\mathcal{K}_u}$. For all $k = 1,...,K$ let $\tilde{\mathcal{T}}^G_k = \mathcal{T}^G_k$ if $k \in \mathcal{K}_u$, $\tilde{\mathcal{T}}^G_k = Z_k$ otherwise. By (39) and Step 3, a.s.,

$$\mathcal{T}^G = \sum_{k=1}^K \omega^G_k \tilde{\mathcal{T}}^G_k + o_p(1). \tag{48}$$

For all $(k,k') \in \mathcal{K}_u^2$, $k \neq k'$, we have $\mathcal{C}_k \cap \mathcal{C}_{k'} = \emptyset$. Thus, by Assumption 8, the $(\mathcal{T}_k^G)_{k\in\mathcal{K}_u}$ and thus the $(\tilde{\mathcal{T}}_k^G)_{k=1,\dots,K}$, are mutually independent. As a result, by Step 2, conditional on $(\boldsymbol{D}_g)_{g\geq 1}$ and a.s.,

$$(\tilde{\mathcal{T}}_1^G, \dots, \tilde{\mathcal{T}}_K^G)' \xrightarrow{d} \mathcal{N}(0, I_K),$$

where I_K denotes the identity matrix of size K. Moreover, $\sum_{k=1}^K \omega_k^{G\,2} = 1$ and the $(\omega_k^G)_{k=1,\dots,K}$ are non-random conditional on $(\boldsymbol{D}_g)_{g\geq 1}$. Then, by Lemma B.5 in D'Haultfœuille and Tuvaandorj (2022), we obtain, conditional on $(\boldsymbol{D}_g)_{g\geq 1}$ and a.s.,

$$\sum_{k=1}^K \omega_k^G \tilde{\mathcal{T}}_k^G \xrightarrow{d} \mathcal{N}(0,1).$$

The result follows from (48) and Slutski's lemma.

3.7.3 Asymptotic validity of $CI_{1-\alpha}$

Let us define $\overline{U}_k := (1/\#\mathcal{C}_k^G) \sum_{g'\in\mathcal{C}_k^G} U_{g',\ell}^G$ and

$$\overline{\sigma}_{G,\ell}^2 := \frac{1}{N_\ell} \sum_{k=1}^K \sum_{g\in\mathcal{C}_k^G} E\left[\left(U_{g,\ell}^G - E[\overline{U}_k|\boldsymbol{D}]\right)^2 \Big| \boldsymbol{D}\right].$$

Remark that

$$\begin{aligned}
\overline{\sigma}_{G,\ell}^2 &= \frac{1}{N_\ell} \sum_{g=1}^G E\left[U_{g,\ell}^{G\,2}|\boldsymbol{D}\right] - \sum_{k=1}^K \frac{\#\mathcal{C}_k^G}{N_\ell} E[\overline{U}_k|\boldsymbol{D}]^2, \\
\widehat{\sigma}_\ell^2 &= \frac{1}{N_\ell} \sum_{g=1}^G U_{g,\ell}^{G\,2} - \sum_{k=1}^K \frac{\#\mathcal{C}_k^G}{N_\ell} \overline{U}_k^2.
\end{aligned}$$

We first show that

$$\widehat{\sigma}_\ell^2 - \overline{\sigma}_{G,\ell}^2 \xrightarrow{P} 0 \quad \text{a.s.}, \tag{49}$$

by showing that

$$\frac{1}{N_\ell} \sum_{g=1}^G \left(U_{g,\ell}^{G\,2} - E\left[U_{g,\ell}^{G\,2}|\boldsymbol{D}\right]\right) \xrightarrow{P} 0, \tag{50}$$

$$\sum_{k=1}^K \frac{\#\mathcal{C}_k^G}{N_\ell} \left[\overline{U}_k^2 - E[\overline{U}_k|\boldsymbol{D}]^2\right] \xrightarrow{P} 0. \tag{51}$$

To obtain (50), note that by the same reasoning as for (41), we have

$$E\left[|U_{g,\ell}^G|^{2+\delta}|\boldsymbol{D}\right] \leq C_5 \|\boldsymbol{\lambda}_{g,\ell}^G\|^{2+\delta}$$

for some $C_5 > 0$. Moreover,

$$E\left[U_{g,\ell}^{G\,2}|\boldsymbol{D}\right] \geq \underline{\rho} \|\boldsymbol{\lambda}_{g,\ell}^G\|^2.$$

Besides, we established below (46) that $\sup_k \|\boldsymbol{\lambda}_{k,\ell}^G\|$ is bounded. Hence, for some $C_6 > 0$ and a.s.,

$$\frac{\sum_{g=1}^G E\left[|U_{g,\ell}^G|^{2+\delta}|\boldsymbol{D}\right]}{\sum_{g=1}^G E\left[U_{g,\ell}^{G\,2}|\boldsymbol{D}\right]} \leq C_6.$$

Then, by the theorem in Hall (1977), a.s.,

$$\frac{\sum_{g=1}^G U_{g,\ell}^{G\,2}}{\sum_{g=1}^G E[U_{g,\ell}^{G\,2}|\boldsymbol{D}]} - 1 \xrightarrow{P} 0. \tag{52}$$

Next, remark that for all $g \in \mathcal{C}_k^G$,

$$E[U_{g,\ell}^{G\,2}|\boldsymbol{D}] \leq C_7 \|\boldsymbol{\lambda}_{k,\ell}^G\|^2,$$

Moreover, using convexity of $x \mapsto x^2$,

$$\|\boldsymbol{\lambda}_{k,\ell}^G\|^2 \leq 8\left[\mathbb{1}\{f_k \leq T_k - \ell\} + \sum_{t<f_k}\left(\frac{N_{t,\ell}^k}{N_t^k}\right)^2\right].$$

Thus,

$$\sum_{g=1}^G E[U_{g,\ell}^{G\,2}|\boldsymbol{D}] \leq 8C_7\left[N_\ell + \sum_{k=1}^K \#\mathcal{C}_k^G \sum_{t<f_k}\left(\frac{N_{t,\ell}^k}{N_t^k}\right)^2\right].$$

Moreover, we saw in Step 3 above that $\limsup_G N_{t,\ell}^k/N_t^k < \infty$ a.s. Also, by definition, $N_t^k \geq \#\mathcal{C}_k^G$ if $f_k > t$. Thus, there exists $C_8 > 0$ such that a.s.,

$$\sum_{g=1}^G E[U_{g,\ell}^{G\,2}|\boldsymbol{D}] \leq 8C_7\left[N_\ell + C_8\sum_{k=1}^K \sum_{t<f_k} N_{t,\ell}^k\right].$$

Finally, notice that $\sum_{t:\exists k: f_k > t} N_{t,\ell}^k = N_\ell$. Thus,

$$\sum_{k=1}^K \sum_{t<f_k} N_{t,\ell}^k \leq KN_\ell.$$

As a result, a.s.,

$$\frac{1}{N_\ell}\sum_{g=1}^G E[U_{g,\ell}^{G\,2}|\boldsymbol{D}] \leq 8C_7 + KC_8.$$

Multiplying (52) by $(1/N_\ell)\sum_{g=1}^G E[U_{g,\ell}^{G\,2}|\boldsymbol{D}]$, we obtain (50).

Now, to obtain (51), first remark that

$$\overline{U}_k^2 - E[\overline{U}_k|\boldsymbol{D}]^2 = (\overline{U}_k - E[\overline{U}_k|\boldsymbol{D}])(\overline{U}_k - E[\overline{U}_k|\boldsymbol{D}] + 2E[\overline{U}_k|\boldsymbol{D}]). \tag{53}$$

Now, we have, for some $C_9 > 0$,

$$\begin{aligned} E[\overline{U}_k|\boldsymbol{D}] &\leq \sup_{g \in \mathcal{C}_k^G} E[\boldsymbol{\lambda}_k^G \boldsymbol{Y}_g \,|\boldsymbol{D}] \\ &\leq \|\boldsymbol{\lambda}_k^G\| \sup_{g \in \mathcal{C}_k^G} E[\|\boldsymbol{Y}_g\| \,|\boldsymbol{D}] \\ &\leq C_9 \|\boldsymbol{\lambda}_k^G\|, \end{aligned}$$

where the last inequality holds by Jensen's inequality and Assumption 10. Hence, $E[\overline{U}_k|\boldsymbol{D}]$ is bounded a.s. Now, because the $(\boldsymbol{Y}_g)$ are independent over groups, $V(\overline{U}_k|\boldsymbol{D}) \leq C_{10}/\#\mathcal{C}_k^G$ a.s., for some $C_{10} > 0$. By the Bienaymé–Chebyshev inequality, this implies that $\overline{U}_k - E[\overline{U}_k|\boldsymbol{D}] = O_p(1/\#\mathcal{C}_k^G))$ a.s. Together with (53) and a.s. boundedness of $E[\overline{U}_k|\boldsymbol{D}]$, this implies that

$$\overline{U}_k^2 - E[\overline{U}_k|\boldsymbol{D}]^2 = O_p(1/\#\mathcal{C}_k^G) \quad \text{a.s.}$$

As a result,

$$\sum_{k=1}^{K} \frac{\#\mathcal{C}_k^G}{N_\ell} \left[\overline{U}_k^2 - E[\overline{U}_k|\boldsymbol{D}]^2\right] = O_p\left(\frac{1}{N_\ell}\right).$$

Equation (51) follows, since $N_\ell \to \infty$ a.s.

Thus, (49) holds. Now, by definition of conditional expectations,

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

Thus,

$$\overline{\sigma}_{G,\ell}^2 \geq \frac{1}{N_\ell} \sum_{g=1}^{G} V(U_{g,\ell}^G|\boldsymbol{D}). \tag{54}$$

Moreover, if $(\boldsymbol{D}_g, \boldsymbol{Y}_g)_{g \geq 1}$ are i.i.d., $E[\overline{U}_k|\boldsymbol{D}] = E\left[U_{g,\ell}^G|\boldsymbol{D}\right]$ and (54) becomes an equality. Now, combining (42) with (44) yields

$$\frac{1}{N_\ell} \sum_{g=1}^{G} V(U_{g,\ell}^G|\boldsymbol{D}) \geq \underline{\rho}.$$

Hence, in view of (49) and (54), $\widehat{\sigma}_\ell^2 > \underline{\rho}/2 > 0$ with probability approaching one. Under this event, we get

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

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

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

By Slutski's lemma and asymptotic normality of DID_ℓ, the term into brackets converges in distribution (and almost surely) to $Z \sim \mathcal{N}(0,1)$. Then, using (54), we obtain

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

Finally, if $(\boldsymbol{D}_g, \boldsymbol{Y}_g)_{g\geq 1}$ are i.i.d., (54) becomes an equality and thus the inequality above also turns into an equality. This completes the proof.

3.8 Proof of Lemma 5

We have:

$$\begin{aligned}E\left[\text{DID}_{g,\ell}^{\text{pl}}\middle|\boldsymbol{D}\right] =&E\left[Y_{g,F_g-1-\ell}(\mathbf{0}_{F_g-1-\ell}) - Y_{g,F_g-1}(\mathbf{0}_{F_g-1})\middle|\boldsymbol{D}\right]\\ &- E\left[\frac{1}{N_{F_g-1+\ell}^g}\sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell}(Y_{g',F_g-1-\ell}(\mathbf{0}_{F_g-1-\ell}) - Y_{g',F_g-1}(\mathbf{0}_{F_g-1})\middle|\boldsymbol{D}\right]\\ =&0. \end{aligned} \tag{55}$$

The first equality follows from the definition of F_g and Assumption 3. The second equality follows from Assumption 4 and the definition of $N_{F_g-1+\ell}^g$. Then, the result follows from the definition of $\text{DID}_\ell^{\text{pl}}$, the fact that $(F_g, S_g, T_g)_{1\leq g\leq G}$, and $(N_\ell^{\text{pl}})_{1\leq \ell\leq L^{\text{pl}}}$ are functions of $\boldsymbol{D}$, the linearity of the conditional expectation operator, and Equation (55).

3.9 Proof of Lemma 6

We prove the result for $g : T_g - F_g + 1 \geq 2$ and $\ell \in \{2, ..., T_g - F_g + 1\}$, the proof for $\ell = 1$ is similar and simpler.

$$\begin{aligned}&E\left[\text{DID}_{g,\ell}^{fd}\middle|\boldsymbol{D}\right]\\ =&E\left[Y_{g,F_g-1+\ell} - Y_{g,F_g-1+\ell-1} - (Y_{g,F_g-1}(\boldsymbol{D}_{g,1,F_g-1}) - Y_{g,F_g-2}(\boldsymbol{D}_{g,1,F_g-2}))\middle|\boldsymbol{D}\right]\\ -&\frac{1}{N_{F_g-1+\ell}^g}E\Bigg[\sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell}(Y_{g',F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell}) - Y_{g',F_g-1+\ell-1}(\boldsymbol{D}_{g,1,F_g-1+\ell-1})\\ &\qquad - (Y_{g',F_g-1}(\boldsymbol{D}_{g,1,F_g-1}) - Y_{g',F_g-2}(\boldsymbol{D}_{g,1,F_g-2})))\middle|\boldsymbol{D}\Bigg]\end{aligned}$$

$$
\begin{aligned}
&=\delta_{g,\ell} - \delta_{g,\ell-1} \\
&+E\left[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell}) - Y_{g,F_g-1+\ell-1}(\boldsymbol{D}_{g,1,F_g-1+\ell-1}) - (Y_{g,F_g-1}(\boldsymbol{D}_{g,1,F_g-1}) - Y_{g,F_g-2}(\boldsymbol{D}_{g,1,F_g-2}))\middle|\boldsymbol{D}\right] \\
&-\frac{1}{N^g_{F_g-1+\ell}}E\Bigg[\sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell}(Y_{g',F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell}) - Y_{g',F_g-1+\ell-1}(\boldsymbol{D}_{g,1,F_g-1+\ell-1}) \\
&\qquad\qquad - (Y_{g',F_g-1}(\boldsymbol{D}_{g,1,F_g-1}) - Y_{g',F_g-2}(\boldsymbol{D}_{g,1,F_g-2})))\Big|\boldsymbol{D}\Bigg] \\
&=\delta_{g,\ell} - \delta_{g,\ell-1}.
\end{aligned}
$$

The first equality follows from the definitions of F_g and $\text{DID}^{fd}_{g,\ell}$ and from Assumption 3. The second equality follows from adding and subtracting $Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell})-Y_{g,F_g-1+\ell-1}(\boldsymbol{D}_{g,1,F_g-1+\ell-1})$ and from the definition of $\delta_{g,\ell}$. The third equality follows from Assumption 12 and the definition of $N^g_{F_g-1+\ell}$.

3.10 Proof of Lemma 7

Let $\tilde{X}_g$ denote the residual from a regression of X_g on indicators for all possible values of $F_g \times D_{g,1} \times S_g$ in the sample of groups such that $F_g - 1 + \ell \leq T_g$. The regression is saturated, so $\tilde{X}_g$ is also a conditional expectation residual. It follows from the Frisch-Waugh theorem that

$$
\widehat{\beta}^{het}_{\ell} = \left(\sum_{g:F_g-1+\ell\leq T_g}\tilde{X}_g\tilde{X}_g'\right)^{-1}\sum_{g:F_g-1+\ell\leq T_g}S_g(Y_{g,F_g-1+\ell} - Y_{g,F_g-1})\tilde{X}_g.
$$

Then we have, for some scalars $(\gamma_{F_g,D_{g,1}})_{g:F_g-1+\ell\leq T_g}$,

$$
\begin{aligned}
&E\left[\sum_{g:F_g-1+\ell\leq T_g}S_g(Y_{g,F_g-1+\ell} - Y_{g,F_g-1})\tilde{X}_g\middle|\boldsymbol{D},\boldsymbol{X}\right] \\
&=\sum_{g:F_g-1+\ell\leq T_g}S_gE\left[Y_{g,F_g-1+\ell} - Y_{g,F_g-1}\middle|\boldsymbol{D},\boldsymbol{X}\right]\tilde{X}_g \\
&=\sum_{g:F_g-1+\ell\leq T_g}S_g\delta_{g,\ell}\tilde{X}_g + \sum_{g:F_g-1+\ell\leq T_g}S_gE\left[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1+\ell}) - Y_{g,F_g-1}(\boldsymbol{D}_{g,1,F_g-1})\middle|\boldsymbol{D},\boldsymbol{X}\right]\tilde{X}_g \\
&=\sum_{g:F_g-1+\ell\leq T_g}\delta_{g,\ell}\tilde{X}_g + \sum_{g:F_g-1+\ell\leq T_g}S_g\gamma_{F_g,D_{g,1}}\tilde{X}_g \\
&=\sum_{g:F_g-1+\ell\leq T_g}\delta_{g,\ell}\tilde{X}_g.
\end{aligned}
$$

The third equality follows from Assumption 15, the fourth follows from the fact $\tilde{X}_g$ is uncorrelated to any function of $(F_g, D_{g,1}, S_g)$. The result follows from the two previous displays and the definition of β^{het}_{ℓ}.

References

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

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica 66*(2), 249–288.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, Volume 3, pp. 1277–1366. Elsevier.

Antecol, H., K. Bedard, and J. Stearns (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., M. Couttenier, D. Rohner, and M. Thoenig (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review 107*(6), 1564–1610.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American economic review 107*(8), 2204–42.

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

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

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

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

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

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

de Chaisemartin, C. and X. D'Haultfœuille (2023). Two-way fixed effects and difference-in-differences estimators with heterogeneous treatment effects and imperfect parallel trends. SSRN Working paper.

de Chaisemartin, C., X. D'Haultfœuille, F. Pasquier, and G. Vazquez-Bare (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.

D'Haultfœuille, X. and P. Tuvaandorj (2022). A robust permutation test for subvector inference in linear regressions. arXiv preprint arXiv:2205.06713.

Di Maggio, M., A. Kermani, B. J. Keys, T. Piskorski, R. Ramcharan, A. Seru, and V. Yao (2017). Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review 107*(11), 3550–88.

Diamond, R., T. McQuade, and F. Qian (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 B. K. Kovak (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 J. Imbs (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., A. Peichl, and S. Siegloch (2018). Do higher corporate taxes reduce wages? micro evidence from germany. *American Economic Review 108*(2), 393–418.

Hall, P. (1977). On the lp convergence of sums of independent random variables. *Mathematical Proceedings of the Cambridge Philosophical Society 82*(3), 439–446.

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

Hoynes, H., D. W. Schanzenbach, and D. Almond (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., S. J. Redding, and E. Rossi-Hansberg (2018). Commuting, migration, and local employment elasticities. *American Economic Review 108*(12), 3855–90.

Mora, R. and I. Reggio (2019). Alternative diff-in-diffs estimators with several pretreatment periods. *Econometric Reviews 38*(5), 465–486.

Munshi, K. and M. Rosenzweig (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 P. K. Schott (2016). The surprisingly swift decline of us manufacturing employment. *American Economic Review 106*(7), 1632–62.

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

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

Shahn, Z. (2023). Subgroup difference in differences to identify effect modification without a control group. arXiv preprint arXiv:2306.11030.

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