Difference-in-Differences Estimators of Intertemporal Treatment Effects

Clément de Chaisemartin* Xavier D'Haultfœuille†

August 11, 2020

Abstract

We consider the estimation of the effect of a policy or treatment, using panel data where different groups of units are exposed to the treatment at different times. We focus on parameters aggregating instantaneous and dynamic treatment effects, with a clear welfare interpretation. We show that under parallel trends conditions, these parameters can be unbiasedly estimated by a weighted average of differences-in-differences, provided that at least one group is always untreated, and another group is always treated. Our estimators are valid if the treatment effect is heterogeneous, contrary to the commonly-used event-study regression.

Keywords: differences-in-differences, panel data, repeated cross-section data, dynamic treatment effects, welfare analysis, event-study regression.

JEL Codes: C21, C23

*University of California at Santa Barbara, clementdechaisemartin@ucsb.edu

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

1 Introduction

We consider the estimation of the effect of a policy or treatment on an outcome, using a panel of groups (indexed by g hereafter) that are exposed to the policy at different times (indexed by t hereafter). It is often appealing to study the dynamic effects of the policy, rather than just focusing on its instantaneous effect at the time of implementation. To do so, a commonly-used method, first proposed by Autor (2003), is to regress the outcome on group fixed effects, time fixed effects, the value of the treatment in group g and period t, and lags of the treatment in group g. Hereafter, we refer to this regression as the event-study regression. Intuitively, the coefficient of the contemporaneous treatment should estimate its instantaneous effect, while the coefficients of the lagged treatments should estimate its dynamic effects. However, Abraham and Sun (2020) have shown that those estimators are only valid if the treatment effect is homogenous over groups and time, the latter assumption being especially unappealing when one is interested in estimating dynamic effects.[1] Instead, we propose to use differences-in-differences (DID) estimators. As the event-study regression, such estimators rely on the standard parallel trends assumption. But unlike it, they remain valid if the treatment effect is heterogeneous.

In our panel data setting, there is a wealth of instantaneous and dynamic treatment effects one could estimate, and some aggregation is in order to improve power. Welfare analysis is a natural guide to perform said aggregation (see Manski, 2005). Specifically, we assume units' utility is additively separable in the outcome, and adopt the perspective of a planner interested in comparing the population's expected average intertemporal utility under the actual treatments received and under the scenario where all groups keep all along the same treatment as in the first period of the panel. Our parameter of interest is the difference between these two expectations, hereafter referred to as the actual-versus-status-quo parameter. It measures the welfare effects of the policy changes that occurred over the period.

We start by focusing on staggered adoption designs, where groups adopt the treatment at different time periods and cannot switch out of the treatment after adoption. Many applications do not fall into this special case, but it is much simpler to analyze than the general case, so we focus on it first for the sake of exposition. We show that if at least one group is still untreated at the end of the panel, the actual-versus-status-quo parameter can be unbiasedly estimated.

Our estimator proceeds in two steps. We start by estimating the average effect of having started to receive the treatment ℓ periods ago, from $\ell = 0$ (corresponding to the instantaneous treatment effect) to the highest value of ℓ observed in the data. For each ℓ, our estimator DID_ℓ is a weighted average, across t, of DID estimators comparing the $t - \ell - 1$ to t outcome evolution, in groups that become treated in $t - \ell$ and in groups not yet treated in t. Then, our estimator of the

[1] Their result is a generalization of that in de Chaisemartin and D'Haultfœuille (2020), who show that a similar result holds for the static two-way fixed effects regression without the lagged treatments.

actual-versus-status-quo parameter is a weighted average of the DID_ℓ estimators. If all groups are treated at the end of the panel, a truncated version of our actual-versus-status-quo parameter can still be unbiasedly estimated, where the truncation happens at the last period when at least one group is still untreated. Finally, we show how to test for the parallel trends condition underlying our estimators, using placebo estimators comparing the outcome evolution between the same groups as above, before groups that eventually become treated do so. In staggered adoption designs, the DID_ℓ and placebo estimators are computed by the Stata `did_multiplegt` package (see de Chaisemartin et al., 2019).

We then consider general designs, where groups may switch in and out of the treatment at any time. We start by showing that if there is at least one group that is always untreated and another group that is always treated, the actual-versus-status-quo parameter can be unbiasedly estimated. Our estimator is a weighted average, across t and ℓ, of DID estimators comparing the $t-\ell-1$ to t outcome evolution, in groups whose treatment first changed in $t-\ell$ and in groups whose treatment has not changed yet in t. Again, if there is no group untreated till the end of the panel, or no group treated till the end of the panel, a truncated version of our parameter can be unbiasedly estimated. We also propose placebo estimators to test the parallel trends conditions underlying our estimators.

In long panels, or in instances where groups change treatment frequently, the truncation may happen early in the panel, and the truncated parameter could then be very different from the original target parameter. Then, we consider an assumption on dynamic effects, and show that under this assumption, one can unbiasedly estimate a parameter that will often be closer to the parameter the planner needs to know to evaluate the policy changes that took place over the period. This assumption amounts to ruling out dynamic effects beyond k-lags, for a given k that the analyst should choose based on the context. Ruling out dynamic effects beyond a certain lag is an assumption that is implicitly made in event-study regressions since without it, such regressions are not identified (Borusyak and Jaravel, 2017; Schmidheiny and Siegloch, 2020).

Abraham and Sun (2020) and Callaway and Sant'Anna (2018) have also proposed DID estimators of instantaneous and dynamic treatment effects in panels with multiple groups and periods. Our paper differs from those on at least three dimensions. First and foremost, those papers only consider staggered adoption designs, while we also consider general designs, where groups may switch in and out of the treatment at any time. Our estimators can also be easily extended to non-binary treatments, unlike theirs. In a survey of all the papers using regressions with group and time fixed effects published by the AER between 2010 and 2012, de Chaisemartin and D'Haultfœuille (2020) find that less than 10% have a staggered adoption design. Therefore, our estimators can be used in a much larger set of empirical applications. Second, our paper is the first to use welfare analysis to guide the aggregation of instantaneous and dynamic treatment effects. Callaway and Sant'Anna (2018) propose several interesting aggregation methods, but

the estimands they propose differ from our actual-versus-status-quo parameter, and they do not have a clear welfare interpretation. The same applies to the aggregation method in Abraham and Sun (2020). Aggregation is especially important outside of staggered designs. With T time periods in the panel, there are only $T+1$ possible treatment trajectories in a staggered design, against 2^T in general designs. Hence, we can expect the estimation of the difference between the outcome of two such trajectories to be very noisy. Third, even in staggered designs, our estimators differ from those in Abraham and Sun (2020) and Callaway and Sant'Anna (2018). To estimate the treatment effect at date t in groups that became treated at date $t-\ell$, we use as controls all groups not yet treated at t, while Callaway and Sant'Anna (2018) use the never treated groups, and Abraham and Sun (2020) use the never treated groups or the groups that become treated last if there are no never treated groups. Our control group is larger, so we can expect our estimators to be more precise.

This paper is also related to previous work of ours. In staggered designs, the DID_0 estimator we propose in this paper is equivalent to the DID_M estimator we proposed in de Chaisemartin and D'Haultfœuille (2020). Outside of staggered designs, the DID_M estimator is equivalent to the estimator we propose in this paper under the assumption of no dynamic effects.

This paper is organized as follows. Section 2 introduces the notation and the assumptions we maintain throughout the paper. Section 3 considers the case of staggered adoption designs. Section 4 considers general designs.

2 Set-up

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

One is interested in measuring the effect of a treatment on some outcome. Throughout the paper we assume that treatment is binary, but our results can easily be generalized to any ordered treatment. Then, for every $(i,g,t) \in \{1,...,N_{g,t}\} \times \{1,...,G\} \times \{1,...,T\}$, let $D_{i,g,t}$ denote the treatment status of observation i in group g at period t. We focus on sharp designs, where the treatment does not vary within (g,t) cells.

Assumption 1 *(Sharp design)* $\forall (i,g,t) \in \{1,...,N_{g,t}\} \times \{1,...,G\} \times \{1,...,T\}$, $D_{i,g,t} = D_{g,t}$.[2]

[2]Assumptions 1, 3, 4, 6, 7, 9-k, and Theorems 1 and 2 have equalities and inequalities involving random variables. Implicitly, these equalities and inequalities are assumed to hold with probability 1.

Assumption 1 is for instance satisfied when the treatment is a group-level variable, for instance a county- or a state-law, or when $N_{g,t} = 1$. Then, let $\boldsymbol{D}_g = (D_{g,1}, ..., D_{g,T})$ be a $1 \times T$ vector stacking the treatments of group g from period 1 to T. For all t, let $\boldsymbol{D}_{g,t} = (D_{g,1}, ..., D_{g,t})$ be a $1 \times t$ vector stacking the treatments of group g from period 1 to t.

For all $\boldsymbol{d} \in \{0,1\}^T$, let $Y_{g,t}(\boldsymbol{d})$ denote the average potential outcome of group g at period t, if the treatments of group g from period 1 to T are equal to $\boldsymbol{d}$. This notation allows for the possibility that group g's outcome at time t be affected by her past and future treatments. Some groups may have already been treated prior to period 1, the first period in the data, and those treatments may still affect some of their period-1-to-T outcomes. However, we cannot estimate such dynamic effects, as treatments and outcomes are not observed for those periods, so we do not account for this potential dependency in our notation. Finally, we let $Y_{g,t} = Y_{g,t}(\boldsymbol{D}_g)$ denote the observed outcome, and for all t we let $\mathbf{0}_t$ denote a vector of t zeros.

Assumption 2 *(Independence between groups) The G vectors $(\boldsymbol{D}_g, (Y_{g,1}(\boldsymbol{d}), ..., Y_{g,T}(\boldsymbol{d}))_{\boldsymbol{d} \in \{0,1\}^T})$ are mutually independent.*

We consider the treatment and potential outcomes of each (g,t) cell as random variables. For instance, aggregate random shocks may affect the potential outcomes of group g at period t, and that cell's treatment may also be random. The expectations below are taken with respect to the distribution of those random variables. Under Assumption 2, the treatments and potential outcomes of a group may be correlated over time, but the potential outcomes and treatments of different groups have to be independent, a commonly made assumption in difference-in-differences (DID) designs (see Bertrand et al., 2004).

Assumption 3 *(No Anticipation) For all g, for all $\boldsymbol{d} \in \{0,1\}^T$, $Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(d_1, ..., d_t)$.*

Assumption 3 requires that a group's current outcome do not depend on her future treatments, the so-called no-anticipation hypothesis, see, e.g., Abbring and Van den Berg (2003). Under Assumption 3, $Y_{g,t} = Y_{g,t}(\boldsymbol{D}_{g,t})$.

Hereafter, we refer to $Y_{g,t}(\mathbf{0}_t)$, the potential outcome that group g will obtain at period t if she remains untreated from period 1 to t as the never-treated potential outcome. We consider two assumptions on that outcome.

Assumption 4 *(Strong exogeneity) $\forall t \geq 2$, $E(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1}) | \boldsymbol{D}_g) = E(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1}))$.*

Assumption 4 requires that the shocks affecting a group's never-treated potential outcome be mean independent of her treatments. For instance, this rules out cases where a group gets treated because it experiences some negative shocks, the so-called Ashenfelter's dip (see Ashenfelter, 1978). Assumption 4 is related to the strong exogeneity condition in panel data models.

Assumption 5 *(Common trends) $\forall t \geq 2$, $E(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1}))$ does not vary across g.*

Assumption 5 requires that in every group, the expectation of the never-treated outcome follow the same evolution over time. It is a generalization of the standard parallel trends assumption in DID models (see, e.g., Abadie, 2005) to our set-up allowing for dynamic effects. Abraham and Sun (2020), Athey and Imbens (2018), and Callaway and Sant'Anna (2018) also consider that assumption.

3 Staggered adoption designs

Throughout this section, we assume that the treatment follows a staggered adoption design, where groups can switch in but not out of the treatment.

Assumption 6 *(Staggered adoption designs) For all g and $t \geq 2$, $D_{g,t} \geq D_{g,t-1}$.*

For any $g \in \{1, ..., G\}$, let

$$F_g = \min\{t : D_{g,t} = 1\}$$

denote the first date at which group g is treated, with the convention that $F_g = T + 1$ if group g is never treated.

Our parameters of interest are motivated by a welfare analysis. The treatment $D_{g,t}$ may for instance correspond to a policy that is costly to implement, and a planner may seek to compare the welfare gains produced by the policy to its cost (see Manski, 2005). We assume that for any possible value $\boldsymbol{d}_t$ of the treatments up to period t, the average utility in group g at period t is equal to $Y_{g,t}(\boldsymbol{d}_t) + U_{g,t}$. If the social planner wants to compare the population's expected average intertemporal utility under the actual treatments received by all groups and under the scenario where all groups are never treated, then she would like to learn

$$\delta_{ATT} = E\left(\sum_{g:F_g \leq T} \sum_{t=F_g}^{T} \frac{N_{g,t}}{N_D} \beta^t (Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t))\right),$$

where β is the planner's discount factor, and $N_D = \sum_{g:F_g \leq T} \sum_{t=F_g}^{T} N_{g,t}$ is the number of units in the treated (g,t) cells. With $\beta = 1$ and assuming no dynamic effects ($Y_{g,t}(d_1, ..., d_t) = Y_{g,t}(d_t)$), δ_{ATT} is equal to the standard average treatment effect on the treated (ATT) parameter. Thus, δ_{ATT} generalizes the ATT to settings with dynamic treatment effects, and it allows for the possibility that the planner may discount later periods relative to earlier ones.

If the social planner wants to compare the population's utility under the actual treatments received and under the scenario where all groups keep the same treatment as in the first period (the status quo scenario), then she would like to learn

$$\delta_{SQ} = E\left(\sum_{g:2 \leq F_g \leq T} \sum_{t=F_g}^{T} \frac{N_{g,t}}{N_S} \beta^t (Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t))\right),$$

where $N_S = \sum_{g:2\leq F_g\leq T}\sum_{t=F_g}^{T} N_{g,t}$ is the number of units in the treated (g,t) cells, excluding the always treated groups. If there are no always treated groups ($F_g > 1$ for all g), $\delta_{ATT} = \delta_{SQ}$. If there are always treated groups, not taking their treatment effect into account may be justified, if the planner wants to evaluate the welfare effects of the policy decisions that took place over the period under consideration, and not of earlier policy decisions.

Before proposing an estimator of δ_{SQ}, we consider parameters that are simpler to estimate. For any $\ell \in \{0,...,T-2\}$, let $N_\ell = \sum_{g:2\leq F_g\leq T-\ell} N_{g,F_g+\ell}$ denote the number of units in the (g,t) cells such that at period t, group g has started receiving the treatment ℓ periods ago. Let

$$\Delta_\ell = \sum_{g:2\leq F_g\leq T-\ell} \frac{N_{g,F_g+l}}{N_\ell}\beta^{F_g+\ell}(Y_{g,F_g+\ell}(\boldsymbol{D}_{g,F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell})) \tag{1}$$

if $N_\ell > 0$, and let $\Delta_\ell = 0$ otherwise. If $\beta = 1$, Δ_ℓ is just the average effect of having been treated for ℓ periods, across all the groups treated for ℓ periods before period T, and excluding the always treated groups. If $\beta < 1$, Δ_ℓ has the same interpretation, except that groups' treatment effect is discounted according to the date when they reach ℓ periods of treatment. Let $L = T - \min_{g:F_g\geq 2} F_g$ denote the number of time periods between the earliest date at which a group goes from untreated to treated and date T. $N_\ell = 0$ if and only if $\ell > L$: if $\ell > L$, the effect of being treated for ℓ periods is not observed for any group. Finally, let $\delta_\ell = E(\Delta_\ell)$.

For any $\ell \in \{0,...,T-2\}$ and $t \in \{\ell+2,...,T\}$, let $N_t^\ell = \sum_{g:F_g=t-\ell} N_{g,t}$ and $N_t^{nt} = \sum_{g:F_g>t} N_{g,t}$. Let

$$\text{DID}_{t,\ell} = \beta^t\left(\sum_{g:F_g=t-\ell}\frac{N_{g,t}}{N_t^\ell}(Y_{g,t} - Y_{g,t-\ell-1}) - \sum_{g:F_g>t}\frac{N_{g,t}}{N_t^{nt}}(Y_{g,t} - Y_{g,t-\ell-1})\right)$$

if $N_t^\ell > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{t,\ell} = 0$ otherwise. $\text{DID}_{t,\ell}$ is the β^t-discounted DID estimator comparing the outcome evolution from period $t-\ell-1$ to t in groups that became treated in $t-\ell$ and in groups still untreated in t. Under Assumptions 4-5, the expectation of the outcome evolution in the latter set of groups is a counterfactual of the evolution that would have taken place in the former set of groups if it had not started receiving the treatment ℓ periods ago. Thus, $\text{DID}_{t,\ell}$ is an unbiased estimator of the effect of having been treated for ℓ periods in those groups. Then, let $N_{\text{DID}_\ell} = \sum_{t\geq\ell+2, N_t^{nt}>0} N_t^\ell$ be the number of units in (g,t) cells such that group g has been treated for ℓ periods at date t, and another group is untreated at t. Let

$$\text{DID}_\ell = \sum_{t=\ell+2}^{T}\frac{N_t^\ell}{N_{\text{DID}_\ell}}\text{DID}_{t,\ell}$$

if $N_{\text{DID}_\ell} > 0$, and let $\text{DID}_\ell = 0$ otherwise. DID_ℓ is our estimator of δ_ℓ. Under Assumption 6, DID_0 with $\beta = 1$ is equal to the DID_M estimator in de Chaisemartin and D'Haultfœuille (2020).The DID_ℓ estimators with $\beta = 1$ are computed by the Stata `did_multiplegt` package.

We have the following relationship between δ_{SQ} and the $(\Delta_\ell)_{0\leq\ell\leq L}$:

$$\begin{aligned} E\left(\sum_{\ell=0}^{L}\frac{N_\ell}{N_S}\Delta_\ell\right) =& E\left(\sum_{\ell=0}^{L}\sum_{g:2\leq F_g\leq T-\ell}\frac{N_{g,F_g+\ell}}{N_S}\beta^{F_g+\ell}(Y_{g,F_g+\ell}(\boldsymbol{D}_{g,F_g+\ell})-Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell}))\right) \\ =& E\left(\sum_{g:2\leq F_g\leq T}\sum_{\ell=0}^{T-F_g}\frac{N_{g,F_g+\ell}}{N_S}\beta^{F_g+\ell}(Y_{g,F_g+\ell}(\boldsymbol{D}_{g,F_g+\ell})-Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell}))\right) \\ =& \delta_{SQ}. \end{aligned} \tag{2}$$

Accordingly, we estimate δ_{SQ} by

$$\widehat{\delta}_{SQ}=\sum_{\ell=0}^{L}\frac{N_{\text{DID}_\ell}}{N_S}\text{DID}_\ell.$$

The following theorem gives conditions under which DID_ℓ and $\widehat{\delta}_{SQ}$ are unbiased estimators of δ_ℓ and δ_{SQ}.

Theorem 1 *Suppose that Assumptions 1-6 hold and $N_S>0$.*

1. *For any $\ell\in\{0,...,T-2\}$, if $\max_{g\in\{1,...,G\}}F_g>\max_{g:2\leq F_g\leq T-\ell}F_g+\ell$, $E\left[DID_\ell\right]=\delta_\ell$.*
2. *If $\max_{g\in\{1,...,G\}}F_g=T+1$, $E\left[\widehat{\delta}_{SQ}\right]=\delta_{SQ}$.*

The condition $\max_{g\in\{1,...,G\}}F_g>\max_{g:2\leq F_g\leq T-\ell}F_g+\ell$ requires that at the last time period when a group reaches ℓ periods of treatment, there is still at least one untreated group. This condition can be tested from the data. Point 1 of Theorem 1 shows that under this condition, the average effect of having been treated for ℓ periods δ_ℓ can be unbiasedly estimated. Then, Point 2 of the theorem shows that δ_{SQ} can also be unbiasedly estimated, provided there is at least one group that is still untreated at date T. When no group is treated at the start of the panel, δ_{SQ} and δ_{ATT} are equal, so δ_{ATT} can also be unbiasedly estimated.

Theorem 1 applies to designs where at least one group is still untreated at the end of the panel. We now consider cases where that condition is not met. Let $NT=\max_{g\in\{1,...,G\}}F_g-1$ denote the last period where at least one group is still untreated. Let $N_S^{trun}=\sum_{g:2\leq F_g\leq NT}\sum_{t=F_g}^{NT}N_{g,t}$ be the number of units in the (g,t) cells that are treated and such that $t\leq NT$, excluding the always treated groups. Let

$$\delta_{SQ}^{trun}=E\left(\sum_{g:2\leq F_g\leq NT}\sum_{t=F_g}^{NT}\frac{N_{g,t}}{N_S^{trun}}\beta^t(Y_{g,t}(\boldsymbol{D}_{g,t})-Y_{g,t}(\mathbf{0}_t))\right)$$

denote the truncated-at-NT version of δ_{SQ}, that only takes into account the treatment effects until period NT. For any $\ell\in\{0,...,T-2\}$, let $N_\ell^{trun}=\sum_{g:2\leq F_g\leq NT-\ell}N_{g,F_g+\ell}$ denote the number

of units in (g,t) cells such that at period t, group g has started receiving the treatment ℓ periods ago and there is still an untreated group, excluding the always treated groups. Let

$$\Delta_\ell^{trun} = \sum_{g:2\leq F_g\leq NT-\ell} \frac{N_{g,F_g+l}}{N_\ell^{trun}} \beta^{F_g+\ell}(Y_{g,F_g+\ell}(\boldsymbol{D}_{g,F_g+\ell}) - Y_{g,F_g+\ell}(\mathbf{0}_{F_g+\ell}))$$

if $N_\ell^{trun}>0$, and let $\Delta_\ell^{trun}=0$ otherwise. Δ_ℓ^{trun} is the truncated-at-NT version of Δ_ℓ, the effect of being treated for ℓ periods. Let δ_ℓ^{trun} denote its expectation. Using the same steps as those use to prove Equation (2), one can show that

$$E\left(\sum_{\ell=0}^{L} \frac{N_\ell^{trun}}{N_S^{trun}} \Delta_\ell^{trun}\right) = \delta_{SQ}^{trun}. \tag{3}$$

Then, we let $\widehat{\delta}_{SQ}^{trun} = \sum_{\ell=0}^{L}(N_{\text{DID}_\ell}/N_S^{trun})\text{DID}_\ell$. The $\widehat{\delta}_{SQ}^{trun}$ estimator with $\beta=1$ is computed by the Stata `did_multiplegt` package.

Theorem 2 *Suppose that Assumptions 1-6 hold and $N_S^{trun}>0$.*

1. *For any $\ell\in\{0,...,T-2\}$, $E\left[DID_\ell\right]=\delta_\ell^{trun}$.*
2. *$E\left[\widehat{\delta}_{SQ}^{trun}\right]=\delta_{SQ}^{trun}$.*

Theorem 2 shows that δ_{SQ}^{trun} can be unbiasedly estimated even when there is no untreated group at period T. However, unlike δ_{SQ}, it is harder to rationalize why a social planner would want to learn δ_{SQ}^{trun}. The treatment effects in δ_{SQ}^{trun} are a subset of those in δ_{SQ}. Accordingly, we propose to use

$$\lambda^{trun} = \frac{\sum_{g:2\leq F_g\leq NT}\sum_{t=F_g}^{NT} N_{g,t}\beta^t}{\sum_{g:2\leq F_g\leq T}\sum_{t=F_g}^{T} N_{g,t}\beta^t},$$

a quantity that can be computed from the data, to assess whether δ_{SQ}^{trun} is an "interesting" parameter. When that ratio is close to 1, most of the treatment effects in δ_{SQ} are also in δ_{SQ}^{trun}, so δ_{SQ}^{trun} may be useful for social choice. When potential outcomes are bounded, one can estimate bounds for δ_{SQ} based on Theorem 2 and λ^{trun}. Notice that if $\beta=1$, $\lambda^{trun}=N_S^{trun}/N_S$.

Finally, we propose placebo estimators to test Assumptions 4 and 5. For any $\ell\in\{0,...,\lfloor\frac{T-3}{2}\rfloor\}$ and $t\in\{2\ell+3,...,T\}$, let

$$\text{DID}_{t,\ell}^{\text{pl}} = \beta^t\left(\sum_{g:F_g=t-\ell}\frac{N_{g,t}}{N_t^\ell}(Y_{g,t-\ell-1}-Y_{g,t-2\ell-2}) - \sum_{g:F_g>t}\frac{N_{g,t}}{N_t^{nt}}(Y_{g,t-\ell-1}-Y_{g,t-2\ell-2})\right)$$

if $N_t^\ell>0$ and $N_t^{nt}>0$, and let $\text{DID}_{t,\ell}^{\text{pl}}=0$ otherwise. Let also $N_{\text{DID}_\ell^{\text{pl}}}=\sum_{t=2\ell+3}^{T}N_t^\ell$, and let

$$\text{DID}_\ell^{\text{pl}} = \sum_{t=2\ell+3}^{T}\frac{N_t^\ell}{N_{\text{DID}_\ell^{\text{pl}}}}\text{DID}_{t,\ell}^{\text{pl}}$$

if $N_{\text{DID}_\ell^{\text{pl}}} > 0$, and let $\text{DID}_\ell^{\text{pl}} = 0$ otherwise. $\text{DID}_{t,\ell}^{\text{pl}}$ compares the outcome evolution in the same two sets of groups as those used in $\text{DID}_{t,\ell}$, but between periods $t-2\ell-2$ and $t-\ell-1$ instead of $t-\ell-1$ and t. Thus, $\text{DID}_{t,\ell}^{\text{pl}}$ is a placebo estimator testing if parallel trends holds for $\ell+1$ periods, the same number of periods over which parallel trends has to hold for $\text{DID}_{t,\ell}$ to be an unbiased estimator of the dynamic treatment effect ℓ periods after starting receiving the treatment. $\text{DID}_\ell^{\text{pl}}$ averages those placebos across t, thus providing us a placebo estimator mimicking DID_ℓ.

Theorem 3 *Suppose that Assumptions 1-6 hold. For any* $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$, $E\left[DID_\ell^{pl}\right] = 0$.

Theorem 3 shows that $E\left[\text{DID}_\ell^{\text{pl}}\right] = 0$ is a testable implication of Assumptions 4 and 5, so one can reject those assumptions when $\text{DID}_\ell^{\text{pl}}$ is significantly different from 0. More informally, it may be the case that, say, $\text{DID}_\ell^{\text{pl}}$ is significantly different from 0 for $\ell \geq 2$, but DID_0^{pl} and DID_1^{pl} are not significantly different from 0. That would suggest that violations of parallel trends may invalidate DID_ℓ for $\ell \geq 2$, but not DID_0 and DID_1.

Finally, note that the "long-difference" placebos we propose here differ from the "first-difference" placebos we proposed in de Chaisemartin and D'Haultfœuille (2020), and that compare the $t-k-2$ to $t-k-1$ outcome evolution in groups becoming treated at t and in groups not yet treated at t, for $k \in \{0, ..., T-2\}$.[3] Both placebo estimators have advantages and drawbacks. The long-difference ones test the common trends assumption underlying the instantaneous and dynamic treatment effect estimators, but one can at most compute $\lfloor \frac{T-3}{2} \rfloor + 1$ of them. On the other hand, the first-difference placebos test only the common trends assumption underlying the instantaneous treatment effect estimator, but one will often be able to compute more of them, using parts of the data that the long-run placebos may not be using.

4 Results in general designs

4.1 Results without ruling out dynamic effects beyond any lag

In this section we no longer assume that treatments follow a staggered adoption design. Groups may switch in or out of the treatment at any date. For every $g \in \{1, ..., G\}$, let

$$S_g = \min\{t \geq 2 : D_{g,t} \neq D_{g,t-1}\}$$

denote the first date at which group g's treatment changes, with the convention that $S_g = T+1$ if group g's treatment never changes. For all t, let $\mathbf{1}_t$ denote a $1 \times t$ vector of ones. Hereafter,

[3] For completeness, we define those placebo estimators in the Web Appendix.

we refer to $Y_{g,t}(\mathbf{1}_t)$, the potential outcome that group g will obtain at period t if she remains treated from period 1 to t, as the always-treated potential outcome.

Again, our parameters of interest are motivated by a welfare analysis. In groups untreated at period 1, a planner interested in evaluating the welfare effects of the policy changes that took place from period 1 to T wants to compare the actual outcome to the never-treated outcome, the outcome that would have been realized without any policy change. This leads us to consider the following parameter:

$$\delta_{SQ,0} = E\left(\sum_{g:S_g\leq T}(1-D_{g,1})\sum_{t=S_g}^{T}\frac{N_{g,t}}{N_{S,0}}\beta^t(Y_{g,t}(\boldsymbol{D}_{g,t})-Y_{g,t}(\mathbf{0}_t))\right),$$

where $N_{S,0}=\sum_{g:S_g\leq T}(1-D_{g,1})\sum_{t=S_g}^{T}N_{g,t}$ is the number of units in the (g,t) cells such that group g was untreated at period 1 and her treatment has changed for the first time at or before t. Under Assumption 6, $\delta_{SQ,0}$ is equal to δ_{SQ} in the previous section, so $\delta_{SQ,0}$ is a generalization of δ_{SQ} to non-staggered designs.

In groups treated at period 1, the planner wants to compare the actual outcome to the always-treated outcome, the outcome that would have been realized without any policy change. This leads us to consider the following parameter:

$$\delta_{SQ,1} = E\left(\sum_{g:S_g\leq T}D_{g,1}\sum_{t=S_g}^{T}\frac{N_{g,t}}{N_{S,1}}\beta^t(Y_{g,t}(\boldsymbol{D}_{g,t})-Y_{g,t}(\mathbf{1}_t))\right),$$

where $N_{S,1}=\sum_{g:S_g\leq T}D_{g,1}\sum_{t=S_g}^{T}N_{g,t}$ is the number of units in the (g,t) cells such that group g was treated at period 1 and her treatment has changed for the first time at or before t.

Finally, the planner may be interested in aggregating those two parameters, to perform a cost-benefit analysis of all the policy changes that occurred over the period. This leads us to consider the following parameter:

$$\delta_{CB} = \frac{N_{S,0}}{N_S}\delta_{SQ,0} - \frac{N_{S,1}}{N_S}\delta_{SQ,1},$$

where $N_S = N_{S,0}+N_{S,1}$. $\delta_{SQ,1}$ enters with a negative sign, because it is the effect of a reduction in exposure to the treatment, while $\delta_{SQ,0}$ is the effect of an increase. When aggregating them, the two parameters have to be put on the same scale.

As in the previous section, we cannot unbiasedly estimate $\delta_{SQ,0}$, $\delta_{SQ,1}$ and δ_{CB} in general, so we consider hereafter their truncated versions. Let $NT = \max_{g:D_{g,1}=0}S_g - 1$ denote the last period at which at least one group has never been treated, and let $AT = \max_{g:D_{g,1}=1}S_g - 1$ denote the

last period at which at least one group has always been treated. Then, let

$$\begin{aligned}
\delta^{trun}_{SQ,0} =& E\left(\sum_{g:S_g\leq NT}(1-D_{g,1})\sum_{t=S_g}^{NT}\frac{N_{g,t}}{N^{trun}_{S,0}}\beta^t(Y_{g,t}(\boldsymbol{D}_{g,t})-Y_{g,t}(\mathbf{0}_t))\right),\\
\delta^{trun}_{SQ,1} =& E\left(\sum_{g:S_g\leq AT}D_{g,1}\sum_{t=S_g}^{AT}\frac{N_{g,t}}{N^{trun}_{S,1}}\beta^t(Y_{g,t}(\boldsymbol{D}_{g,t})-Y_{g,t}(\mathbf{1}_t))\right),\\
\delta^{trun}_{CB} =& \frac{N^{trun}_{S,0}}{N^{trun}_{S}}\delta^{trun}_{SQ,0}-\frac{N^{trun}_{S,1}}{N^{trun}_{S}}\delta^{trun}_{SQ,1},
\end{aligned}$$

where

$$\begin{aligned}
N^{trun}_{S,0} =& \sum_{g:S_g\leq NT}(1-D_{g,1})\sum_{t=S_g}^{NT}N_{g,t},\\
N^{trun}_{S,1} =& \sum_{g:S_g\leq AT}D_{g,1}\sum_{t=S_g}^{AT}N_{g,t},\\
N^{trun}_{S} =& N^{trun}_{S,0}+N^{trun}_{S,1}.
\end{aligned}$$

$\delta^{trun}_{SQ,0}$ (resp. $\delta^{trun}_{SQ,1}$) is a version of $\delta_{SQ,0}$ (resp. $\delta_{SQ,1}$) truncated at NT (resp. AT). Similarly, δ^{trun}_{CB} is a truncated version of δ_{CB}. As in the previous section, let

$$\lambda^{trun}=\frac{\sum_{g:S_g\leq NT}(1-D_{g,1})\sum_{t=S_g}^{NT}N_{g,t}\beta^t+\sum_{g:S_g\leq AT}D_{g,1}\sum_{t=S_g}^{AT}N_{g,t}\beta^t}{\sum_{g:S_g\leq T}(1-D_{g,1})\sum_{t=S_g}^{T}N_{g,t}\beta^t+\sum_{g:S_g\leq T}D_{g,1}\sum_{t=S_g}^{T}N_{g,t}\beta^t}$$

denote the "'proportion" of δ_{CB}'s treatment effects that are also in δ^{trun}_{CB}.

$\delta^{trun}_{SQ,0}$ can be unbiasedly estimated under the same assumptions as in the previous section. On the other hand, new assumptions are needed to unbiasedly estimate $\delta^{trun}_{SQ,1}$.

Assumption 7 *(Strong exogeneity for the always treated outcome)* $\forall t\geq 2$, $E(Y_{g,t}(\mathbf{1}_t)-Y_{g,t-1}(\mathbf{1}_{t-1})|\boldsymbol{D}_g)=E(Y_{g,t}(\mathbf{1}_t)-Y_{g,t-1}(\mathbf{1}_{t-1}))$.

Assumption 7 is the equivalent of Assumption 4, for the always-treated potential outcome. It requires that the shocks affecting a group's $Y_{g,t}(\mathbf{1}_t)$ be mean independent of that group's treatment sequence.

Assumption 8 *(Common trends for the always treated outcome)* $\forall t\geq 2$, $E(Y_{g,t}(\mathbf{1}_t)-Y_{g,t-1}(\mathbf{1}_{t-1}))$ *does not vary across* g.

Again, Assumption 8 is the equivalent of Assumption 5, for the always-treated potential outcome. It requires that between each pair of consecutive periods, the expectation of the always-treated outcome follow the same evolution over time in every group.

For any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$, let

$$N_t^{\ell,+} = \sum_{g:S_g=t-\ell} (1-D_{g,1})N_{g,t},$$
$$N_t^{\ell,-} = \sum_{g:S_g=t-\ell} D_{g,1}N_{g,t},$$
$$N_t^{nt} = \sum_{g:S_g>t} (1-D_{g,1})N_{g,t},$$
$$N_t^{at} = \sum_{g:S_g>t} D_{g,1}N_{g,t}.$$

Let

$$\text{DID}_{t,\ell}^{+} = \beta^t \left(\sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} (Y_{g,t} - Y_{g,t-\ell-1}) - \sum_{g:S_g>t} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{nt}} (Y_{g,t} - Y_{g,t-\ell-1}) \right)$$

if $N_t^{\ell,+} > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{t,\ell}^{+} = 0$ otherwise. $\text{DID}_{t,\ell}^{+}$ is the β^t-discounted DID estimator comparing the outcome evolution from period $t-\ell-1$ to t in groups untreated in period 1 and whose treatment changed for the first time in $t-\ell$ and in groups untreated in period 1 and whose treatment has not changed yet in t. Under Assumptions 4-5, the expectation of the outcome evolution in the latter set of groups is a counterfactual of the evolution that would have taken place in the former set of groups if it had remained untreated till t. Thus, $\text{DID}_{t,\ell}^{+}$ is an unbiased estimator of the effect of not having remained untreated till t in those groups.

Let

$$\text{DID}_{t,\ell}^{-} = \beta^t \left(\sum_{g:S_g=t-\ell} D_{g,1} \frac{N_{g,t}}{N_t^{\ell,-}} (Y_{g,t} - Y_{g,t-\ell-1}) - \sum_{g:S_g>t} D_{g,1} \frac{N_{g,t}}{N_t^{at}} (Y_{g,t} - Y_{g,t-\ell-1}) \right)$$

if $N_t^{\ell,-} > 0$ and $N_t^{at} > 0$, and let $\text{DID}_{t,\ell}^{-} = 0$ otherwise. $\text{DID}_{t,\ell}^{-}$ is the β^t-discounted DID estimator comparing the outcome evolution from period $t-\ell-1$ to t in groups treated in period 1 and whose treatment changed for the first time in $t-\ell$ and in groups treated in period 1 and whose treatment has not changed yet in t. Under Assumptions 7-8, $\text{DID}_{t,\ell}^{-}$ is an unbiased estimator of the effect of not having remained treated till t in the former groups.

Then, let

$$N_{\text{DID}_\ell^+} = \sum_{t \geq \ell+2, N_t^{nt}>0} N_t^{\ell,+}$$
$$N_{\text{DID}_\ell^-} = \sum_{t \geq \ell+2, N_t^{at}>0} N_t^{\ell,-}.$$

Finally, let

$$\text{DID}_\ell^+ = \sum_{t=\ell+2}^{T} \frac{N_t^{\ell,+}}{N_{\text{DID}_\ell^+}} \text{DID}_{t,\ell}^+$$

if $N_{\text{DID}_\ell^+} > 0$ and let $\text{DID}_\ell^+ = 0$ otherwise, and let

$$\text{DID}_\ell^- = \sum_{t=\ell+2}^{T} \frac{N_t^{\ell,-}}{N_{\text{DID}_\ell^-}} \text{DID}_{t,\ell}^-$$

if $N_{\text{DID}_\ell^-} > 0$ and let $\text{DID}_\ell^- = 0$ otherwise. Under Assumption 6, DID_ℓ^+ is equal to DID_ℓ in the previous section. Finally, we let

$$\begin{aligned}
\widehat{\delta}_{SQ,0}^{trun} &= \sum_{\ell=0}^{L} \frac{N_{\text{DID}_\ell^+}}{N_{S,0}^{trun}} \text{DID}_\ell^+, \\
\widehat{\delta}_{SQ,1}^{trun} &= \sum_{\ell=0}^{L} \frac{N_{\text{DID}_\ell^-}}{N_{S,1}^{trun}} \text{DID}_\ell^-, \\
\widehat{\delta}_{CB}^{trun} &= \frac{N_{S,0}^{trun}}{N_S^{trun}} \widehat{\delta}_{SQ,0}^{trun} - \frac{N_{S,1}^{trun}}{N_S^{trun}} \widehat{\delta}_{SQ,1}^{trun}.
\end{aligned}$$

Theorem 4 *Suppose that Assumptions 1-3 hold.*

1. *If Assumptions 4 and 5 also hold,* $E\left[\widehat{\delta}_{SQ,0}^{trun}\right] = \delta_{SQ,0}^{trun}$.
2. *If Assumptions 7 and 8 also hold,* $E\left[\widehat{\delta}_{SQ,1}^{trun}\right] = \delta_{SQ,1}^{trun}$.
3. *If Assumptions 4-5 and 7-8 also hold,* $E\left[\widehat{\delta}_{CB}^{trun}\right] = \delta_{CB}^{trun}$.

If there is at least one group untreated from period 1 to T, $\delta_{SQ,0}^{trun} = \delta_{SQ,0}$, so Point 1 of Theorem 4 implies that $\delta_{SQ,0}$ can be unbiasedly estimated. Similarly, if there is at least one group treated from period 1 to T, $\delta_{SQ,1}^{trun} = \delta_{SQ,1}$, so Point 2 of Theorem 4 implies that $\delta_{SQ,1}$ can be unbiasedly estimated. Accordingly, if there is both a never treated and an always treated group, δ_{CB}, the parameter the planner needs to know to evaluate the policy changes that took place during the period under consideration, can also be unbiasedly estimated. If there is no never-treated or no always-treated group, δ_{CB} can no longer be unbiasedly estimated, but one can still unbiasedly estimate δ_{CB}^{trun}, whose closedness to δ_{CB} can be assessed by computing λ^{trun}. Whether λ^{trun} is close to 1 or not depends on whether NT (resp. AT), the number of periods for which at least one group remains untreated (resp. treated) is close to T or not. Accordingly, δ_{CB}^{trun} and δ_{CB} should be close in short panels, or in instances where groups rarely change treatment.

Theorem 4 above relies on Assumptions 4, 5, 7, and 8. We now define placebo estimators one can use to test those assumptions. For any $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$ and $t \in \{2\ell+3, ..., T\}$, let

$$\text{DID}^{+,\text{pl}}_{t,\ell} = \beta^t \left(\sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N^{\ell,+}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right.$$
$$\left. - \sum_{g:S_g>t} (1-D_{g,1}) \frac{N_{g,t}}{N^{nt}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right)$$

if $N^{\ell,+}_t > 0$ and $N^{nt}_t > 0$, and let $\text{DID}^{+,\text{pl}}_{t,\ell} = 0$ otherwise. Let also $N_{\text{DID}^{+,\text{pl}}_\ell} = \sum_{t=2\ell+3}^{T} N^{\ell,+}_t$, and let

$$\text{DID}^{+,\text{pl}}_\ell = \sum_{t=2\ell+3}^{T} \frac{N^{\ell,+}_t}{N_{\text{DID}^{+,\text{pl}}_\ell}} \text{DID}^{+,\text{pl}}_{t,\ell}$$

if $N_{\text{DID}^{+,\text{pl}}_\ell} > 0$, and let $\text{DID}^{+,\text{pl}}_\ell = 0$ otherwise. $\text{DID}^{+,\text{pl}}_{t,\ell}$ compares the outcome evolution in the same two sets of groups as those used in $\text{DID}^{+}_{t,\ell}$, but between periods $t-2\ell-2$ and $t-\ell-1$ instead of $t-\ell-1$ and t. Thus, $\text{DID}^{+,\text{pl}}_{t,\ell}$ tests if parallel trends holds for $\ell+1$ periods, the same number of periods over which parallel trends has to hold for $\text{DID}^{+}_{t,\ell}$ to be an unbiased estimator of the effect of having changed treatment ℓ periods ago.

Similarly, for any $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$ and $t \in \{2\ell+3, ..., T\}$, let

$$\text{DID}^{-,\text{pl}}_{t,\ell} = \beta^t \left(\sum_{g:S_g=t-\ell} D_{g,1} \frac{N_{g,t}}{N^{\ell,-}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) - \sum_{g:S_g>t} D_{g,1} \frac{N_{g,t}}{N^{at}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right)$$

if $N^{\ell,-}_t > 0$ and $N^{at}_t > 0$, and let $\text{DID}^{-,\text{pl}}_{t,\ell} = 0$ otherwise. Let also $N_{\text{DID}^{-,\text{pl}}_\ell} = \sum_{t=2\ell-3}^{T} N^{\ell,-}_t$, and let

$$\text{DID}^{-,\text{pl}}_\ell = \sum_{t=2\ell-3}^{T} \frac{N^{\ell,-}_t}{N_{\text{DID}^{-,\text{pl}}_\ell}} \text{DID}^{-,\text{pl}}_{t,\ell}$$

if $N_{\text{DID}^{-,\text{pl}}_\ell} > 0$, and let $\text{DID}^{-,\text{pl}}_\ell = 0$ otherwise.

Theorem 5 *Suppose that Assumptions 1-3 hold and $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$.*

1. *If Assumptions 4 and 5 also hold, $E\left[DID^{+,pl}_\ell\right] = 0$.*
2. *If Assumptions 7 and 8 also hold, $E\left[DID^{-,pl}_\ell\right] = 0$.*

Theorem 5 shows that $E\left[\text{DID}^{+,\text{pl}}_\ell\right] = 0$ (resp. $E\left[\text{DID}^{-,\text{pl}}_\ell\right] = 0$) is a testable implication of Assumptions 4 and 5 (resp. Assumptions 7 and 8), so one can reject those assumptions when $\text{DID}^{\text{pl}}_{+,\ell}$ (resp. $\text{DID}^{\text{pl}}_{-,\ell}$) is significantly different from 0. One could consider instead first-difference

placebo estimators, comparing the $t-k-2$ to $t-k-1$ outcome evolution in groups changing treatment for the first time in t and in groups that have not changed treatment yet in t. The trade-off between the first- and long-difference placebo estimators are the same as in staggered adoption designs. These alternative estimators are defined and discussed in our Web Appendix.

4.2 Results ruling out dynamic effects beyond a certain lag

On the other hand, λ^{trun} may be low in long panels, and in instances where groups often change treatment. Then, δ_{CB}^{trun} may be very different from δ_{CB}. We now consider an assumption on dynamic treatment effects indexed by $k \in \{0, ..., T-1\}$. If this assumption is plausible for some k, one may be able to unbiasedly estimate a parameter close to that the planner needs to know for policy evaluation.

Assumption 9-k *(Treatments from up to k-periods ago can affect the current outcome)*
For all g, all $t \geq k+1$, and all $(d_1, ..., d_t) \in \{0,1\}^t$, $Y_{g,t}(d_1, ..., d_t) = Y_{g,t}(d_{t-k}, ..., d_t)$.

Assumption 9-k is equivalent to ruling out dynamic effects beyond the kth-lagged treatment, an assumption often implicitly made in event-study regressions (see Autor, 2003). For instance, if $k=1$, only the period-t and period-$t-1$ treatments can affect the period t-outcome. Assumption 9-k is plausible in instances where the treatment is unlikely to have very long-run effects.

Under Assumption 9-k, if a group's treatment changes at some point, this change may have an effect for at most k periods thereafter. Accordingly, in her cost-benefit analysis, the planner only needs to take into account the effect of that change at the period it takes place and the k following periods, relative to the scenario where that change had not taken place. This motivates the following generalization of the δ_{CB} parameter introduced in the previous subsection. For all g, $t \geq 2$, and $k \geq 0$, let

$$S_{g,t,k} = \min\{t' \in \{\max(t-k, 2), ..., t\} : D_{g,t'} \neq D_{g,t'-1}\}$$

be the least recent date included between $t-k$ and t (or 2 and t if $t-k<2$) at which group g's treatment changed. We let $S_{g,t,k}=0$ if group g's treatment did not change treatment between $\max(t-k,2)-1$ and t, or if $t=1$. Let $N_{S,k} = \sum_{(g,t):S_{g,t,k}\geq 2} N_{g,t}$ denote the number of units in the (g,t) cells such that g's treatment changed at least once over the k periods before t. Finally, let

$$\begin{aligned}\delta_{CB,k} =& E\left(\sum_{(g,t):S_{g,t,k}\geq 2} (1 - D_{g,S_{g,t,k}-1}) \frac{N_{g,t}}{N_{S,k}} \beta^t (Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) \right. \\ & \left. - D_{g,S_{g,t,k}-1} \sum_{(g,t):S_{g,t,k}\geq 2} \frac{N_{g,t}}{N_{S,k}} \beta^t (Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)})) \right).\end{aligned}$$

$\delta_{CB,k}$ is a discounted weighted average of the effect of having changed treatment, in all the (g,t) cells such that g's treatment changed at least once over the last k periods before t. Remarking that $S_{g,t,T-1} = S_g 1\{t \geq S_g\}$, one can show that $\delta_{CB,T-1}$ is equal to δ_{CB} in the previous subsection. Also, $\delta_{CB,0}$ with $\beta = 1$ is equal to the average treatment effect among the switchers considered by de Chaisemartin and D'Haultfœuille (2020).

As in the previous sections, we cannot unbiasedly estimate $\delta_{CB,k}$ in general, so we consider a truncated version of it. Let

$$\begin{aligned} C_{g,t,k} =& S_{g,t,k} \times 1\left\{S_{g,S_{g,t,k}-1,k-1} = 0, \exists g' \in \{1,...,G\} : S_{g',t,k} = S_{g',S_{g,t,k}-1,k-1} = 0,\right. \\ & \left. D_{g',S_{g,t,k}-1} = D_{g,S_{g,t,k}-1}\right\}. \end{aligned}$$

Thus, $C_{g,t,k} \neq 0$ (and then $C_{g,t,k} \geq 2$) under the following conditions. First, g's treatment has changed at least once over the k periods before t (so that $S_{g,t,k} > 0$) and did not change for at least $k-1$ periods before that change (so that $S_{g,S_{g,t,k}-1,k-1} = 0$). Second, at least another group g' had the same treatment as g before g's treatment changed ($D_{g',S_{g,t,k}-1} = D_{g,S_{g,t,k}-1}$) and did not experience any change in its treatment from the $k-1$th period before g's treatment changed until period t ($S_{g',t,k} = S_{g',S_{g,t,k}-1,k-1} = 0$). Under Assumptions 4, 5, 7, 8, and 9-k, g' can act as a control to infer the outcome evolution g would have experienced until period t if her treatment had not changed. Let $N^{trun}_{S,k} = \sum_{(g,t):C_{g,t,k}\geq 2} N_{g,t}$ denote the number of units in the (g,t) cells such that g's treatment changed at least once over the k periods before t, and there is at least another group g' that can act as a control for g. Finally, let

$$\begin{aligned} \delta^{trun}_{CB,k} =& E\left(\sum_{(g,t):C_{g,t,k}\geq 2} (1 - D_{g,S_{g,t,k}-1}) \frac{N_{g,t}}{N^{trun}_{S,k}} \beta^t (Y_{g,t}(D_{g,\max(t-k,1)},...,D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)})\right. \\ & \left. - D_{g,S_{g,t,k}-1} \sum_{(g,t):C_{g,t,k}\geq 2} \frac{N_{g,t}}{N^{trun}_{S,k}} \beta^t (Y_{g,t}(D_{g,\max(t-k,1)},...,D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)}))\right). \end{aligned}$$

$\delta^{trun}_{CB,k}$ is a version of $\delta_{CB,k}$ truncated from the treatment effect in all the (g,t) cells such that g's treatment changed at least once over the k periods before t but there is no other group g' that can act as a control to estimate the effect of that change. As previously, let

$$\lambda^{trun}_k = \frac{\sum_{(g,t):C_{g,t,k}\geq 2} N_{g,t}\beta^t}{\sum_{(g,t):S_{g,t,k}\geq 2} N_{g,t}\beta^t}$$

denote the "'proportion" of $\delta_{CB,k}$'s treatment effects that are also in $\delta^{trun}_{CB,k}$.

$\delta^{trun}_{CB,T-1} = \delta^{trun}_{CB}$, so if $k = T-1$, the estimators we propose are the same as in the previous

subsection. We therefore assume that $k \leq T-2$. For any $\ell \in \{0,...,k\}$ and $t \in \{\ell+2,...,T\}$, let

$$\begin{aligned}
N_t^{\ell,+} &= \sum_{g:S_{g,t,k}=t-\ell,S_{g,t-\ell-1,k-1}=0} (1-D_{g,t-\ell-1})N_{g,t} \\
N_t^{\ell,-} &= \sum_{g:S_{g,t,k}=t-\ell,S_{g,t-\ell-1,k-1}=0} D_{g,t-\ell-1}N_{g,t} \\
N_t^{\ell,nt} &= \sum_{g:S_{g,t,k}=0,S_{g,t-\ell-1,k-1}=0} (1-D_{g,t-\ell-1})N_{g,t} \\
N_t^{\ell,at} &= \sum_{g:S_{g,t,k}=0,S_{g,t-\ell-1,k-1}=0} D_{g,t-\ell-1}N_{g,t}.
\end{aligned}$$

Let

$$\begin{aligned}
\text{DID}_{t,\ell}^{+} =& \beta^t \left(\sum_{g:S_{g,t,k}=t-\ell,S_{g,t-\ell-1,k-1}=0} (1-D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} (Y_{g,t}-Y_{g,t-\ell-1}) \right. \\
& \left. - \sum_{g:S_{g,t,k}=0,S_{g,t-\ell-1,k-1}=0} (1-D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,nt}} (Y_{g,t}-Y_{g,t-\ell-1}) \right)
\end{aligned}$$

if $N_t^{\ell,+} > 0$ and $N_t^{\ell,nt} > 0$, and let $\text{DID}_{t,\ell}^{+} = 0$ otherwise. $\text{DID}_{t,\ell}^{+}$ is the β^t-discounted DID estimator comparing the outcome evolution from period $t-\ell-1$ to t in groups untreated from period $t-\ell-1-k$ to $t-\ell-1$ and treated in $t-\ell$ to the same evolution in groups untreated from period $t-\ell-1-k$ to t. Under Assumptions 4-5 and 9-k, the expectation of the outcome evolution in the latter set of groups is a counterfactual of the evolution that would have taken place in the former set of groups if it had remained untreated till t. Thus, $\text{DID}_{t,\ell}^{+}$ is an unbiased estimator of the effect of not having remained untreated till t in those groups. Let

$$\begin{aligned}
\text{DID}_{t,\ell}^{-} =& \beta^t \left(\sum_{g:S_{g,t,k}=t-\ell,S_{g,t-\ell-1,k-1}=0} D_{g,t-\ell-1} \frac{N_{g,t}}{N_t^{\ell,-}} (Y_{g,t}-Y_{g,t-\ell-1}) \right. \\
& \left. - \sum_{g:S_{g,t,k}=0,S_{g,t-\ell-1,k-1}=0} D_{g,t-\ell-1} \frac{N_{g,t}}{N_t^{\ell,at}} (Y_{g,t}-Y_{g,t-\ell-1}) \right)
\end{aligned}$$

if $N_t^{\ell,-} > 0$ and $N_t^{\ell,at} > 0$, and let $\text{DID}_{t,\ell}^{-} = 0$ otherwise. $\text{DID}_{t,\ell}^{-}$ is the β^t-discounted DID estimator comparing the outcome evolution from period $t-\ell-1$ to t in groups treated from period $t-\ell-1-k$ to $t-\ell-1$ and untreated in $t-\ell$ to the same evolution in groups treated from period $t-\ell-1-k$ to t. Under Assumptions 7-8 and 9-k, $\text{DID}_{t,\ell}^{-}$ is an unbiased estimator of the effect of not having remained treated till t in the former set of groups.

Then, let $N_{\text{DID}_\ell} = \sum_{t\geq\ell+2,N_t^{\ell,nt}>0} N_t^{\ell,+} + \sum_{t\geq\ell+2,N_t^{\ell,at}>0} N_t^{\ell,-}$. Let

$$\text{DID}_\ell = \sum_{t=\ell+2}^{T} \left(\frac{N_t^{\ell,+}}{N_{\text{DID}_\ell}} \text{DID}_{t,\ell}^{+} - \frac{N_t^{\ell,-}}{N_{\text{DID}_\ell}} \text{DID}_{t,\ell}^{-} \right)$$

if $N_{\text{DID}_\ell} > 0$, and let $\text{DID}_\ell = 0$ otherwise. For $k = 0$ and $\beta = 1$, DID_0 is equal to the DID_M estimator in de Chaisemartin and D'Haultfœuille (2020). Finally, let $\widehat{\delta}^{trun}_{CB,k} = \sum_{\ell=0}^{k}(N_{\text{DID}_\ell}/N^{trun}_{S,k})\text{DID}_\ell$.

Theorem 6 *Suppose that Assumptions 2-5, 7, 8, and 9-k hold and* $N^{trun}_{S,k} > 0$. *Then,*

$$E\left[\widehat{\delta}^{trun}_{CB,k}\right] = \delta^{trun}_{CB,k}.$$

Suppose Assumption 9-0 is plausible. If for all (g,t) such that $t \geq 2$ and $D_{g,t} \neq D_{g,t-1}$, there is another g' such that $D_{g',t} = D_{g',t-1} = D_{g,t-1}$, then $\delta^{trun}_{CB,0} = \delta_{CB,0}$. Therefore, Theorem 6 implies that the parameter the planner needs to know to evaluate the policy changes that took place over the period can be unbiasedly estimated. Note also that in this case, with $k = 0$ and $\beta = 1$ Theorem 6 is equivalent to Theorem 3 in de Chaisemartin and D'Haultfœuille (2020).

Now, suppose Assumption 9-0 is implausible, but Assumption 9-1 is. Moreover, assume that the following two conditions hold:

1. for all $t \geq 3$, all (g,t) such that $D_{g,t} \neq D_{g,t-1}$ are also such that $D_{g,t-1} = D_{g,t-2}$ and there is another g' such that $D_{g',t+1} = D_{g',t} = D_{g',t-1} = D_{g',t-2} = D_{g,t-2}$
2. for all g such that $D_{g,2} \neq D_{g,1}$, there is another g' such that $D_{g',2} = D_{g',1} = D_{g,1}$.

Then $\delta^{trun}_{CB,1} = \delta_{CB,1}$, and Theorem 6 implies that $\delta_{CB,1}$ can be unbiasedly estimated. Notice that the conditions under which $\delta^{trun}_{CB,1} = \delta_{CB,1}$ are stronger than those under which $\delta^{trun}_{CB,0} = \delta_{CB,0}$. Therefore, it is more likely that only the truncated parameter can be unbiasedly estimated when one works under Assumption 9-1, than when one works under Assumption 9-0. Similarly, when only the truncated parameter can be estimated in both cases, that parameter should be closer to the untruncated one under Assumption 9-0 than under Assumption 9-1 ($\lambda^{trun}_1 < \lambda^{trun}_0$). The same conclusion applies to higher values of k: one should typically have that $k \mapsto \lambda^{trun}_k$ is decreasing. Then, there is a trade-off between the plausibility of the assumptions one imposes, and the relevance of the parameter that can be estimated under those assumptions.

As discussed previously some groups may have already been treated prior to period 1, and those treatments may still affect some of their period-1-to-T outcomes. However, under Assumption 9-k, one can circumvent this problem, by redefining the DID_ℓ estimators above by considering only time periods after $k+1$.

Finally, Theorem 6 above relies on Assumptions 4, 5, 7, and 8. Under Assumption 9-k, the placebo estimators one can use to test those assumptions differ from those in the previous

section. For any $\ell \in \{0, ..., \min(\lfloor \frac{T-3}{2} \rfloor, k)\}$ and $t \in \{2\ell+3, ..., T\}$, let

$$\begin{aligned}\text{DID}^{+,\text{pl}}_{t,\ell} =& \beta^t \left(\sum_{g: S_{g,t,k}=t-\ell, S_{g,t-\ell-1,k+\ell}=0} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N^{\ell,+}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right. \\ & \left. - \sum_{g: S_{g,t,k}=0, S_{g,t-\ell-1,k+\ell}=0} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N^{\ell,nt}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right)\end{aligned}$$

if $N^{\ell,+}_t > 0$ and $N^{\ell,nt}_t > 0$, and let $\text{DID}^{+,\text{pl}}_{t,\ell} = 0$ otherwise. Note that $\{g : S_{g,t-\ell-1,k+\ell} = 0\}$ is the subset of $\{g : S_{g,t-\ell-1,k-1} = 0\}$ (the latter being used in $\text{DID}^{+}_{t,\ell}$) for which $D_{g,t-\ell-1} = ... = D_{t-2\ell-2-k}$. Let also $N_{\text{DID}^{+,\text{pl}}_\ell} = \sum_{t=2\ell+3}^{T} N^{\ell,+}_t$, and let

$$\text{DID}^{+,\text{pl}}_{\ell} = \sum_{t=2\ell+3}^{T} \frac{N^{\ell,+}_t}{N_{\text{DID}^{+,\text{pl}}_\ell}} \text{DID}^{+,\text{pl}}_{t,\ell}$$

if $N_{\text{DID}^{+,\text{pl}}_\ell} > 0$, and let $\text{DID}^{+,\text{pl}}_{\ell} = 0$ otherwise. Similarly, let

$$\begin{aligned}\text{DID}^{-,\text{pl}}_{t,\ell} =& \beta^t \left(\sum_{g: S_{g,t,k}=t-\ell, S_{g,t-\ell-1,k+\ell}=0} D_{g,t-\ell-1} \frac{N_{g,t}}{N^{\ell,-}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right. \\ & \left. - \sum_{g: S_{g,t,k}=0, S_{g,t-\ell-1,k+\ell}=0} D_{g,t-\ell-1} \frac{N_{g,t}}{N^{\ell,at}_t} (Y_{g,t-\ell-1} - Y_{g,t-2\ell-2}) \right)\end{aligned}$$

if $N^{\ell,-}_t > 0$ and $N^{\ell,at}_t > 0$, and let $\text{DID}^{-}_{t,\ell} = 0$ otherwise. Let also $N_{\text{DID}^{-,\text{pl}}_\ell} = \sum_{t=2\ell-3}^{T} N^{\ell,-}_t$, and let

$$\text{DID}^{-,\text{pl}}_{\ell} = \sum_{t=2\ell-3}^{T} \frac{N^{\ell,-}_t}{N_{\text{DID}^{-,\text{pl}}_\ell}} \text{DID}^{-,\text{pl}}_{t,\ell}$$

if $N_{\text{DID}^{-,\text{pl}}_\ell} > 0$, and let $\text{DID}^{-,\text{pl}}_{\ell} = 0$ otherwise.

Theorem 7 *Suppose that Assumptions 1-3 and 9-k hold, and* $\ell \in \{0, ..., \min(\lfloor \frac{T-3}{2} \rfloor, k)\}$.

1. *If Assumptions 4 and 5 also hold,* $E\left[DID^{+,pl}_\ell\right] = 0$.
2. *If Assumptions 7 and 8 also hold,* $E\left[DID^{-,pl}_\ell\right] = 0$.

Theorem 7 shows that under Assumption 9-k, $E\left[\text{DID}^{+,\text{pl}}_\ell\right] = 0$ (resp. $E\left[\text{DID}^{-,\text{pl}}_\ell\right] = 0$) is a testable implication of Assumptions 4 and 5 (resp. Assumptions 7 and 8), so one can reject those assumptions when $\text{DID}^{\text{pl}}_{+,\ell}$ (resp. $\text{DID}^{\text{pl}}_{-,\ell}$) is significantly different from 0. Again, one could instead define first-difference placebo estimators under Assumption 9-k. These alternative placebos are defined and discussed in our Web Appendix.

References

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

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

Abraham, S. and Sun, L. (2020), Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Working Paper.

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

Athey, S. and Imbens, G. W. (2018), Design-based analysis in difference-in-differences settings with staggered adoption, Technical report, National Bureau of Economic Research.

Autor, D. H. (2003), 'Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing', *Journal of labor economics* **21**(1), 1–42.

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

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

Callaway, B. and Sant'Anna, P. H. (2018), Difference-in-differences with multiple time periods and an application on the minimum wage and employment. arXiv e-print 1803.09015.

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

de Chaisemartin, C., D'Haultfoeuille, X. and Guyonvarch, Y. (2019), 'DID_MULTIPLEGT: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods', Statistical Software Components, Boston College Department of Economics.
URL: *https://ideas.repec.org/c/boc/bocode/s458643.html*

Manski, C. F. (2005), *Social choice with partial knowledge of treatment response*, Princeton University Press.

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

A Appendix: proofs

A.1 Proofs of Theorems 1 and 2

Let $\boldsymbol{D} = (\boldsymbol{D}_g)_{g=1,...,G}$. We first prove the following lemma.

Lemma 1 *If Assumptions 2-6 hold, then for any* $\ell \in \{0, ..., T-2\}$ $E\left[DID_\ell|\boldsymbol{D}\right] = E\left[\Delta_\ell^{trun}|\boldsymbol{D}\right]$.

Proof: $\text{DID}_\ell 1\{L < \ell\} = \Delta_\ell^{trun} 1\{L < \ell\} = 0$, so to prove the result, it is sufficient to show that $1\{L \geq \ell\} E\left[\text{DID}_\ell|\boldsymbol{D}\right] = 1\{L \geq \ell\} E\left[\Delta_\ell|\boldsymbol{D}\right]$.

We consider an arbitrary $\ell \in \{0, ..., L\}$. By Assumption 5, for all $t \geq 2$ there is a real number ψ_t such that $\psi_t = E(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1}))$ for all g. Then, for all g and all $t \geq \ell + 2$,

$$E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})] = \sum_{k=0}^{\ell} \psi_{t-k}. \tag{4}$$

Then, for all $t \in \{\ell + 2, ..., T\}$ such that $N_t^\ell > 0$ and $N_t^{nt} > 0$,

$$\begin{aligned}
&E[\text{DID}_{t,\ell}|\boldsymbol{D}] \\
=&\beta^t \left(\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} E[Y_{g,t} - Y_{g,t-\ell-1}|\boldsymbol{D}] - \sum_{g:F_g>t} \frac{N_{g,t}}{N_t^{nt}} E[Y_{g,t} - Y_{g,t-\ell-1}|\boldsymbol{D}] \right) \\
=&\beta^t \left(\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} E\left[Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right] \right. \\
&\left. + \sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} E\left[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})|\boldsymbol{D}\right] - \sum_{g:F_g>t} \frac{N_{g,t}}{N_t^{nt}} E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})|\boldsymbol{D}] \right) \\
=&\beta^t \left(\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} E\left[Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right] \right. \\
&\left. + \sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} E\left[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})\right] - \sum_{g:F_g>t} \frac{N_{g,t}}{N_t^{nt}} E[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})] \right) \\
=&\beta^t E\left[\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} (Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)) \middle| \boldsymbol{D} \right].
\end{aligned} \tag{5}$$

The first equality holds by the definition of $\text{DID}_{t,\ell}$ and the fact that $N_t^\ell > 0$ and $N_t^{nt} > 0$. The second equality holds by Assumptions 3 and 6. The third equality follows from Assumptions 2

and 4. The last equality follows from Equation (4). Then,

$$
\begin{aligned}
& 1\{L \geq \ell\} E[\mathrm{DID}_\ell | \boldsymbol{D}] \\
=& 1\{L \geq \ell, N_{\mathrm{DID}_\ell} > 0\} \left(\sum_{t=\ell+2}^{T} \frac{N_t^\ell}{N_{\mathrm{DID}_\ell}} E[\mathrm{DID}_{t,\ell} | \boldsymbol{D}] \right) \\
=& 1\{L \geq \ell, N_{\mathrm{DID}_\ell} > 0\} \left(\sum_{t=\ell+2}^{NT} \beta^t E \left[\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_\ell^{trun}} (Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)) \middle| \boldsymbol{D} \right] \right) \\
=& 1\{L \geq \ell, N_{\mathrm{DID}_\ell} > 0\} E\left[\Delta_\ell^{trun} | \boldsymbol{D} \right] \\
=& 1\{L \geq \ell\} E\left[\Delta_\ell^{trun} | \boldsymbol{D} \right]. \qquad (6)
\end{aligned}
$$

The first equality follows from the definition of DID_ℓ. The second equality follows from the following facts. First, $\mathrm{DID}_{t,\ell} = 0$ for $t > NT$. Second, if $t \leq NT$, we have $N_t^{nt} > 0$ and thus Equation (5) holds if $N_t^\ell > 0$. Third, if $N_t^\ell = 0$, we have $\sum_{g:F_g=t-\ell} N_{g,t}(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)) = \mathrm{DID}_{t,\ell} = 0$. Finally, we use $N_\ell^{trun} = N_{\mathrm{DID}_\ell}$. The last equality follows from $N_\ell^{trun} = N_{\mathrm{DID}_\ell}$, and $\Delta_\ell^{trun} = 0$ if $N_\ell^{trun} = 0$ □

Turning to Theorem 2, Point 1 follows from Lemma 1 and the law of iterated expectations. Point 2 follows from $N_{\mathrm{DID}_\ell} = N_\ell^{trun}$, Lemma 1, the law of iterated expectations and (3).

Finally, we prove Theorem 1. Point 1 follows from Point 1 of Theorem 2, and the fact that $\max_{g\in\{1,\ldots,G\}} F_g > \max_{g:2\leq F_g\leq T-\ell} F_g + \ell$ implies that $\delta_\ell^{trun} = \delta_\ell$. Point 2 follows from Point 2 of Theorem 2, and the fact that $\max_{g\in\{1,\ldots,G\}} F_g = T+1$ implies that $\delta_{SQ}^{trun} = \delta_{SQ}$.

A.2 Proof of Theorem 3

Following the same steps as those used to obtain (5), we get, whenever $N_t^\ell > 0$ and $N_t^{nt} > 0$,

$$
\begin{aligned}
E[\mathrm{DID}_{t,\ell}^{\mathrm{pl}} | \boldsymbol{D}] =& \beta^t E \left[\sum_{g:F_g=t-\ell} \frac{N_{g,t}}{N_t^\ell} (Y_{g,t-\ell-1}(\boldsymbol{D}_{g,t-\ell-1}) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})) \middle| \boldsymbol{D} \right] \\
=& 0,
\end{aligned}
$$

where the second equality follows since by definition, $F_g = t-\ell$ implies that $\boldsymbol{D}_{g,t-\ell-1} = \mathbf{0}_{t-\ell-1}$. The result follows using the same reasoning as that used to obtain (6).

A.3 Proof of Theorem 4

1. First, note that for any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$ such that $N_t^{\ell,+} > 0$ and $N_t^{nt} > 0$,

$$
\begin{aligned}
&E\left(\mathrm{DID}_{t,\ell}^{+}|\boldsymbol{D}\right) \\
=&\beta^t \left(\sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t} - Y_{g,t-\ell-1}|\boldsymbol{D}\right) - \sum_{g:S_g>t} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{nt}} E\left(Y_{g,t} - Y_{g,t-\ell-1}|\boldsymbol{D}\right) \right) \\
=&\beta^t \left(\sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right) \right. \\
&+ \sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})|\boldsymbol{D}\right) \\
&\left. - \sum_{g:S_g>t} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{nt}} E\left(Y_{g,t}(\mathbf{0}_t) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})|\boldsymbol{D}\right) \right) \\
=&\beta^t \sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right). \qquad (7)
\end{aligned}
$$

The first equality follows from the definition of $\mathrm{DID}_{t,\ell}^{+}$, $N_t^{\ell,+} > 0$ and $N_t^{nt} > 0$. The second equality follows from Assumption 3. The third equality follows from Assumptions 2 and 4 and Equation (4). Then,

$$
\begin{aligned}
\sum_{\ell=0}^{L} \frac{N_{\mathrm{DID}_\ell^+}}{N_{S,0}^{trun}} E\left(\mathrm{DID}_\ell^+|\boldsymbol{D}\right) &= \sum_{\ell=0}^{L} \sum_{t=\ell+2}^{T} \frac{N_t^{\ell,+}}{N_{S,0}^{trun}} E\left(\mathrm{DID}_{t,\ell}^+|\boldsymbol{D}\right) \\
&= \sum_{\ell=0}^{L} \sum_{t=\ell+2}^{NT} \beta^t \sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_{S,0}^{trun}} E\left(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right) \\
&= \sum_{g:S_g \le NT} (1-D_{g,1}) \sum_{t=S_g}^{NT} \beta^t \frac{N_{g,t}}{N_{S,0}^{trun}} E\left(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right). \qquad (8)
\end{aligned}
$$

The first equality follows from the definition of DID_ℓ^+, and from the fact $\mathrm{DID}_{t,\ell}^+ = 0$ for all $t \in \{\ell+2, ..., T\}$ if $N_{\mathrm{DID}_\ell^+} = 0$. The second equality follows from the following facts. First, if $t > NT$, $\mathrm{DID}_{t,\ell}^+ = 0$. Second, if $t \le NT$, $N_t^{nt} > 0$ by definition of NT. Then we use Equation (7) if $N_t^{\ell,+} > 0$, and if $N_t^{\ell,+} = 0$,

$$
\mathrm{DID}_{t,\ell}^{+} = \sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_{S,0}^{trun}} E\left(Y_{g,t}(\boldsymbol{D}_{g,t}) - Y_{g,t}(\mathbf{0}_t)|\boldsymbol{D}\right) = 0.
$$

Then Point 1 follows from Equation (8) and the law of iterated expectations. Point 2 follows

from the same reasoning, using Assumptions 7 and 8 instead of Assumptions 4 and 5. Point 3 follows directly from Points 1 and 2.

A.4 Proof of Theorem 5

Following the same steps as those used to obtain (7), we get, whenever $N_t^{\ell,+} > 0$ and $N_t^{nt} > 0$,

$$\begin{aligned} E[\mathrm{DID}_{t,\ell}^{+,\mathrm{pl}}|\boldsymbol{D}] =& \beta^t \sum_{g:S_g=t-\ell} (1-D_{g,1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left[Y_{g,t-\ell-1}(\boldsymbol{D}_{g,t-\ell-1}) - Y_{g,t-\ell-1}(\mathbf{0}_{t-\ell-1})|\boldsymbol{D}\right] \\ =& 0, \end{aligned}$$

where the second equality follows since by definition, $S_g = t - \ell$ and $D_{g,1} = 0$ imply that $\boldsymbol{D}_{g,t-\ell-1} = \mathbf{0}_{t-\ell-1}$. Then, Point 1 follows using the same reasoning as that used to obtain (8). Point 2 can be obtained similarly.

A.5 Proof of Theorem 6

Under Assumption 9-k, it follows from Equation (4) that for all g, all $\ell \geq 0$ and all $t \geq \ell + 2$,

$$E[Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) - Y_{g,t-\ell-1}(\mathbf{0}_{\min(t-\ell-1,k+1)})] = \sum_{j=0}^{\ell} \psi_{t-j}. \tag{9}$$

For any $\ell \in \{0, ..., k\}$ and $t \in \{\ell + 2, ..., T\}$ such that $N_t^{\ell,+} > 0$ and $N_t^{\ell,nt} > 0$,

$$
\begin{aligned}
& E\left(\mathrm{DID}_{t,\ell}^{+} | \boldsymbol{D}\right) \\
=& \beta^t \left(\sum_{\substack{g: S_{g,t,k} = t-\ell, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t} - Y_{g,t-\ell-1} | \boldsymbol{D}\right) \right. \\
& \left. - \sum_{\substack{g: S_{g,t,k} = 0, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,nt}} E\left(Y_{g,t} - Y_{g,t-\ell-1} | \boldsymbol{D}\right) \right) \\
=& \beta^t \left(\sum_{\substack{g: S_{g,t,k} = t-\ell, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) | \boldsymbol{D}\right) \right. \\
& + \sum_{\substack{g: S_{g,t,k} = t-\ell, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) - Y_{g,t-\ell-1}(\mathbf{0}_{\min(t-\ell-1,k+1)}) | \boldsymbol{D}\right) \\
& \left. - \sum_{\substack{g: S_{g,t,k} = 0, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,nt}} E\left(Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) - Y_{g,t-\ell-1}(\mathbf{0}_{\min(t-\ell-1,k+1)}) | \boldsymbol{D}\right) \right) \\
=& \beta^t \sum_{\substack{g: S_{g,t,k} = t-\ell, \\ S_{g,t-\ell-1,k-1}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) | \boldsymbol{D}\right). \qquad (10)
\end{aligned}
$$

The first equality follows from the definition of $\mathrm{DID}_{t,\ell}^{+}$ and $N_t^{\ell,+} > 0$ and $N_t^{\ell,nt} > 0$. The second equality follows from Assumptions 3 and 9-k and the definitions of $S_{g,t,k}$ and $S_{g,t-\ell-1,k-1}$. The third equality follows from Assumptions 2 and 4 and Equation (9).

Similarly, one can show that for any $\ell \in \{0, ..., k\}$ and $t \in \{\ell + 2, ..., T\}$ such that $N_t^{\ell,-} > 0$ and $N_t^{\ell,at} > 0$,

$$
\begin{aligned}
& E\left(\mathrm{DID}_{t,\ell}^{-} | \boldsymbol{D}\right) \\
=& \beta^t \sum_{\substack{g: S_{g,t,k} = t-\ell, \\ S_{g,t-\ell-1,k-1}=0}} D_{g,t-\ell-1} \frac{N_{g,t}}{N_t^{\ell,-}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)}) | \boldsymbol{D}\right). \qquad (11)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\sum_{\ell=0}^{k} \frac{N_{\mathrm{DID}_\ell}}{N_{S,k}^{trun}} E\left(\mathrm{DID}_\ell | \boldsymbol{D}\right) \\
=&\sum_{\ell=0}^{k} \sum_{t=\ell+2}^{T} \left(\frac{N_t^{\ell,+}}{N_{S,k}^{trun}} E\left(\mathrm{DID}_{t,\ell}^{+} | \boldsymbol{D}\right) - \frac{N_t^{\ell,-}}{N_{S,k}^{trun}} E\left(\mathrm{DID}_{t,\ell}^{-} | \boldsymbol{D}\right) \right) \\
=&\sum_{\ell=0}^{k} \sum_{t=\ell+2}^{T} \beta^t \sum_{g: C_{g,t,k}=t-\ell} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) | \boldsymbol{D}\right) \\
-&\sum_{\ell=0}^{k} \sum_{t=\ell+2}^{T} \beta^t \sum_{g: C_{g,t,k}=t-\ell} D_{g,t-\ell-1} \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)}) | \boldsymbol{D}\right) \\
=&\sum_{g: C_{g,t,k} \geq 2} (1 - D_{g,t-\ell-1}) \beta^t \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) | \boldsymbol{D}\right) \\
-&\sum_{g: C_{g,t,k} \geq 2} D_{g,t-\ell-1} \beta^t \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)}) | \boldsymbol{D}\right). \qquad (12)
\end{aligned}
$$

The first equality follows from the definition of DID_ℓ, and from the fact $\mathrm{DID}_{t,\ell}^{+} = \mathrm{DID}_{t,\ell}^{-} = 0$ for all $t \in \{\ell+2, ..., T\}$ if $N_{\mathrm{DID}_\ell} = 0$. The second equality follows from the following facts. First,

$$
\mathrm{DID}_{t,\ell}^{+} = \beta^t \sum_{g: C_{g,t,k}=t-\ell} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{0}_{\min(t,k+1)}) | \boldsymbol{D}\right) = 0
$$

if $N_t^{\ell,+} = 0$ or $N_t^{\ell,nt} = 0$ and

$$
\mathrm{DID}_{t,\ell}^{-} = \beta^t \sum_{g: C_{g,t,k}=t-\ell} D_{g,t-\ell-1} \frac{N_{g,t}}{N_{S,k}^{trun}} E\left(Y_{g,t}(D_{g,\max(t-k,1)}, ..., D_{g,t}) - Y_{g,t}(\mathbf{1}_{\min(t,k+1)}) | \boldsymbol{D}\right) = 0
$$

if $N_t^{\ell,-} = 0$ or $N_t^{\ell,at} = 0$. Second, if $N_t^{\ell,+} > 0$ and $N_t^{\ell,nt} > 0$ (resp. $N_t^{\ell,-} > 0$ and $N_t^{\ell,at} > 0$), Equation (10) (resp. (11)) holds.

The result follows from Equation (12) and the law of iterated expectations.

A.6 Proof of Theorem 7

Following the same steps as those used to obtain (10), we get, whenever $N_t^{\ell,+} > 0$ and $N_t^{\ell,nt} > 0$,

$$
\begin{aligned}
E[\mathrm{DID}_{t,\ell}^{+,\mathrm{pl}} | \boldsymbol{D}] =& \beta^t \sum_{\substack{g: S_{g,t,k}=t-\ell, \\ S_{g,t-\ell-1,k+\ell}=0}} (1 - D_{g,t-\ell-1}) \frac{N_{g,t}}{N_t^{\ell,+}} E\left[Y_{g,t-\ell-1}(D_{g,\max(t-\ell-1-k,1)}, ..., D_{g,t-\ell-1})\right. \\
&\left. - Y_{g,t-\ell-1}(\mathbf{0}_{\min(t-\ell-1,k+1)}) | \boldsymbol{D}\right] \\
=& 0,
\end{aligned}
$$

where the second equality follows since by definition, $S_{g,t-\ell-1,k+\ell} = 0$ and $D_{g,t-\ell-1} = 0$ imply that $(D_{g,\max(t-\ell-1-k,1)}, ..., D_{g,t-\ell-1}) = \mathbf{0}_{\min(t-\ell-1,k+1)}$. Then, Point 1 follows using the same reasoning as that used to obtain (12). Point 2 can be obtained similarly.