Difference-in-Differences Estimators of Intertemporal Treatment Effects[*]

Clément de Chaisemartin[†]   Xavier D'Haultfœuille[‡]

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, as a way to evaluate the welfare effects of the policies that occurred over the duration of the panel. We show that under common trends conditions, these parameters can be unbiasedly estimated by weighted sums 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. We also propose estimators of a dynamic linear model, with group-specific but time-invariant effects of the current and lagged treatments, which may be used to evaluate ex-ante the effect of future policies.

Keywords: differences-in-differences, dynamic treatment effects, heterogenous treatment effects, panel data, policy evaluation, welfare analysis.

JEL Codes: C21, C23

1 Introduction

We consider the estimation of the effect of a policy or treatment on an outcome. To do so, we use a panel of groups, indexed by g, that are exposed to the policy at different time periods, indexed by t. The treatment of group g at period t may have an effect on that group's period-t outcome, but it may also have an effect on its future outcomes.

To estimate the treatment's instantaneous and dynamic effects, a commonly-used method is to regress the outcome on group fixed effects, time fixed effects, the contemporaneous value of the treatment, and lags of the treatment. Intuitively, the coefficient of the contemporaneous treatment should estimate its instantaneous effect, while the coefficients of the lagged treatments should estimate its dynamic effects. In staggered adoption designs where groups can switch in but not out of the treatment, researchers have estimated a slightly different regression, where the treatments are replaced by indicators for the number of time periods since a group has started receiving the treatment. Hereafter, we refer to those regressions as event-study regressions. Sun and Abraham (2020) have shown that the coefficients in the second regression are not robust to heterogeneous treatment effects across groups and over time,[1] and could be misleading even under an additive dynamic treatment effect model with constant effects. Schmidheiny and Siegloch (2020) have shown that the two event-study regressions are numerically identical up to a linear bijection. Together with the result in Sun and Abraham (2020), this implies that the first event-study regression is also not robust to heterogeneous or dynamic effects. Instead, we propose to use differences-in-differences (DID) estimators. As the event-study regressions, our estimators rely on the standard common trends assumption, but unlike them they are robust to heterogeneous and dynamic effects.

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). We assume that units' utility is additively separable in the outcome, and that the cost of each treatment period is constant over time and across units. We adopt the perspective of a utilitarian planner interested in assessing whether groups' actual treatments led to a welfare increase, relative to the *status quo* scenario were they would have kept all along the same treatment as in the first period of the panel. In other words, the planner seeks to assess whether the treatment changes that occurred over the period under consideration increased welfare.

In groups untreated at period 1, the status quo treatment is to never be treated. Let c denote the cost of treatment. The welfare of those groups is larger than under the status quo if and only if $\Delta_+ > c$, where Δ_+ is the discounted sum of the difference between those groups' actual

---

[1]This 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.

outcome and their outcome if they had never been treated, divided by the discounted sum of treatments received. Our parameter of interest for those groups is then $\delta_+ = E[\Delta_+]$. We define similarly δ_- for groups treated at period 1 and refer to δ_+ and δ_- as the actual-versus-status-quo parameters. On top of their welfare interpretation, those parameters also tell us the average increase in outcome produced by a one-unit increase in the treatment, over the duration of the panel. Specifically, δ_+ is equal to the ratio of two quantities. Its numerator is the average effect of not having remained untreated throughout the panel, across all observations in (g, t) cells such that group g was initially untreated but became treated at least once at or before t. As we allow for dynamic effects, the outcome of all those observations may be affected by that first switch from untreated to treated, and may differ from their never treated outcome. Then, this numerator is scaled by the average treatment of the same observations. By scaling the average "intention-to-treat" effect of first switches on the outcome by their average "first-stage" effect on the treatment, δ_+ measures the average increase in outcome produced by a one-unit increase in the treatment. δ_-^{tru} has a similar interpretation.

We then build unbiased estimators of these parameters. The logic is similar for both, so let us focus on δ_+. For every t and ℓ, we form a DID estimator comparing the $t - \ell - 1$ to t outcome evolution, in groups treated for the first time in $t - \ell$ and in groups untreated from period 1 to t. Under our common trends assumption, this DID estimator is unbiased for the effect of that first treatment switch, ℓ periods after it took place. Our estimator of δ_+ is a weighted sum of those estimators, across t and ℓ. If at least one group is untreated from period 1 to T, our estimator is unbiased for δ_+. Otherwise, it is unbiased for a truncated version of δ_+, where the truncation happens at the last period t where at least one group has been untreated from period 1 to t. Similarly, we propose an estimator of δ_-, and show that it is unbiased if at least one group is treated from period 1 to T. If not, our estimator is unbiased for a truncated version of δ_-.

We also propose placebo estimators of the common trends assumption, that compare the outcome evolution between the same groups as above, before groups switching treatment do so. Contrary to the standard common trends test in event-study regressions, our test is robust to heterogeneous and dynamic effects.

Finally, the approach proposed so far evaluates ex-post the effects of the policies that took place during the panel. Instead, one may be interested in using the panel to evaluate and compare ex-ante future policies. Such an exercise must rely on the assumption that treatment effects do not vary over time. Otherwise, the panel cannot inform us on future treatment effects. Accordingly, we assume that potential outcomes follow an additive dynamic effect model with time-invariant but group-specific coefficients of the current and lagged treatments, and show that many of those coefficients can be unbiasedly estimated by linear combinations of DID estimators, comparing again the first-time switchers to the not-yet switchers. We then discuss how estimators of those coefficients can be used for ex-ante policy evaluation.

Related literature, and outline of the paper

In de Chaisemartin and D'Haultfœuille (2020), we considered the estimation of the instantaneous treatment effect, with a group-level panel data set, ruling out dynamic effects. In staggered designs, the DID_M estimator we proposed therein is robust to dynamic effects, and it is in fact equal to $\text{DID}_{+,0}$, one of the instantaneous treatment effect estimators we propose in this paper. Outside of staggered designs, the DID_M estimator is not robust to dynamic effects. Thus, we improve on our earlier work, by proposing an estimator of the instantaneous treatment effect robust to dynamic effects, estimators of dynamic effects, and a principled method to aggregate all those effects.

Sun and Abraham (2020) and Callaway and Sant'Anna (2020) have also proposed DID estimators robust to heterogeneous effects in panel data sets. Our paper differs from those on three important dimensions. First and foremost, those papers focus on binary treatments, and on staggered designs where groups can switch in but not out of the treatment. Our results, on the other hand, apply to any design, and to non-binary treatments. Of all the papers using regressions with group and time fixed effects published by the AER between 2010 and 2012, we found in de Chaisemartin and D'Haultfœuille (2020) that less than 10% have a binary treatment and staggered design. Thus, our estimators can be used in a larger set of empirical applications.

Second, our paper uses welfare analysis to guide the aggregation of instantaneous and dynamic treatment effects. We thus complement Callaway and Sant'Anna (2020), who propose several other interesting aggregation methods. Aggregation is especially important outside of staggered designs. With T periods and a binary treatment, there are only $T + 1$ possible treatment trajectories in a staggered design, against 2^T in general designs. Hence, estimators of the effect produced by every trajectory observed will be much noisier in general designs.

Third, in staggered designs, the control group used by our DID estimators differs from that in Sun and Abraham (2020). To estimate the treatment effect at date t in groups treated at date $t - \ell$, we use as controls all groups not yet treated at t, while they use the never treated groups or the groups treated last if there are no never treated groups. Our control group is larger, so our estimators ought to be more precise. Callaway and Sant'Anna (2020) propose to use either the never treated or the not yet treated. In de Chaisemartin and D'Haultfœuille (2020), we also proposed to use the not yet treated as controls to estimate the instantaneous treatment effect.

Finally, another related paper is Bojinov et al. (2020), who propose estimators of instantaneous and dynamic treatment effects, in panel experiments. Their approach is applicable when the treatment is randomized, while ours is applicable when it is not.

The paper is organized as follows. Section 2 introduces the notation, our assumptions, and our parameters of interest. Section 3 presents our estimators, and shows that they are unbiased. It

also presents the placebo estimators. We study the additive dynamic effect model in Section 4. Finally, in our Web Appendix we consider several extensions of the estimators proposed in Section 3.

2 Set-up and parameters of interest

2.1 Notation and assumptions

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

One is interested in measuring the effect of a treatment on some outcome. We start by assuming that treatment is binary, but show in Section 1.3 of the Web Appendix that our results can be generalized to ordered treatments. 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, and for all $\boldsymbol{d} \in \{0, 1\}^T$, let $Y_{i,g,t}(\boldsymbol{d})$ denote the potential outcome of observation i in group g at period t, if her treatments from period 1 to T are equal to $\boldsymbol{d}$. This dynamic potential outcome framework is similar to that in Robins (1986). It allows for the possibility that observations' outcome at time t be affected by their past and future treatments. Some observations may have already been treated prior to period 1, the first period in the data, and those treatments may still affect some of their period-1-to-T outcomes. However, we cannot estimate such dynamic effects, as treatments and outcomes are not observed for those periods, so we do not account for this potential dependency in our notation.

We 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]

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

---

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

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

Assumption 2 requires that a group's current outcome do not depend on her future treatments, the so-called no-anticipation hypothesis. Abbring and Van den Berg (2003) have discussed that assumption in the context duration models, and Malani and Reif (2015), Botosaru and Gutierrez (2018), and Sun and Abraham (2020) have discussed it in the context of DID models.

In Assumption 3 below, we require that there is at least one group going from untreated to treated at a date where another group has been untreated all along, or at least one group going from treated to untreated at a date where another group has been treated all along. This only rules out pathological applications, where groups untreated at period 1 either all remain untreated till period T or all get treated for the first time at the same date, and groups treated at period 1 either all remain treated till period T or all get untreated for the first time at the same date. For any $g \in \{1, ..., G\}$, let $F_{g,1} = \min\{t : D_{g,t} = 1\}$ denote the first date at which group g is treated, with the convention that $F_{g,1} = T+1$ if group g is never treated. Similarly, let $F_{g,0} = \min\{t : D_{g,t} = 0\}$ denote the first date at which group g is untreated, with the convention that $F_{g,0} = T + 1$ if group g is always treated.

Assumption 3 *(Non-pathological design) At least one of the two following statements hold:*

1. *There exists $(g, g') \in \{1, ..., G\}^2$ such that $1 < F_{g,1} < F_{g',1}$.*

2. *There exists $(g, g') \in \{1, ..., G\}^2$ such that $1 < F_{g,0} < F_{g',0}$.*

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.

Finally, we let $\boldsymbol{0}$ and $\boldsymbol{1}$ respectively denote $1 \times T$ vectors of zeros and ones. Hereafer, we refer to $Y_{g,t}(\boldsymbol{0})$ and $Y_{g,t}(\boldsymbol{1})$ as group g's never- and always-treated potential outcomes at period t. Our estimators rely on the following assumptions on $Y_{g,t}(\boldsymbol{0})$ and $Y_{g,t}(\boldsymbol{1})$.

Assumption 4 *(Independent groups and strong exogeneity) $\forall t \geq 2$ and $\forall g \in \{1, ..., G\}$,*

1. $E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0})|\boldsymbol{D}_1, ..., \boldsymbol{D}_G) = E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0})|\boldsymbol{D}_g)$.

2. $E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0})|\boldsymbol{D}_g) = E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0}))$.

Point 1 of Assumption 4 requires that conditional on group g's treatment, the shocks affecting its never-treated outcome be mean independent of other groups' treatments. This holds if the potential outcomes and treatments of different groups are independent, a commonly-made assumption in DID analysis, where standard errors are usually clustered at the group level (see

Bertrand et al., 2004). Point 2 is related to the strong exogeneity condition in panel data models. It requires that the shocks affecting group g's never-treated outcome be mean independent of group g's treatments. For instance, this rules out cases where a group gets treated because it experiences negative shocks, the so-called Ashenfelter's dip (see Ashenfelter, 1978).

Assumption 5 *(Common trends)* $\forall t \geq 2$, $E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}))$ *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 common trends assumption in DID models (see, e.g., Abadie, 2005) to our set-up allowing for dynamic effects. Sun and Abraham (2020), Athey and Imbens (2018), and Callaway and Sant'Anna (2020) also consider that assumption.

Assumption 6 *(Independent groups and strong exogeneity for the always treated outcome)* $\forall t \geq 2$ *and* $\forall g \in \{1, ..., G\}$,

1. $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})|\mathbf{D}_1, ..., \mathbf{D}_G) = E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})|\mathbf{D}_g)$.

2. $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})|\mathbf{D}_g) = E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}))$.

Assumption 7 *(Common trends for the always treated outcome)* $\forall t \geq 2$, $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}))$ *does not vary across g.*

Assumptions 6 and 7 are the equivalent of Assumptions 4 and 5, for the always-treated potential outcome.

2.2 Parameters of interest

Our parameters of interest are benefit-to-cost ratios, that a planner may use for treatment choice (see Manski, 2005). Assume that the instantaneous utility for observation i in group g at period t with treatment vector $\mathbf{d}_g = (d_{g,1}, ..., d_{g,T})$ is

$$U_{i,g,t}(\mathbf{d}_g) = Y_{i,g,t}(\mathbf{d}_g) - cd_{g,t} + u_{i,g,t},$$

where c is the instantaneous cost of treatment, assumed to be constant across groups and over time. With treatments $\mathbf{d} = (\mathbf{d}_1, ..., \mathbf{d}_G)$ in groups 1 to g, the utilitarian social welfare in any set of groups $\mathcal{G}$ is

$$W(\mathbf{d}, \mathcal{G}) = \sum_{g \in \mathcal{G}, t} N_{g,t} \beta^t Y_{g,t}(\mathbf{d}_g) - c \sum_{g \in \mathcal{G}, t} N_{g,t} \beta^t d_{g,t} + \sum_{i,g \in \mathcal{G}, t} u_{i,g,t}, \tag{1}$$

where $\beta \in (0, 1]$ is the planner's discount factor.

Let $\boldsymbol{D} = (\boldsymbol{D}_1, ..., \boldsymbol{D}_G)$ be a vector stacking the treatments received by groups 1 to G from periods 1 to T. The planner may want to compare $\boldsymbol{D}$ to the scenario where each observation keeps her period 1 treatment till period T, the status-quo scenario. Thus, she can determine if the treatment changes that occurred over the period increased welfare or not. This analysis can be conducted separately for groups untreated and treated at period 1, whose status quo treatment differs. In groups untreated at period 1 ($F_{g,1} > 1$), the status quo treatment is to never be treated. The actual treatment $\boldsymbol{D}$ is beneficial compared to never being treated if and only if $W((\boldsymbol{D}_1, ..., \boldsymbol{D}_G), g : F_{g,1} > 1) > W((\boldsymbol{0}, ..., \boldsymbol{0}), g : F_{g,1} > 1)$, i.e. if and only if

$$\Delta_+ := \frac{\sum_{(g,t):F_{g,1}>1} N_{g,t}\beta^t (Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}))}{\sum_{(g,t):F_{g,1}>1} N_{g,t}\beta^t D_{g,t}} > c.$$

Accordingly, the planner may be interested in learning $\delta_+ = E\left(\Delta_+\right).$[3]

In groups treated at period 1 ($F_{g,0} > 1$), the status quo treatment is to always be treated. The actual treatment $\boldsymbol{D}$ is beneficial compared to always being treated if and only if

$$\Delta_- := \frac{\sum_{(g,t):F_{g,0}>1} N_{g,t}\beta^t (Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{1}))}{\sum_{(g,t):F_{g,0}>1} N_{g,t}\beta^t (D_{g,t} - 1)} < c,$$

where the inequality goes in the opposite direction than above because the actual treatments of the initially treated groups represent a decrease in exposure with respect to always being treated. Accordingly, the planner may be interested in learning $\delta_- = E\left(\Delta_-\right).$

It may not always be possible to estimate δ_+. To see why, let us start by simplifying the expression of that parameter. For never treated groups ($F_{g,1} = T + 1$), $Y_{g,t}(\boldsymbol{D}_g) = Y_{g,t}(\boldsymbol{0})$ for every t. For groups initially untreated but treated at least once, ($2 \le F_{g,1} \le T$), under Assumption 2 $Y_{g,t}(\boldsymbol{D}_g) = Y_{g,t}(\boldsymbol{0})$ for $t < F_{g,1}$. Accordingly,

$$\delta_+ = E\left(\frac{\sum_{g:2\le F_{g,1}\le T}\sum_{t=F_{g,1}}^T N_{g,t}\beta^t (Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}))}{\sum_{g:2\le F_{g,1}\le T}\sum_{t=F_{g,1}}^T N_{g,t}\beta^t D_{g,t}}\right).$$

Thus, estimating δ_+ requires estimating the never-treated outcome $Y_{g,t}(\boldsymbol{0})$ of all the (g,t)s entering the previous equation. For that purpose, the DID estimator we propose below uses the fact that $Y_{g,t}(\boldsymbol{0}) = Y_{g,F_{g,1}-1}(\boldsymbol{0}) + Y_{g,t}(\boldsymbol{0}) - Y_{g,F_{g,1}-1}(\boldsymbol{0})$, and estimates $Y_{g,t}(\boldsymbol{0}) - Y_{g,F_{g,1}-1}(\boldsymbol{0})$ by the outcome evolution from $F_{g,1} - 1$ to t in groups never treated from period 1 to t. If there is no group never treated from period 1 to T, this strategy is not feasible at every t. Then, estimating δ_+ is not feasible under the common trends condition we introduced above, and would require imposing supplementary assumptions.

---

[3]Strictly speaking, the planner may rather be interested in learning $E\left(\Delta_+|\boldsymbol{D}\right)$, because $E(W((\boldsymbol{D}_1, ..., \boldsymbol{D}_G), g : F_{g,1} > 1)|\boldsymbol{D}) > E(W((\boldsymbol{0}, ..., \boldsymbol{0}), g : F_{g,1} > 1)|\boldsymbol{D})$ if and only if $E\left(\Delta_+|\boldsymbol{D}\right) > c$, while we do not have that $E(W((\boldsymbol{D}_1, ..., \boldsymbol{D}_G), g : F_{g,1} > 1)) > E(W((\boldsymbol{0}, ..., \boldsymbol{0}), g : F_{g,1} > 1))$ if and only if $\delta_+ > c$. It turns out that the estimator $\widehat{\delta}_+^{\text{tru}}$ we propose below is also unbiased conditional on $\boldsymbol{D}$, so it may also be used to learn $E\left(\Delta_+|\boldsymbol{D}\right)$.

In that case, we consider a truncated version of δ_+. Let $NT = \max_{g \in \{1,\ldots,G\}} F_{g,1} - 1$ denote the last period where there is still a group that has been untreated since period 1, and let

$$\delta_+^{\text{tru}} = E\left(\frac{\sum_{g:2 \leq F_{g,1} \leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t} \beta^t (Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}))}{\sum_{g:2 \leq F_{g,1} \leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t} \beta^t D_{g,t}}\right)$$

denote the truncated-at-NT version of δ_+, that only takes into account the effects and costs of treatment until period NT. δ_+^{tru} includes all the treatment effects that can be estimated under our DID assumptions. Let

$$\lambda_+^{\text{tru}} = \frac{\sum_{g:2 \leq F_{g,1} \leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t} \beta^t D_{g,t}}{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=F_{g,1}}^{T} N_{g,t} \beta^t D_{g,t}} \tag{2}$$

denote the proportion of treatment effects in δ_+ that are also in δ_+^{tru}. When λ_+^{tru} is close to 1, δ_+ and δ_+^{tru} are unlikely to differ by much. Formally, assume that for all (g,t), $\underline{y} \leq Y_{g,t}(\boldsymbol{0}) \leq \overline{y}$, where $\underline{y}$ and $\overline{y}$ are two real numbers. Then, $\delta_+ \in [\underline{\delta}_+, \overline{\delta}_+]$, with

$$\underline{\delta} = E\left[\lambda_+^{\text{tru}} \delta_+^{\text{tru}} + (1 - \lambda_+^{\text{tru}}) \frac{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=NT+1}^{T} N_{g,t} \beta^t (Y_{g,t}(\boldsymbol{D}_g) - \overline{y})}{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=F_{g,1}}^{T} N_{g,t} \beta^t D_{g,t}}\right], \tag{3}$$

$$\overline{\delta} = E\left[\lambda_+^{\text{tru}} \delta_+^{\text{tru}} + (1 - \lambda_+^{\text{tru}}) \frac{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=NT+1}^{T} N_{g,t} \beta^t (Y_{g,t}(\boldsymbol{D}_g) - \underline{y})}{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=F_{g,1}}^{T} N_{g,t} \beta^t D_{g,t}}\right]. \tag{4}$$

Similarly, if there is no group treated at all periods, δ_- cannot be estimated under the common trends condition we consider below. Then, let $AT = \max_{g \in \{1,\ldots,G\}} F_{g,0} - 1$ denote the last period where there is still a group that has been treated since period 1, and let

$$\delta_-^{\text{tru}} = E\left(\frac{\sum_{g:2 \leq F_{g,0} \leq AT} \sum_{t=F_{g,0}}^{AT} N_{g,t} \beta^t (Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{1}))}{\sum_{g:2 \leq F_{g,0} \leq AT} \sum_{t=F_{g,0}}^{AT} N_{g,t} \beta^t (D_{g,t} - 1)}\right)$$

denote the truncated-at-AT version of δ_-. Let also

$$\lambda_-^{\text{tru}} = \frac{\sum_{g:2 \leq F_{g,0} \leq AT} \sum_{t=F_{g,0}}^{AT} N_{g,t} \beta^t (D_{g,t} - 1)}{\sum_{g:2 \leq F_{g,0} \leq T} \sum_{t=F_{g,0}}^{T} N_{g,t} \beta^t (D_{g,t} - 1)} \tag{5}$$

denote the proportion of treatment effects in δ_- that are also in δ_-^{tru}. As above, δ_- and δ_-^{tru} are likely to be close when λ_-^{tru} is close to 1.

δ_+^{tru} and δ_-^{tru} can be interpreted as instrumental variable (IV) estimands measuring some average of the change in outcome created by a one-unit change in treatment. To simplify, let us first assume that $\beta = 1$. Let

$$N^+ = \#\{(i,g,t) : 2 \leq F_{g,1} \leq t \leq NT\},$$

$$\Delta_+^{ITT} = \frac{1}{N^+} \sum_{(i,g,t):2 \leq F_{g,1} \leq t \leq NT} (Y_{i,g,t}(\boldsymbol{D}_g) - Y_{i,g,t}(\boldsymbol{0})),$$

$$\Delta_+^{FS} = \frac{1}{N^+} \sum_{(i,g,t):2 \leq F_{g,1} \leq t \leq NT} D_{i,g,t}.$$

N^+ is the number of observations in (g,t) cells such that group g has switched from untreated to treated for the first time at or before t. As we allow for dynamic effects, the outcome of all those observations may be affected by that first switch, and may differ from their never treated outcome. Accordingly, Δ_+^{ITT} is the average difference between their observed and never-treated outcome. Δ_+^{FS} is the average treatment of the same observations. In staggered designs, $\Delta_+^{FS} = 1$, as groups switching from untreated to treated remain treated thereafter. Outside of staggered designs, $\Delta_+^{FS} < 1$ as some groups may revert to being untreated. If $\beta = 1$,

$$\delta_+^{\text{tru}} = E\left(\frac{\Delta_+^{ITT}}{\Delta_+^{FS}}\right).$$

δ_+^{tru} scales the average "intention-to-treat" effect of first switches on the outcome by their average "first-stage" effect on the treatment, thus ensuring it measures some average of the change in outcome created by a one-unit change in treatment. If $\beta < 1$, δ_+^{tru} has a similar interpretation, except that every (i,g,t) receives a weight proportional to β^t in the definition of the intention-to-treat and first-stage effects. δ_-^{tru} has a similar interpretation.

There are also special cases where δ_+^{tru} and δ_-^{tru} simply reduce to standard average treatment effects. For instance, if $T = 2$,

$$\delta_+^{\text{tru}} = E\left(\frac{1}{N_+} \sum_{(i,g):D_{i,g,2}>D_{i,g,1}} (Y_{i,g,t}(0,1) - Y_{i,g,t}(0,0))\right),$$

where $N_+ = \#\{(i,g) : D_{i,g,2} > D_{i,g,1}\}$. δ_+^{tru} is the average effect of having been treated rather than untreated at period 2, among all observations going from untreated to treated from period 1 to 2. Similarly, δ_-^{tru} is the average effect of having been treated rather than untreated at period 2, among all observations going from treated to untreated from period 1 to 2:

$$\delta_-^{\text{tru}} = E\left(\frac{1}{N_-} \sum_{(i,g):D_{i,g,2}<D_{i,g,1}} (Y_{i,g,t}(1,1) - Y_{i,g,t}(1,0))\right),$$

where $N_- = \#\{(i,g) : D_{i,g,2} < D_{i,g,1}\}$. Then, δ_+^{tru} and δ_-^{tru} are similar to the switcher's average treatment effect considered in de Chaisemartin and D'Haultfœuille (2020). Another special case where δ_+^{tru} may reduce to a standard average treatment effect are staggered adoption designs, where groups can switch in but not out of the treatment:

$$D_{g,t} \geq D_{g,t-1} \text{ for all } g \text{ and } t \geq 2. \tag{6}$$

If (6) holds, $\beta = 1$, no groups are always treated ($1 < F_{g,1}$ for all g), and there are no dynamic effects ($Y_{g,t}(d_1,...,d_t) = Y_{g,t}(d_t)$), then

$$\delta_+^{\text{tru}} = E\left(\frac{1}{\sum_{(i,g,t)} D_{i,g,t}} \sum_{(i,g,t):D_{i,g,t}=1} (Y_{i,g,t}(1) - Y_{i,g,t}(0))\right),$$

10

so δ_+^{tru} is just the average treatment effect on the treated. Thus, δ_+^{tru} and δ_-^{tru} can be seen as generalizations of standard average treatment effect parameters to settings with multiple periods and dynamic effects, that aim to guide treatment choice in such settings.

3 Estimators and placebo estimators

3.1 Estimators

We start by proposing an unbiased estimator of δ_+^{tru}. Our estimator is a weighted average of difference-in-difference estimators. First, for any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$, let $N_{t,\ell}^1 = \sum_{g:F_{g,1}=t-\ell} N_{g,t}$ denote the number of observations in groups treated for the first time at period $t - \ell$. Let $N_t^{nt} = \sum_{g:F_{g,1}>t} N_{g,t}$ denote the number of observations in groups untreated from period 1 to t. We define

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

if $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$, and we let $\text{DID}_{+,t,\ell} = 0$ otherwise. $\text{DID}_{+,t,\ell}$ is the DID estimator comparing the outcome evolution from period $t - \ell - 1$ to t in groups treated for the first time in $t - \ell$ and in groups untreated from period 1 to t. Under Assumptions 4-5, the latter evolution is a counterfactual of the evolution that would have taken place in the former set of groups if it had not switched treatment for the first time ℓ periods ago. Thus, $\text{DID}_{+,t,\ell}$ is an unbiased estimator of the effect of that first switch, ℓ periods after it took place. In staggered designs, groups treated for the first time in $t - \ell$ remain treated till t, so $\text{DID}_{+,t,\ell}$ is an unbiased estimator of the cumulative effect of having been treated for $\ell+1$ periods, in groups reaching $\ell+1$ periods of treatment at period t. Without staggered adoption, some groups treated for the first time in $t - \ell$ may remain treated till t, while other groups may immediately go back to being untreated. Accordingly, $\text{DID}_{+,t,\ell}$ may estimate a mixture of different treatment effects.

Then, we define

$$\text{DID}_{+,t,\ell}^D = \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1}(D_{g,t} - D_{g,t-\ell-1}) - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}}(D_{g,t} - D_{g,t-\ell-1})$$

if $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{+,t,\ell}^D = 0$ otherwise. $\text{DID}_{+,t,\ell}^D$ is a DID estimator similar to $\text{DID}_{+,t,\ell}$, except that the outcome is replaced by the treatment. Given the definition of $F_{g,1}$, the previous equation simplifies to $\text{DID}_{+,t,\ell}^D = \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} D_{g,t}$.

Next, let $L_{nt} = NT - \min_{g:F_{g,1}\geq 2} F_{g,1}$ denote the number of time periods between the earliest date at which a group goes from untreated to treated and the last period at which a group has

been untreated all along. Note that $L_{nt} \geq 0$ under Point 1 of Assumption 3. For $\ell \in \{0, ..., L_{nt}\}$, we let

$$\text{DID}_{+,\ell} = \frac{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t \text{DID}_{+,t,\ell}}{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t}. \tag{7}$$

$\text{DID}_{+,\ell}$ is a weighted average of the $(\text{DID}_{+,t,\ell})_{t \in \{\ell+2,...,NT\}}$. We establish in the proof of Theorem 1 below that with $\beta = 1$, $\text{DID}_{+,\ell}$ is an unbiased estimator of the average effect of having switched from untreated to treated for the first time ℓ periods ago:

$$E[\text{DID}_{+,\ell}] = E\left[\frac{1}{N_\ell^1} \sum_{t=\ell+2}^{NT} \sum_{(i,g):F_{g,1}=t-\ell} Y_{i,g,t}(\boldsymbol{D}_g) - Y_{i,g,t}(\boldsymbol{0})\right],$$

with $N_\ell^1 = \sum_{t=\ell+2}^{NT} \sum_{(i,g)} \mathbb{1}\{F_{g,1} = t - \ell\}$.[4] Similarly, let

$$\text{DID}_{+,\ell}^D = \frac{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t \text{DID}_{+,t,\ell}^D}{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t}$$

be a weighted average of the $(\text{DID}_{+,t,\ell}^D)_{t \in \{\ell+2,...,NT\}}$.

Finally, our estimator of δ_+^{tru} is the ratio of weighted averages of the $\text{DID}_{+,\ell}$ and $\text{DID}_{+,\ell}^D$ estimators. For every $\ell \in \{0, ..., L_{nt}\}$, let

$$w_{+,\ell} = \frac{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t}{\sum_{\ell=0}^{L_{nt}} \sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t}.$$

$w_{+,\ell}$ is proportional to the discounted number of observations $\text{DID}_{+,\ell}$ and $\text{DID}_{+,\ell}^D$ apply to. Then, let

$$\widehat{\delta}_+^{\text{tru}} = \frac{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}}{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^D}.$$

Following the interpretation of δ_+^{tru} as an IV estimand outlined in the previous section, $\widehat{\delta}_+^{\text{tru}}$ may be interpreted as an IV estimator: its numerator estimates the average effect of first switches on the outcome, while its denominator estimates the average effect of first switches on the treatment. Notice that in staggered designs, $\text{DID}_{+,t,\ell}^D = 1$ for all (t, ℓ). Accordingly, $\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^D = 1$, so $\widehat{\delta}_+^{\text{tru}}$ is just a weighted average of the $\text{DID}_{+,\ell}$ estimators.

Then, we propose an unbiased estimator of δ_-^{tru}. First, for any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$, let $N_{t,\ell}^0 = \sum_{g:F_{g,0}=t-\ell} N_{g,t}$ denote the number of observations in groups untreated for the first time at period $t - \ell$. Let $N_t^{at} = \sum_{g:F_{g,0}>t} N_{g,t}$ denote the number of observations in groups treated from period 1 to t. We define

$$\text{DID}_{-,t,\ell} = \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}}(Y_{g,t} - Y_{g,t-\ell-1}) - \sum_{g:F_{g,0}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^0}(Y_{g,t} - Y_{g,t-\ell-1})$$

---

[4]Notice that for $\ell < \ell'$, $\text{DID}_{+,\ell}$ and $\text{DID}_{+,\ell'}$ do not apply to the same groups, as there may be fewer groups for which the treatment effect can be estimated ℓ' periods after their first switch than ℓ periods after it. If this is a concern, one may compute those estimators for the same sets of groups.

if $N_{t,\ell}^0 > 0$ and $N_t^{at} > 0$, and we let $\text{DID}_{-,t,\ell} = 0$ otherwise. $\text{DID}_{-,t,\ell}$ compares the outcome evolution from period $t-\ell-1$ to t in groups treated from period 1 to t and in groups untreated for the first time in $t-\ell$. Under Assumptions 6-7, the former evolution is a counterfactual of the evolution that would have taken place in the latter set of groups if it had not switched treatment for the first time ℓ periods ago. Thus, $\text{DID}_{-,t,\ell}$ is an unbiased estimator of the effect of that first switch, ℓ periods after it took place. As that switch corresponds to a decrease in exposure to treatment, the outcome evolution in the switching groups enters with a negative sign in $\text{DID}_{-,t,\ell}$, while that of the control groups enters with a positive sign. This ensures that $\text{DID}_{-,t,\ell}$ estimates the effect of an increase in exposure. Similarly, we define

$$\text{DID}_{-,t,\ell}^D = \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}}(D_{g,t} - D_{g,t-\ell-1}) - \sum_{g:F_{g,0}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^0}(D_{g,t} - D_{g,t-\ell-1})$$

if $N_{t,\ell}^0 > 0$ and $N_t^{at} > 0$, and we let $\text{DID}_{-,t,\ell}^D = 0$ otherwise.

Next, let $L_{at} = 1 - \min_{g:F_{g,0}\geq 2} F_{g,0}$ denote the number of time periods between the earliest date at which a group goes from treated to untreated and the last period at which a group has always been treated. Note that $L_{at} \geq 0$ under Point 2 of Assumption 3. For $\ell \in \{0, ..., L_{at}\}$, we let

$$\text{DID}_{-,\ell} = \frac{\sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t \text{DID}_{-,t,\ell}}{\sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t}. \tag{8}$$

Similarly, let

$$\text{DID}_{-,\ell}^D = \frac{\sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t \text{DID}_{-,t,\ell}^D}{\sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t}.$$

Finally, for every $\ell \in \{0, ..., L_{at}\}$, let

$$w_{-,\ell} = \frac{\sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t}{\sum_{\ell=0}^{L_{at}} \sum_{t=\ell+2}^{AT} N_{t,\ell}^0 \beta^t},$$

and

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

Theorem 1 *Suppose that Assumptions 1-2 hold.*

1. *If Point 1 of Assumption 3 and Assumptions 4-5 also hold, $E\left[\widehat{\delta}_+^{\text{tru}}\right] = \delta_+^{\text{tru}}$.*

2. *If Point 2 of Assumption 3 and Assumptions 6-7 also hold, $E\left[\widehat{\delta}_-^{\text{tru}}\right] = \delta_-^{\text{tru}}$.*

Theorem 1 shows that $\widehat{\delta}_+^{\text{tru}}$ and $\widehat{\delta}_-^{\text{tru}}$ are unbiased estimators of δ_+^{tru} and δ_-^{tru}, respectively. It also follows from Theorem 1 and (3)-(4) that when potential outcomes are bounded by $\underline{y}$ and $\overline{y}$, the

bounds $\underline{\delta}_+$ and $\overline{\delta}_+$ on δ_+ (and similarly for δ_-) can be unbiasedly estimated by

$$
\widehat{\underline{\delta}} = \lambda_+^{\text{tru}} \widehat{\delta}_+^{\text{tru}} + (1 - \lambda_+^{\text{tru}}) \frac{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=NT+1}^{T} N_{g,t} \beta^t (Y_{g,t} - \overline{y})}{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=F_{g,1}}^{T} N_{g,t} \beta^t D_{g,t}},
$$

$$
\widehat{\overline{\delta}} = \lambda_+^{\text{tru}} \widehat{\delta}_+^{\text{tru}} + (1 - \lambda_+^{\text{tru}}) \frac{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=NT+1}^{T} N_{g,t} \beta^t (Y_{g,t} - \underline{y})}{\sum_{g:2 \leq F_{g,1} \leq T} \sum_{t=F_{g,1}}^{T} N_{g,t} \beta^t D_{g,t}}.
$$

Finally, it follows from Theorem 1 that δ_+ and δ_- can sometimes be unbiasedly estimated. When there is at least one never-treated group ($NT = T$), $\delta_+ = \delta_+^{\text{tru}}$, so $\widehat{\delta}_+^{\text{tru}}$ is unbiased for δ_+. Similarly, when there is at least one always-treated group ($AT = T$), $\delta_- = \delta_-^{\text{tru}}$, so $\widehat{\delta}_-^{\text{tru}}$ is unbiased for δ_-.

When $\beta = 1$, both $\widehat{\delta}_+^{\text{tru}}$ and $\widehat{\delta}_-^{\text{tru}}$ estimate some average of the effect of increasing treatment by one unit. Then, one may average them, to form a potentially more precise estimator. Accordingly, we define the following averaged estimator:

$$
\widehat{\delta}^{\text{tru}} = w_+ \widehat{\delta}_+^{\text{tru}} + (1 - w_+) \widehat{\delta}_-^{\text{tru}},
$$

where

$$
w_+ = \frac{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^D}{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^D + \sum_{\ell=0}^{L_{at}} w_{-,\ell} \text{DID}_{-,\ell}^D}
$$

is a weight proportional to the "first stage" attached to $\widehat{\delta}_+^{\text{tru}}$. Similarly, if one wishes to obtain a more precise estimator of the effect of having switched treatment for the first time ℓ periods ago, one can average the $\text{DID}_{+,\ell}$ and $\text{DID}_{-,\ell}$ estimators as follows:

$$
\text{DID}_\ell = w_+^\ell \text{DID}_{+,\ell} + \left(1 - w_+^\ell\right) \text{DID}_{-,\ell},
$$

where

$$
w_+^\ell = \frac{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1}{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 + \sum_{t=\ell+2}^{AT} N_{t,\ell}^0}
$$

is a weight proportional to the number of observations $\text{DID}_{+,\ell}$ applies to.

3.2   Placebo estimators

We first propose placebo estimators of the assumptions underlying $\widehat{\delta}_+^{\text{tru}}$. First, for any $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$ and $t \in \{2\ell + 3, ..., T\}$, let

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

14

if $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{+,t,\ell}^{\text{pl}} = 0$ otherwise. $\text{DID}_{+,t,\ell}^{\text{pl}}$ compares the outcome evolution in groups treated for the first time in $t - \ell$ and in groups untreated from period 1 to t, as $\text{DID}_{+,t,\ell}$, 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 common trends holds for $\ell + 1$ periods.

Next, let $L_{nt}^{\text{pl}} = \max\{\ell : \exists g : F_{g,1} + \ell \le NT, F_{g,1} - \ell - 2 \ge 1\}$. L_{nt}^{pl} is the largest ℓ such that there is a group for which the effect of switching from untreated to treated for the first time ℓ periods ago can be estimated ($F_{g,1} + \ell \le NT$), and for which one can form a placebo estimator comparing that group's outcome evolution to that of the untreated groups over the $\ell + 1$ periods before that switch ($F_{g,1} - \ell - 2 \ge 1$). Under Point 1 of Assumption 3, $-1 \le L_{nt}^{\text{pl}} \le \lfloor \frac{NT-3}{2} \rfloor$. One may have $L_{nt}^{\text{pl}} = -1$, if all groups that switch from untreated to treated for the first time do so at period 2. Then, none of the placebos defined below can be computed, as one does not observe the outcome evolution of any group before it switches from untreated to treated. Outside of that special case, $L_{nt}^{\text{pl}} \ge 0$, so one can at least compute one of the $\text{DID}_{+,\ell}^{\text{pl}}$ estimators below.

Finally, we let

$$\text{DID}_{+,\ell}^{\text{pl}} = \frac{\sum_{t=2\ell+3}^{NT} N_{t,\ell}^1 \beta^t \text{DID}_{+,t,\ell}^{\text{pl}}}{\sum_{t=2\ell+3}^{NT} N_{t,\ell}^1 \beta^t}$$

if $\ell \le L_{nt}^{\text{pl}}$, and we let $\text{DID}_{+,\ell}^{\text{pl}} = 0$ otherwise. $\text{DID}_{+,\ell}^{\text{pl}}$ is a placebo estimator mimicking $\text{DID}_{+,\ell}$.

We then propose placebo estimators of the assumptions underlying $\hat{\delta}_{-}^{\text{tru}}$. First, for any $\ell \in \{0, ..., \lfloor \frac{T-3}{2} \rfloor\}$ and $t \in \{2\ell + 3, ..., T\}$, let

$$\text{DID}_{-,t,\ell}^{\text{pl}} = \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}}(Y_{g,t-2\ell-2} - Y_{g,t-\ell-1}) - \sum_{g:F_{g,0}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^0}(Y_{g,t-2\ell-2} - Y_{g,t-\ell-1})$$

if $N_{t,\ell}^0 > 0$ and $N_t^{at} > 0$, and let $\text{DID}_{-,t,\ell}^{\text{pl}} = 0$ otherwise. $\text{DID}_{-,t,\ell}^{\text{pl}}$ compares the outcome evolution in groups treated from period 1 to t and in groups going from treated to untreated in period $t - \ell$, as $\text{DID}_{-,t,\ell}$, but between periods $t - 2\ell - 2$ and $t - \ell - 1$ instead of $t - \ell - 1$ and t.

Next, let $L_{at}^{\text{pl}} = \max\{\ell : \exists g : F_{g,0} + \ell \le AT, F_{g,0} - \ell - 2 \ge 1\}$. L_{at}^{pl} is the largest ℓ such that there is a group for which the effect of having switched from treated to untreated for the first time ℓ periods ago can be estimated ($F_{g,0} + \ell \le NT$), and for which one can form a placebo estimator comparing that group's outcome evolution over the $\ell + 1$ periods before it got treated to that of always treated groups ($F_{g,1} - \ell - 2 \ge 1$). Under Point 2 of Assumption 3, $-1 \le L_{at}^{\text{pl}} \le \lfloor \frac{AT-3}{2} \rfloor$. One may have $L_{at}^{\text{pl}} = -1$, if all groups that switch from treated to untreated do so at period 2. Then, none of the placebos defined below can be computed, as one does not observe the outcome evolution of any group before it switches from treated to untreated. Outside of that special case, $L_{at}^{\text{pl}} \ge 0$, so one can at least compute one of the $\text{DID}_{-,\ell}^{\text{pl}}$ estimators below.

Finally, we let

$$\mathrm{DID}^{\mathrm{pl}}_{-,\ell} = \frac{\sum_{t=2\ell+3}^{AT} N^0_{t,\ell} \beta^t \mathrm{DID}^{\mathrm{pl}}_{-,t,\ell}}{\sum_{t=2\ell+3}^{AT} N^0_{t,\ell} \beta^t}$$

if $\ell \leq L^{\mathrm{pl}}_{at}$, and we let $\mathrm{DID}^{\mathrm{pl}}_{-,\ell} = 0$ otherwise. $\mathrm{DID}^{\mathrm{pl}}_{-,\ell}$ is a placebo estimator mimicking $\mathrm{DID}_{-,\ell}$.

Theorem 2 *Suppose that Assumptions 1-2 hold.*

1. *If Point 1 of Assumption 3 and Assumptions 4-5 also hold, for any $\ell \in \{0,...,\lfloor \frac{T-3}{2} \rfloor\}$,*
   $E\left[DID^{pl}_{+,\ell}\right] = 0$.

2. *If Point 2 of Assumption 3 and Assumptions 6-7 also hold, for any $\ell \in \{0,...,\lfloor \frac{T-3}{2} \rfloor\}$,*
   $E\left[DID^{pl}_{-,\ell}\right] = 0$.

Theorem 2 shows that $E\left[\mathrm{DID}^{\mathrm{pl}}_{+,\ell}\right] = 0$ (resp. $E\left[\mathrm{DID}^{\mathrm{pl}}_{-,\ell}\right] = 0$) is a testable implication of Assumptions 2, 4, and 5 (resp. Assumptions 2, 6, and 7). If $E\left[\mathrm{DID}^{\mathrm{pl}}_{-,\ell}\right] = 0$ is rejected but $E\left[\mathrm{DID}^{\mathrm{pl}}_{+,\ell}\right] = 0$ is not, that would indicate that the assumptions underlying $\widehat{\delta}^{\mathrm{tru}}_{-}$ are violated while those underlying $\widehat{\delta}^{\mathrm{tru}}_{+}$ are not.

The placebo estimators $\mathrm{DID}^{\mathrm{pl}}_{+,\ell}$ and $\mathrm{DID}^{\mathrm{pl}}_{-,\ell}$ we consider here do not exhaust all the testable implications of our assumptions. We focus on them, because when we reject $E\left[\mathrm{DID}^{\mathrm{pl}}_{+,\ell}\right] = 0$ (resp. $E\left[\mathrm{DID}^{\mathrm{pl}}_{-,\ell}\right] = 0$), the value of $\mathrm{DID}^{\mathrm{pl}}_{+,\ell}$ (resp. $\mathrm{DID}^{\mathrm{pl}}_{-,\ell}$) may be used to estimate $\mathrm{DID}_{+,\ell}$'s (resp. $\mathrm{DID}_{-,\ell}$'s) bias, under a condition stated below. Specifically, fix $\ell \in \{0,...,\min(\lfloor \frac{T-3}{2} \rfloor, L_{nt})\}$. Instead of Assumption 5, assume that for all (g,g',t) such that $t - \ell \geq 2$, $F_{g,1} = t - \ell$, and $F_{g',1} > t$,

$$E\left[Y_{g,t}(\mathbf{0}) - Y_{g,t-\ell-1}(\mathbf{0})\right] - E\left[Y_{g',t}(\mathbf{0}) - Y_{g',t-\ell-1}(\mathbf{0})\right]$$

does not depend on (g,g',t), meaning that the differential trend between groups treated for the first time in $t - \ell$ and groups untreated from period 1 to t does not vary over time and across groups. Under Assumption 5, the quantity in the previous display is equal to 0, so the previous condition is weaker than Assumption 5. Under this weaker condition, one can show that $-E[\mathrm{DID}^{\mathrm{pl}}_{+,\ell}]$ is equal to the bias of $\mathrm{DID}_{+,\ell}$, so the placebos can be used to estimate the bias of $\mathrm{DID}_{+,\ell}$ and, in turn, of $\widehat{\delta}^{\mathrm{tru}}_{+}$. Similarly, if, for all (g,g',t) such that $t - \ell \geq 2$, $F_{g,1} = t - \ell$ and $F_{g',1} > t$, the sign of

$$E\left[Y_{g,t}(\mathbf{0}) - Y_{g,t-\ell-1}(\mathbf{0})\right] - E\left[Y_{g',t}(\mathbf{0}) - Y_{g',t-\ell-1}(\mathbf{0})\right]$$

does not depend on (g,g',t), then the sign of the bias of $\mathrm{DID}_{+,\ell}$ is equal to the sign of $-E[\mathrm{DID}^{\mathrm{pl}}_{+,\ell}]$.

The "long-difference" placebos we propose here differ from the "first-difference" placebos we proposed in de Chaisemartin and D'Haultfœuille (2020) and that we extend in the Web Appendix to allow for dynamic effects. The first-difference placebo estimators of the assumptions

underlying $\widehat{\delta}_+^{\mathrm{tru}}$ compare the $t - 1 - k$ to $t - k$ outcome evolution in groups treated for the first time at period t and in groups untreated from period 1 to t, for $k \geq 1$. The long-difference placebos test if common trends holds over several periods, while the short-difference ones only test if it holds over pairs of consecutive periods. If treated and untreated groups follow different linear trends, differential trends will be larger, and easier to detect, over several periods than over two consecutive periods. Then, the long-difference placebos may lead to a more powerful test of common trends. As indicated above, they are also informative on the bias of $\mathrm{DID}_{+,\ell}$ under weaker conditions than Assumption 5.

On the other hand, the first-difference placebos may be useful to specifically test Assumption 2, the no-anticipation assumption. Assume for instance that those placebos are statistically insignificant for every $k > 1$, but the one for $k = 1$ is statistically significant. This indicates that groups switching and not switching treatment are on common trends, except between the two last periods before the switching groups switch. This may be interpreted as evidence that Assumptions 4-5 and 6-7 hold, but Assumption 2 fails. Groups are on common trends, but the statistically significant placebo at $k = 1$ indicates that groups' period-$t + 1$ treatment affects their period-t outcome. In that case, group g's actual period-t treatment is $\tilde{D}_{g,t} = D_{g,t+1}$, and one can just compute the estimators above with $\tilde{D}_{g,t}$ instead of $D_{g,t}$.

3.3 Extensions

We now briefly review some of the extensions in our online Appendix. First, we show how covariates $X_{g,t}$ can be included in the estimation. Then, we replace Assumption 5 by the requirement that $E[Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}) - (X_{g,t} - X_{g,t-1})'\theta_0]$ does not vary across g, meaning that groups can experience differential trends provided those differential trends are fully explained by changes in their covariates. The idea, then, is to estimate θ_0 by regressing $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ in the sample of not-yet-treated cells and define our estimators as above, replacing $Y_{g,t} - Y_{g,t-1}$ by $Y_{g,t} - Y_{g,t-1} - (X_{g,t} - X_{g,t-1})'\widehat{\theta}_0$.

Second, we show how one can allow for different trends across sets of groups. This mirrors the common practice of, e.g., allowing for state-specific trends in county-level two-way fixed effect regressions. Third, we extend our analysis to non-binary treatments. Fourth, we examine the benefits of ruling out the effect of past treatments beyond k lags, for some $k \geq 0$. Doing so provides a solution to the initial conditions problem that treatments prior to the start of the panel $(D_{g,t})_{t \leq 0}$ may not be known and may affect potential outcomes. Finally, we show that our results extend to fuzzy designs, provided some groups are fully untreated at the start of the panel.

4 Estimating an additive dynamic effect model

The approach proposed so far evaluates ex-post the effects of the policies that took place during the panel. Instead, one may be interested in using the panel to evaluate ex-ante future policies. Such an exercise can only be conducted under a restriction on how the treatment effect can vary over time. Otherwise, the panel cannot inform us on future treatment effects. In this section, we consider an assumption requiring that treatment effects be constant over time, we show that various treatment effect parameters can be estimated under that assumption, and we finally discuss how the corresponding estimators can be used for ex-ante policy evaluation.

We start by considering the following assumption.

Assumption 8 *(Additive dynamic effect model with constant effects over time) For all (g, t), there are real numbers $(\alpha_{g,k})_{k \in \{0,...,T-1\}}$ such that for all $(d_1, ..., d_t) \in \{0, 1\}^t$,*

$$Y_{g,t}(\boldsymbol{d}) = Y_{g,t}(\boldsymbol{0}) + \sum_{k=0}^{t-1} \alpha_{g,k} d_{t-k}.$$

$\alpha_{g,k}$ represents the effect of group g's treatment k periods ago on her current outcome. Assumption 8 requires that those effects be constant over time, though they can vary across groups. Assumption 8 also rules out interaction effects between a group's past and current treatments.

For every $g : 1 < F_{g,1} \leq NT$, and for every $t \in \{F_{g,1}, ..., NT\}$, let

$$\text{DID}_{g,t} = Y_{g,t} - Y_{g,F_{g,1}-1} - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} (Y_{g,t} - Y_{g,F_{g,1}-1}).$$

$\text{DID}_{g,t}$ compares the outcome evolution from period $F_{g,1} - 1$ to t in group g and in groups untreated from period 1 to t. Then, let $\text{DID}_g = (\text{DID}_{g,F_{g,1}}, ..., \text{DID}_{g,NT})'$, let $\alpha_g = (\alpha_{g,0}, ..., \alpha_{g,NT-F_{g,1}})'$ and let $\boldsymbol{D}_{g,F_{g,1}}$ be a $(NT - F_{g,1} + 1) \times (NT - F_{g,1} + 1)$ lower triangular matrix whose first line is the vector $(D_{g,F_{g,1}}, 0, ..., 0)$, whose second line is the vector $(D_{g,F_{g,1}+1}, D_{g,F_{g,1}}, 0, ..., 0)$,..., and whose last line is the vector $(D_{g,NT}, D_{g,NT-1}, ..., D_{g,F_{g,1}})$.

Similarly, for every $g : 1 < F_{g,0} \leq AT$, and for every $t \in \{F_{g,0}, ..., AT\}$, let

$$\text{DID}_{g,t} = Y_{g,t} - Y_{g,F_{g,0}-1} - \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}} (Y_{g,t} - Y_{g,F_{g,0}-1}).$$

$\text{DID}_{g,t}$ compares the outcome evolution from period $F_{g,0} - 1$ to t in group g and in groups treated from period 1 to t. Then, let $\text{DID}_g = (\text{DID}_{g,F_{g,0}}, ..., \text{DID}_{g,AT})'$, let $\alpha_g = (\alpha_{g,0}, ..., \alpha_{g,AT-F_{g,0}})'$ and let $\boldsymbol{D}_{g,F_{g,0}}$ be a $(AT - F_{g,0} + 1) \times (AT - F_{g,0} + 1)$ lower triangular matrix whose first line is the vector $(D_{g,F_{g,0}} - 1, 0, ..., 0)$, whose second line is the vector $(D_{g,F_{g,0}+1} - 1, D_{g,F_{g,0}} - 1, 0, ..., 0)$,..., and whose last line is the vector $(D_{g,AT} - 1, D_{g,AT-1} - 1, ..., D_{g,F_{g,0}} - 1)$.

Theorem 3 *Suppose that Assumptions 1-2 and 8 hold.*

1. *If Point 1 of Assumption 3 and Assumptions 4-5 also hold, for every $g : 1 < F_{g,1} \leq NT$,*
$$\alpha_g = E\left(\boldsymbol{D}_{g,F_{g,1}}^{-1} DID_g\right).$$

2. *If Point 2 of Assumption 3 and Assumptions 6-7 also hold, for every $g : 1 < F_{g,0} \leq AT$,*
$$\alpha_g = E\left(\boldsymbol{D}_{g,F_{g,0}}^{-1} DID_g\right).$$

It directly follows from Theorem 3 that averages of the $\alpha_{g,k}$s can also be unbiasedly estimated. For all $k \in \{0, ..., \max(L_{nt}, L_{at})\}$, let $\mathcal{G}_k = \{g : 1 < F_{g,1} + k \leq NT\} \cup \{g : 1 < F_{g,0} + k \leq AT\}$ be the set of groups for which $\alpha_{g,k}$ can be unbiasedly estimated, let $N_{\mathcal{G}_k} = \sum_{g \in \mathcal{G}_k} \sum_{t=1}^T N_{g,t}$ be the number of observations in those groups, let $\widehat{\alpha}_{g,k}$ denote the $k+1$th coordinate of $\boldsymbol{D}_{g,F_{g,1}}^{-1} DID_g$ (resp. $\boldsymbol{D}_{g,F_{g,0}}^{-1} DID_g$) for groups in $\mathcal{G}_k$ untreated (resp. treated) at period 1, and let

$$\alpha_k = \frac{1}{N_{\mathcal{G}_k}} \sum_{g \in \mathcal{G}_k} \left(\sum_{t=1}^T N_{g,t}\right) \alpha_{g,k}$$

$$\widehat{\alpha}_k = \frac{1}{N_{\mathcal{G}_k}} \sum_{g \in \mathcal{G}_k} \left(\sum_{t=1}^T N_{g,t}\right) \widehat{\alpha}_{g,k}.$$

As points 1 and 2 in Theorem 3 also hold conditional on $\boldsymbol{D}$, one has

$$\alpha_k = E\left(\widehat{\alpha}_k\right).$$

The estimators $(\widehat{\alpha}_k)_{k \in \{0,...,\max(L_{nt}, L_{at})\}}$ are similar to those empirical researchers routinely compute when they regress the outcome on the contemporaneous and lagged treatments, together with group and time effects (see e.g. Autor, 2003). A key difference, however, is that the estimators $(\widehat{\alpha}_k)_{k \in \{0,...,\max(L_{nt}, L_{at})\}}$ are robust to dynamic effects and to heterogeneous effects across groups.

The estimators $(\widehat{\alpha}_k)_{k \in \{0,...,\max(L_{nt}, L_{at})\}}$ may be used for ex-ante evaluation of future policies. To simplify the discussion, let us assume that the number of units in each group does not depend on t: for every (g,t), $N_{g,t} = N_g$. We start by assuming that the planner wants to know if groups should be treated or untreated at period $T+1$. Under Assumption 8, the difference between the average expected outcome at period $T+1$ of groups belonging to $\mathcal{G}_0$ if they are treated in period $T+1$ and if they are untreated is equal to α_0. Accordingly, under (1), treating those groups at period $T+1$ is welfare improving if and only if $\alpha_0 > c$. The planner may then use $\widehat{\alpha}_0$ to estimate if treating those groups at $T+1$ is welfare improving. The planner could also consider more complicated policy decisions. She may for instance want to know which of the four policies would lead to the highest welfare in periods $T+1$ and $T+2$: $D_{g,T+1} = D_{g,T+2} = 1$, $D_{g,T+1} = 1, D_{g,T+2} = 0$, $D_{g,T+1} = 0, D_{g,T+2} = 1$, or $D_{g,T+1} = 0, D_{g,T+2} = 0$. Let

$$\alpha_0^{\text{bal}} = \frac{1}{\sum_{g \in \mathcal{G}_1} N_g} \sum_{g \in \mathcal{G}_1} N_g \alpha_{g,0}$$

be the average effect of the current treatment across groups for which the effect of the first lag of the treatment can be estimated. Under Assumption 8 and (1), treating groups belonging to $\mathcal{G}_1$ at $T+1$ and $T+2$ is optimal if and only if

$$\alpha_0^{\text{bal}} + \beta(\alpha_0^{\text{bal}} + \alpha_1) - (1+\beta)c \geq \max\left[\alpha_0^{\text{bal}} + \beta\alpha_1 - c, \beta\alpha_0^{\text{bal}} - \beta c, 0\right].$$

Similarly, treating groups belonging to $\mathcal{G}_1$ at $T+1$ but not at $T+2$ is optimal if and only if

$$\alpha_0^{\text{bal}} + \beta\alpha_1 - c \geq \max\left[\alpha_0^{\text{bal}} + \beta(\alpha_0^{\text{bal}} + \alpha_1) - (1+\beta)c, \beta\alpha_0^{\text{bal}} - \beta c, 0\right],$$

etc. One can estimate α_0^{bal} and α_1 to estimate the optimal policy, for groups belonging to $\mathcal{G}_1$.

A few comments on this estimation of the optimal policy are in order. First, notice that the longer the policy horizon one considers, the smaller the set of groups for which the optimal policy can be estimated. Second, we have considered homogeneous policies where each group receives the same treatments. This is because estimating optimal group-level policies would require using the noisy group-level estimators $\widehat{\alpha}_{g,k}$, which may yield poor results. Finally, performing inference on, say, the average outcome under the estimated optimal policy may give rise to a winner's curse phenomenon similar to that studied by Andrews et al. (2019). The solutions proposed therein may also apply here, though showing it goes beyond the scope of this paper.

On top of the assumption that treatment effects do not vary over time, which may be hard to avoid if one wants to use the panel to do ex-ante evaluation, Assumption 8 also rules out interaction effects between the current and past treatments. The point identification results in Theorem 3 would break down without that assumption. Specifically, Point 1 (resp. 2) of Theorem 3 shows that for groups untreated (resp. treated) at period 1, $NT - F_{g,1} + 1$ (resp. $AT - F_{g,0} + 1$) causal effects are just identified by a linear system of $NT - F_{g,1} + 1$ (resp. $AT - F_{g,0} + 1$) equations and unknowns under Assumption 8. Accordingly, if one were to allow for interaction effects, one would need to restrict effects heterogeneity across groups, at least with the DID estimation strategy we consider.

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.

Andrews, I., Kitagawa, T. and McCloskey, A. (2019), Inference on winners, Technical report, National Bureau of Economic Research.

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.

Bojinov, I., Rambachan, A. and Shephard, N. (2020), 'Panel experiments and dynamic causal effects: A finite population perspective', *arXiv preprint arXiv:2003.09915* .

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

Callaway, B. and Sant'Anna, P. H. (2020), 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* **110**(9), 2964–2996.

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

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

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.

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.

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

A   Appendix: proofs

A.1   Proof of Theorem 1

Proof of Point 1

First, by Assumption 5, for all $t \geq 2$ there is a real number ψ_t such that $\psi_t = E(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0}))$ for all g. Then, for all g and all $t \geq \ell + 2$,

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

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

$$
\begin{aligned}
&E\left(\mathrm{DID}_{+,t,\ell} | \boldsymbol{D}\right) \\
&= \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t} - Y_{g,t-\ell-1} | \boldsymbol{D}\right) - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} E\left(Y_{g,t} - Y_{g,t-\ell-1} | \boldsymbol{D}\right) \\
&= \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\mathbf{0}) | \boldsymbol{D}\right) \\
&\quad + \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-\ell-1}(\mathbf{0}) | \boldsymbol{D}\right) \\
&\quad - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-\ell-1}(\mathbf{0}) | \boldsymbol{D}\right) \\
&= \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\mathbf{0}) | \boldsymbol{D}\right). \tag{10}
\end{aligned}
$$

The first equality follows from the definition of $\mathrm{DID}_{+,t,\ell}$, and $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$. The second equality follows from Assumption 2. The third equality follows from Assumption 4 and (9).

By definition of NT and because there exists $(g, g') \in \{1, ..., G\}^2$ such that $1 < F_{g,1} < F_{g',1}$, $N_t^{nt} > 0$ for all $2 \leq t \leq NT$. We adopt the convention that a sum over an empty set is equal to 0. Then, Equation (10) implies

$$N_{t,\ell}^1 E\left(\mathrm{DID}_{+,t,\ell} | \boldsymbol{D}\right) = \sum_{g:F_{g,1}=t-\ell} N_{g,t} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\mathbf{0}) | \boldsymbol{D}\right).$$

Moreover, by definition of L_{nt}, we have $\sum_{t=\ell+2}^{NT} \beta^t N_{t,\ell}^1 > 0$ for all $\ell \in \{0, ..., L_{nt}\}$. Then,

$$E\left(\mathrm{DID}_{+,\ell} | \boldsymbol{D}\right) = \frac{1}{\sum_{t=\ell+2}^{NT} \beta^t N_{t,\ell}^1} \sum_{t=\ell+2}^{NT} \sum_{g:F_{g,1}=t-\ell} \beta^t N_{g,t} E[Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\mathbf{0}) | \boldsymbol{D}]. \tag{11}$$

Using again that $1 < F_{g,1} < F_{g',1}$ for some (g, g'), we have

$$\sum_{g:2 \leq F_{g,1} \leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t} \beta^t D_{g,t} > 0.$$

23

As a result,

$$\frac{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} E\left(\text{DID}_{+,\ell}|\boldsymbol{D}\right)}{\sum_{\ell=0}^{L_{nt}} w_{+,\ell}\text{DID}_{+,\ell}^D} = \frac{\sum_{\ell=0}^{L_{nt}} \sum_{t=\ell+2}^{NT} \sum_{g:F_{g,1}=t-\ell} N_{g,t}\beta^t E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0})|\boldsymbol{D}\right)}{\sum_{\ell=0}^{L_{nt}} \sum_{t=\ell+2}^{NT} \sum_{g:F_{g,1}=t-\ell} N_{g,t}\beta^t D_{g,t}}$$

$$= \frac{\sum_{g:2\leq F_{g,1}\leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t}\beta^t E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0})|\boldsymbol{D}\right)}{\sum_{g:2\leq F_{g,1}\leq NT} \sum_{t=F_{g,1}}^{NT} N_{g,t}\beta^t D_{g,t}}. \quad (12)$$

The first equality follows from (11) and the definition of $w_{+,\ell}$. The second equality follows after some algebra.

Finally, Point 1 follows from (12) and the law of iterated expectations.

Proof of Point 2

The proof of Point 2 is similar to that of Point 1 and is therefore omitted.

A.2   Proof of Theorem 2

Following the same steps as those used to obtain (10), we get, whenever $N_{t,\ell}^1 > 0$, $N_t^{nt} > 0$,

$$E[\text{DID}_{+,t,\ell}^{\text{pl}}|\boldsymbol{D}] = \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t-\ell-1}(\boldsymbol{D}_g) - Y_{g,t-\ell-1}(\boldsymbol{0})|\boldsymbol{D}\right)$$

$$= 0,$$

where the second equality follows since $F_{g,1} = t-\ell$ and Assumption 2 imply that $Y_{g,t-\ell-1}(\boldsymbol{D}_g) = Y_{g,t-\ell-1}(\boldsymbol{0})$. Then, Point 1 follows using the same reasoning as that used to obtain (11). Point 2 can be obtained similarly.

A.3   Proof of Theorem 3

We only prove Point 1, the proof of Point 2 is symmetric. Using the same steps as those used to prove (10), one can show that for every $g : 1 < F_{g,1} \leq NT$, and for every $t \in \{F_{g,1}, ..., NT\}$

$$E\left(\text{DID}_{g,t}|\boldsymbol{D}\right) = E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0})|\boldsymbol{D}\right).$$

Under Assumption 8, this implies that

$$E\left(\text{DID}_{g,t}|\boldsymbol{D}\right) = \sum_{k=0}^{t-F_{g,1}} \alpha_{g,k} D_{g,t-k},$$

which in turn implies that

$$E\left(\text{DID}_g|\boldsymbol{D}\right) = \boldsymbol{D}_{g,F_{g,1}}\alpha_g.$$

The result follows from the fact that $\boldsymbol{D}_{g,F_{g,1}}$ is a lower triangular matrix with 1s on the diagonal, so it is invertible, and from the law of iterated expectations.

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

Clément de Chaisemartin[*]   Xavier D'Haultfœuille[†]

December 22, 2020

Abstract

In this web appendix, we first consider extensions of the estimators described in Section 3.1. Second, we discuss first-difference placebo estimators, which differ from those considered in Section 3.2.

1 Extensions of our main estimators

1.1 Including covariates

Often times, researchers want to control for a vector of covariates $X_{g,t}$ in their estimation. In this section, we propose estimators controlling for covariates. To do so, we need to slightly modify our assumptions. Hereafter, we let $\boldsymbol{X}_g = (X'_{g,1}, ..., X'_{g,T})$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_G)$.

Assumption 9 *(Independence between groups, strong exogeneity and common trends with covariates) There is a vector θ_0 of same dimension as $X_{g,t}$ such that $\forall(g,t)$, $t \geq 2$,*

1. *$E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0})|\boldsymbol{D}, \boldsymbol{X}) = E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0})|\boldsymbol{D}_g, \boldsymbol{X}_g)$.*

2. *$E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0}) - (X_{g,t} - X_{g,t-1})'\theta_0|\boldsymbol{D}_g, \boldsymbol{X}_g) = E(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0}) - (X_{g,t} - X_{g,t-1})'\theta_0)$.*

3. *$E\left(Y_{g,t}(\boldsymbol{0}) - Y_{g,t-1}(\boldsymbol{0}) - (X_{g,t} - X_{g,t-1})'\theta_0\right)$ does not vary across g.*

---
[*]University of California at Santa Barbara, clementdechaisemartin@ucsb.edu
[†]CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr

Points 1 and 2 are the same as in Assumption 4, except that they include the covariates in the conditioning. Point 3 is a common trends condition, on the "residualized" never-treated potential outcome $Y_{g,t}(\mathbf{0}) - X'_{g,t}\theta_0$. In de Chaisemartin and D'Haultfœuille (2020), we showed that Assumption 9 underlies two-way fixed effects regressions with covariates. Assumption 9 requires that there exist θ_0 and λ_t such that

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

Then, Assumption 9 allows for the possibility that groups experience different evolutions of their never-treated outcome over time, but it requires that those differential evolutions be accounted for by a linear model in $X_{g,t} - X_{g,t-1}$, the change in a group's covariates. An interesting special case is when the control variables are group-specific linear trends. Then, Assumption 9 requires that

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

for some constants θ_{0g} and λ_t. From $t-1$ to t, the evolution of the never-treated outcome in group g should deviate from its group-specific linear trend θ_{0g} by an amount λ_t common to all groups. Then, Assumption 9 is a "common deviation from linear trends" assumption, which may be more plausible than the standard common trends assumption.

Assumption 10 *(Independence between groups, strong exogeneity and common trends with covariates for the always treated outcome) There is a vector θ_1 of same dimension as $X_{g,t}$ such that $\forall (g,t)$, $t \geq 2$,*

1. $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})|\mathbf{D}, \mathbf{X}) = E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})|\mathbf{D}_g, \mathbf{X}_g)$.

2. $E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}) - (X_{g,t} - X_{g,t-1})'\theta_1|\mathbf{D}_g, \mathbf{X}_g) = E(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}) - (X_{g,t} - X_{g,t-1})'\theta_1)$.

3. $E\left(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1}) - (X_{g,t} - X_{g,t-1})'\theta_1\right)$ *does not vary across g.*

Here as well, Assumption 10 allows for the possibility that groups experience different evolutions of their always-treated outcome over time, but it requires that those differential evolutions be accounted for by a linear model in $X_{g,t} - X_{g,t-1}$.

Let $\widehat{\theta}_0$ (resp. $\widehat{\theta}_1$) denote the coefficient of $X_{g,t} - X_{g,t-1}$ in the OLS regression of $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ and time fixed effects, in the sample of all (g,t) such that $F_{g,1} > t$ (resp. $F_{g,0} > t$). For any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$, let

$$\text{DID}^X_{+,t,\ell} = \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N^1_{t,\ell}}(Y_{g,t} - Y_{g,t-\ell-1} - (X_{g,t} - X_{g,t-\ell-1})'\widehat{\theta}_0)$$

$$- \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N^{nt}_t}(Y_{g,t} - Y_{g,t-\ell-1} - (X_{g,t} - X_{g,t-\ell-1})'\widehat{\theta}_0).$$

if $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{+,t,\ell}^X = 0$ otherwise. $\text{DID}_{+,t,\ell}^X$ is similar to $\text{DID}_{+,t,\ell}$, except that instead of the outcome evolution, it uses the part of that evolution that is not due to a change in covariates.

Similarly, let

$$\text{DID}_{-,t,\ell}^X = \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}}(Y_{g,t} - Y_{g,t-\ell-1} - (X_{g,t} - X_{g,t-\ell-1})'\widehat{\theta}_1)$$
$$- \sum_{g:F_{g,0}=t-\ell} \frac{N_{g,t}}{N_t^{\ell,-}}(Y_{g,t} - Y_{g,t-\ell-1} - (X_{g,t} - X_{g,t-\ell-1})'\widehat{\theta}_1)$$

if $N_t^{\ell,-} > 0$ and $N_t^{at} > 0$, and let $\text{DID}_{-,t,\ell}^X = 0$ otherwise.

Then, for all $\ell \in \{0, ..., L_{nt}\}$, we let

$$\text{DID}_{+,\ell}^X = \frac{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t \text{DID}_{+,t,\ell}^X}{\sum_{t=\ell+2}^{NT} N_{t,\ell}^1 \beta^t},$$

and we let

$$\widehat{\delta}_+^{tru,X} = \frac{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^X}{\sum_{\ell=0}^{L_{nt}} w_{+,\ell} \text{DID}_{+,\ell}^D}.$$

We define $\text{DID}_{-,\ell}^X$ and $\widehat{\delta}_-^{tru,X}$ similarly.

Theorem 4 *Suppose that Assumptions 1-2 hold.*

1. *If Point 1 of Assumption 3 and Assumption 9 also hold, $E\left[\widehat{\delta}_+^{tru,X}\right] = \delta_+^{tru}$.*

2. *If Point 2 of Assumption 3 and Assumption 10 also hold, $E\left[\widehat{\delta}_-^{tru,X}\right] = \delta_-^{tru}$.*

Theorem 4 shows that if groups experience different evolutions of their never- and always-treated outcome over time, one can still estimate δ_+^{tru} and δ_-^{tru}, provided that those differential evolutions are accounted for by a linear model in $X_{g,t} - X_{g,t-1}$. With respect to the estimators in Section 3, $\widehat{\delta}_+^{tru,X}$ (resp. $\widehat{\delta}_-^{tru,X}$) replace the outcome evolution of all the (g,t) cells entering in $\widehat{\delta}_+^{tru}$ (resp. $\widehat{\delta}_-^{tru}$) by the residuals from a regression of $Y_{g,t} - Y_{g,t-1}$ on $X_{g,t} - X_{g,t-1}$ and time fixed effects among all the (g,t) cells used as controls by $\widehat{\delta}_+^{tru}$ (resp. $\widehat{\delta}_-^{tru}$). One can follow the exact same steps to include covariates in the placebo estimators proposed in Section 3.2.

1.2 Allowing for different trends across sets of groups

In some cases, controlling for covariates may be insufficient to account for differences in trends between groups. Then, a common remedy in static or dynamic two-way fixed effect regressions

consists in including interactions between time dummies and dummies for sets of groups. For instance, if groups are US counties, one can allow for state-specific trends. A similar idea can be pursued in our context. Let us index sets of groups by $s \in \mathcal{S}$. In this set-up, we modify our assumptions as follows.

Assumption 11 *(Non-pathological design, in at least one set of groups) At least one of the two following statements hold:*

1. *There is at least one $s \in \mathcal{S}$ such that there exists $(g, g') \in s^2$ such that $1 < F_{g,1} < F_{g',1}$.*

2. *There is at least one $s \in \mathcal{S}$ such that there exists $(g, g') \in s^2$ such that $1 < F_{g,0} < F_{g',0}$.*

Assumption 11 requires that there is at least one set of groups with at least one group going from untreated to treated at a date where another group has been untreated all along, or at least one set of groups with at least one group going from treated to untreated at a date where another group has been treated all along. Let $\mathcal{S}_{+} = \{s \in \mathcal{S} : \exists (g, g') \in s^2 : 1 < F_{g,1} < F_{g',1}\}$ (resp. $\mathcal{S}_{-} = \{s \in \mathcal{S} : \exists (g, g') \in s^2 : 1 < F_{g,0} < F_{g',0}\}$) denote the sets of groups satisfying the first (resp. second) point of Assumption 11.

Assumption 12 *(Common trends with set-specific trends) For all $s \in \mathcal{S}$, $E\left(Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})\right)$ does not vary across $g \in s$.*

Assumption 13 *(Common trends with set-specific trends for the always treated outcome) For all $s \in \mathcal{S}$, $E\left(Y_{g,t}(\mathbf{1}) - Y_{g,t-1}(\mathbf{1})\right)$ does not vary across $g \in s$.*

Assumptions 12 (resp. 13) is a weakening of Assumption 5 (resp. 7), as it only requires that the never-treated (resp. always-treated) potential outcome of groups in the same set of groups follow the same evolution over time.

Under those assumptions, the treatment effects one can identify differ from those identified in Section 3.1, as only groups in the same set of groups can act as controls for each other. For instance, there may be sets of groups where all the groups switching treatment do so at the same date, in which case none of the treatment effects in those sets of groups can be identified. There may also be sets of groups where the last group switching treatment does so early in the panel, in which case long-run treatment effects cannot be estimated in those sets of groups. We modify the parameters δ_{+}^{tru} and δ_{-}^{tru} accordingly. First, let us define

$$NT^s = \max_{g \in s} F_{g,1} - 1,$$

$$AT^s = \max_{g \in s} F_{g,0} - 1.$$

Then, we define $\delta_+^{\text{tru},\mathcal{S}}$ and $\delta_-^{\text{tru},\mathcal{S}}$ as:

$$\delta_+^{\text{tru},\mathcal{S}} = E\left(\frac{\sum_{s\in\mathcal{S}_+}\sum_{g\in s:2\leq F_{g,1}\leq NT^s}\sum_{t=F_{g,1}}^{NT^s} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g)-Y_{g,t}(\boldsymbol{0}))}{\sum_{s\in\mathcal{S}_+}\sum_{g\in s:2\leq F_{g,1}\leq NT^s}\sum_{t=F_{g,1}}^{NT^s} N_{g,t}\beta^t D_{g,t}}\right),$$

$$\delta_-^{\text{tru},\mathcal{S}} = E\left(\frac{\sum_{s\in\mathcal{S}_-}\sum_{g\in s:2\leq F_{g,0}\leq AT^s}\sum_{t=F_{g,0}}^{AT^s} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g)-Y_{g,t}(\boldsymbol{1}))}{\sum_{s\in\mathcal{S}_-}\sum_{g\in s:2\leq F_{g,0}\leq AT^s}\sum_{t=F_{g,0}}^{AT^s} N_{g,t}\beta^t(D_{g,t}-1)}\right).$$

Note that if $\mathcal{S}_+ = \mathcal{S}$ and $NT^s = NT$ for all s (resp. $\mathcal{S}_- = \mathcal{S}$ and $AT^s = AT$ for all s), we have $\delta_+^{\text{tru},\mathcal{S}} = \delta_+^{\text{tru}}$ (resp. $\delta_-^{\text{tru},\mathcal{S}} = \delta_-^{\text{tru}}$).

The estimators of $\delta_+^{\text{tru},\mathcal{S}}$ and $\delta_-^{\text{tru},\mathcal{S}}$ follow the same logic as those of δ_+^{tru} and δ_-^{tru}, except that we first consider difference-in-differences within sets, before aggregating over the different sets s. Because the construction is similar for $\delta_+^{\text{tru},\mathcal{S}}$ and $\delta_-^{\text{tru},\mathcal{S}}$, we focus on $\delta_+^{\text{tru},\mathcal{S}}$ hereafter. Let $N_{t,\ell}^{1,s} = \sum_{g\in s:F_{g,1}=t-\ell} N_{g,t}$ and $N_t^{nt,s} = \sum_{g\in s:F_{g,1}>t} N_{g,t}$. Then, we define

$$\text{DID}_{+,t,\ell}^s = \sum_{g\in s:F_{g,1}=t-\ell}\frac{N_{g,t}}{N_{t,\ell}^{1,s}}(Y_{g,t}-Y_{g,t-\ell-1}) - \sum_{g\in s:F_{g,1}>t}\frac{N_{g,t}}{N_t^{nt,s}}(Y_{g,t}-Y_{g,t-\ell-1})$$

if $N_{t,\ell}^{1,s} > 0$ and $N_t^{nt,s} > 0$, and we let $\text{DID}_{+,t,\ell}^s = 0$ if $N_{t,\ell}^{1,s} = 0$ or $N_t^{nt,s} = 0$. Hence, $\text{DID}_{+,t,\ell}^s$ is defined as $\text{DID}_{+,t,\ell}$, but within set s. Then, we let $L_{nt}^s = NT^s - \min_{g\in s:F_{g,1}\geq 2} F_{g,1}$. Note that $L_{nt}^s \geq 0$ for all $s \in \mathcal{S}_+$. Then, for all $s \in \mathcal{S}_+$ and $\ell \in \{0,...,L_{nt}^s\}$, we let

$$\text{DID}_{+,\ell}^s = \frac{\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t\text{DID}_{+,t,\ell}^s}{\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t}.$$

We define similarly

$$\text{DID}_{+,t,\ell}^{D,s} = \sum_{g\in s:F_{g,1}=t-\ell}\frac{N_{g,t}}{N_{t,\ell}^{1,s}}(D_{g,t}-D_{g,t-\ell-1}) - \sum_{g\in s:F_{g,1}>t}\frac{N_{g,t}}{N_t^{nt,s}}(D_{g,t}-D_{g,t-\ell-1}),$$

$$\text{DID}_{+,\ell}^{D,s} = \frac{\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t\text{DID}_{+,t,\ell}^{D,s}}{\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t}.$$

Finally, we let

$$\widehat{\delta}_+^{\text{tru},\mathcal{S}} = \frac{\sum_{s\in\mathcal{S}_+}\omega^s\sum_{\ell=0}^{L_{nt}^s} w_{+,\ell}^s\text{DID}_{+,\ell}^s}{\sum_{s\in\mathcal{S}_+}\omega^s\sum_{\ell=0}^{L_{nt}^s} w_{+,\ell}^s\text{DID}_{+,\ell}^{D,s}},$$

where the weights are defined as

$$w_{+,\ell}^s = \frac{\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t}{\sum_{\ell=0}^{L_{nt}^s}\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t},$$

$$\omega^s = \frac{\sum_{\ell=0}^{L_{nt}^s}\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t}{\sum_{s\in\mathcal{S}_+}\sum_{\ell=0}^{L_{nt}^s}\sum_{t=\ell+2}^{NT^s} N_{t,\ell}^{1,s}\beta^t}.$$

We define $\widehat{\delta}_-^{\text{tru},\mathcal{S}}$ similarly.

Theorem 5 *Suppose that Assumptions 1-2 hold.*

1. *If Point 1 of Assumption 11, Assumptions 4 and 12 also hold, $E\left[\widehat{\delta}_{+}^{tru,\mathcal{S}}\right] = \delta_{+}^{tru,\mathcal{S}}$.*

2. *If Point 2 of Assumption 11, Assumptions 6 and 13 also hold, $E\left[\widehat{\delta}_{-}^{tru,\mathcal{S}}\right] = \delta_{-}^{tru,\mathcal{S}}$.*

1.3   Non-binary treatments

In this subsection, we consider the case where treatment is not binary but takes values in $\{0, ..., \overline{d}\}$ for $\overline{d} \geq 1$. For every $d \in \{0, ..., \overline{d}\}$, we let $\boldsymbol{d} = d \times \boldsymbol{1}$. $Y_{g,t}(\boldsymbol{d})$ is the potential outcome of group g at period t if she receives treatment d throughout. We generalize some of our assumptions as follows.

Assumption 14 *(Independent groups and strong exogeneity) $\forall d \in \{0, ..., \overline{d}\}$, $\forall t \geq 2$, and $\forall g \in \{1, ..., G\}$,*

1. $E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d})|\boldsymbol{D}_1, ..., \boldsymbol{D}_G) = E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d})|\boldsymbol{D}_g).$

2. $E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d})|\boldsymbol{D}_g) = E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d})).$

Assumption 15 *(Common trends) $\forall t \geq 2$, $E(Y_{g,t}(\boldsymbol{d}) - Y_{g,t-1}(\boldsymbol{d}))$ does not vary across g.*

Assumptions 14 and 15 generalize Assumptions 4 and 5 to all the $Y_{g,t}(\boldsymbol{d})$ potential outcomes.

In Assumption 16 below, we require that for at least one $d \in \{0, ..., \overline{d}\}$, there is at least one group leaving for the first time treatment d at a date where another group has had treatment d all along. For any $d \in \{0, ..., \overline{d}\}$ and $g \in \{1, ..., G\}$, let $F_{g,\neq d} = \min\{t : D_{g,t} \neq d\}$ denote the first date at which group g does not receive treatment d, with the convention that $F_{g,\neq d} = T + 1$ if group g always receives treatment d. Let also $I_{g,d} = 1\{\sum_{t=1}^{T} N_{g,t}\beta^t D_{g,t} > \sum_{t=1}^{T} N_{g,t}\beta^t d\}$ be an indicator for whether group g's actual treatments were more costly than keeping treatment d from period 1 to T. If $\beta = 1$, $I_{g,d}$ is equal to 1 if group g received on average more than d units of treatment from period 1 to T.

Assumption 16 *(Non-pathological design) At least one of the two following statements hold:*

1. $\exists d \in \{0, ..., \overline{d}\}$ *such that* $1 < F_{g,\neq d}I_{g,d} < F_{g',\neq d}$ *for some* $(g, g') \in \{1, ..., G\}^2.$

2. $\exists d \in \{0, ..., \overline{d}\}$ *such that* $1 < F_{g,\neq d}(1 - I_{g,d}) < F_{g',\neq d}$ *for some* $(g, g') \in \{1, ..., G\}^2.$

Then, let

$$\Delta_{+,d} = \frac{\sum_{(g,t):F_{g,\neq d}>1,I_{g,d}=1} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{d}))}{\sum_{(g,t):F_{g,\neq d}>1,I_{g,d}=1} N_{g,t}\beta^t(D_{g,t} - d)},$$

and let $\delta_{+,d} = E\left(\Delta_{+,d}\right)$. Under Equation (1), $W((\boldsymbol{D}_1,...,\boldsymbol{D}_G), g : F_{g,\neq d} > 1, I_{g,d} = 1) > W((\boldsymbol{d},...,\boldsymbol{d}), g : F_{g,\neq d} > 1, I_{g,d} = 1)$ if and only if $\Delta_{+,d} > c$, so the planner may be interested in learning $\delta_{+,d}$, to assess if in groups with treatment d at period 0 and with more costly treatments than keeping treatment d throughout, the actual treatments they received led to a welfare increase relative to the status quo. Similarly, let

$$\Delta_{-,d} = \frac{\sum_{(g,t):F_{g,\neq d}>1,I_{g,d}=0} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{d}))}{\sum_{(g,t):F_{g,\neq d}>1,I_{g,d}=0} N_{g,t}\beta^t(D_{g,t} - d)},$$

and let $\delta_{-,d} = E\left(\Delta_{-,d}\right)$.

It may not be possible to estimate $\delta_{+,d}$ and $\delta_{-,d}$. In that case, we consider truncated versions of those parameters. For every $d \in \{0,...,\overline{d}\}$, let $AT_d = \max_{g\in\{1,...,G\}} F_{g,\neq d} - 1$ denote the last period where there is still a group that has received treatment d since period 1. Let

$$\Delta_{+,d}^{\text{tru}} = \frac{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=1} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{d}))}{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=1} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)}$$

$$\Delta_{-,d}^{\text{tru}} = \frac{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=0} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{d}))}{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=0} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)},$$

and let $\delta_{+,d}^{\text{tru}} = E\left(\Delta_{+,d}^{\text{tru}}\right)$ and $\delta_{-,d}^{\text{tru}} = E\left(\Delta_{-,d}^{\text{tru}}\right)$ denote truncated-at-AT_d versions of $\delta_{+,d}$ and $\delta_{-,d}$.

When the treatment takes many values, it may be useful to aggregate the $\delta_{+,d}^{\text{tru}}$ (resp. $\delta_{-,d}^{\text{tru}}$) parameters. For every $d \in \{0,...,\overline{d}\}$, let

$$v_{+,d} = \frac{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=1} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)}{\sum_{d=0}^{\overline{d}} \sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=1} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)}$$

$$v_{-,d} = \frac{\sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=0} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)}{\sum_{d=0}^{\overline{d}} \sum_{g:2\leq F_{g,\neq d}\leq AT_d, I_{g,d}=0} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t}\beta^t(D_{g,t} - d)},$$

and let

$$\delta_+^{\text{tru}} = E\left(\sum_{d=0}^{\overline{d}} v_{+,d}\Delta_{+,d}^{\text{tru}}\right)$$

$$\delta_-^{\text{tru}} = E\left(\sum_{d=0}^{\overline{d}} v_{-,d}\Delta_{-,d}^{\text{tru}}\right).$$

The planner may be interested in learning δ_+^{tru} (resp. δ_-^{tru}), to assess if in groups with more (resp. less) costly treatments than the status-quo ones, the actual treatments they received until period AT_d led to a welfare increase. When $\beta = 1$, δ_+^{tru} and δ_-^{tru} can also be interpreted as the average change in outcome created by a one-unit change in treatment in those groups.

When $\bar{d} = 1$, δ_+^{tru} and δ_-^{tru} above are equal to the δ_+^{tru} and δ_-^{tru} parameters in the previous section, so they generalize those parameters to settings with non binary-treatments. A difference, however, is that δ_+^{tru} and δ_-^{tru} may now aggregate together effects of increases and decreases of the treatment. To see this, consider the following example. Suppose that $\bar{d} = 2$, $\beta = 1$, $N_{g,t} = 1$, and $T = 4$, and assume that all g receiving 1 unit of treatment at period 1, respectively receive 2, 2, and 0 units at period 2, 3, and 4. Suppose also that

$$Y_{g,t}(\boldsymbol{d}_g) = Y_{g,t}(\boldsymbol{0}) + \alpha_0 \mathbb{1}\{d_{g,t} > 0\} + \alpha_1 d_{g,t-1} + \varepsilon_{g,t}$$

for any $\boldsymbol{d}_g = (d_{g,1}, ..., d_{g,T})$. α_0 represents the effect of the current treatment, which is assumed to depend only on whether one receives a strictly positive number of treatment units. α_1 represents the effect of the lagged treatment, which is assumed to be linear in the number of units received. Then, $\delta_{+,1} = \Delta_{+,1} = 2\alpha_1 - \alpha_0$, which could be strictly negative even if α_1 and α_0 are both strictly positive. More generally, if non-monotonic treatment trajectories are frequent in the data, δ_+^{tru} and δ_-^{tru} could be of a different sign than the treatment's instantaneous and dynamic effects, even if those effects are all of the same sign. To avoid this issue, one can drop, in the definition of the parameters of interest and in the corresponding estimators, groups whose lowest treatment is strictly lower than their period-1 treatment, and whose highest treatment is strictly higher. That said, we keep our focus on δ_+^{tru} and δ_-^{tru} here, because they have a simpler interpretation in terms of policy evaluation than those other parameters.

We propose unbiased estimators of δ_+^{tru} and δ_-^{tru}. First, for any $d \in \{0, ..., \bar{d}\}$, $\ell \in \{0, ..., T-2\}$, and $t \in \{\ell+2, ..., T\}$, let $N_{t,\ell}^{\neq d,+} = \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=1} N_{g,t}$ (resp. $N_{t,\ell}^{\neq d,-} = \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=0} N_{g,t}$) be the number of observations in groups leaving treatment d for the first time at period $t-\ell$ and with more (resp. less) costly treatments than the status quo treatments $\boldsymbol{d}$. Let $N_t^{=d} = \sum_{g:F_{g,\neq d}>t} N_{g,t}$ denote the number of observations in groups with treatment d from period 1 to t. We let

$$\text{DID}_{+,d,t,\ell} = \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=1} \frac{N_{g,t}}{N_{t,\ell}^{\neq d,+}}(Y_{g,t} - Y_{g,t-\ell-1}) - \sum_{g:F_{g,\neq d}>t} \frac{N_{g,t}}{N_t^{=d}}(Y_{g,t} - Y_{g,t-\ell-1})$$

$$\text{DID}_{+,d,t,\ell}^D = \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=1} \frac{N_{g,t}}{N_{t,\ell}^{\neq d,+}}(D_{g,t} - D_{g,t-\ell-1}) - \sum_{g:F_{g,\neq d}>t} \frac{N_{g,t}}{N_t^{=d}}(D_{g,t} - D_{g,t-\ell-1})$$

if $N_{t,\ell}^{\neq d,+} > 0$ and $N_t^{=d} > 0$, and let $\text{DID}_{+,d,t,\ell} = \text{DID}_{+,d,t,\ell}^D = 0$ otherwise. $\text{DID}_{+,d,t,\ell}$ compares the outcome evolution from period $t-\ell-1$ to t in groups leaving treatment d for the first time in $t-\ell$ and with more costly treatments than the status quo treatments $\boldsymbol{d}$, and in groups with treatment d from period 1 to t. Under Assumptions 14-15, the latter evolution is a counterfactual of the evolution that would have taken place in the former set of groups if it had not left treatment d for the first time ℓ periods ago. $\text{DID}_{+,d,t,\ell}^D$ is similar to $\text{DID}_{+,d,t,\ell}$, except that the outcome is

replaced by the treatment. We let

$$
\mathrm{DID}_{-,d,t,\ell} = \sum_{g:F_{g,\neq d}>t} \frac{N_{g,t}}{N_t^{=d}}(Y_{g,t}-Y_{g,t-\ell-1}) - \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=0} \frac{N_{g,t}}{N_{t,\ell}^{\neq d,-}}(Y_{g,t}-Y_{g,t-\ell-1})
$$

$$
\mathrm{DID}^D_{-,d,t,\ell} = \sum_{g:F_{g,\neq d}>t} \frac{N_{g,t}}{N_t^{=d}}(D_{g,t}-D_{g,t-\ell-1}) - \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=0} \frac{N_{g,t}}{N_{t,\ell}^{\neq d,-}}(D_{g,t}-D_{g,t-\ell-1})
$$

if $N_{t,\ell}^{\neq d,-}>0$ and $N_t^{=d}>0$, and let $\mathrm{DID}_{-,d,t,\ell}=\mathrm{DID}^D_{-,d,t,\ell}=0$ otherwise. $\mathrm{DID}_{-,d,t,\ell}$ and $\mathrm{DID}^D_{-,d,t,\ell}$ have a similar interpretation as $\mathrm{DID}_{+,d,t,\ell}$ and $\mathrm{DID}^D_{+,d,t,\ell}$, but apply to groups with less costly treatments than the status quo.

Next, let $L_{+,d} = AT_d - \min_{g:F_{g,\neq d}\geq 2,I_{g,d}=1} F_{g,\neq d}$ (resp. $L_{-,d} = AT_d - \min_{g:F_{g,\neq d}\geq 2,I_{g,d}=0} F_{g,\neq d}$) denote the number of time periods between the earliest date at which a group with treatments more (resp. less) costly than $\boldsymbol{d}$ leaves treatment d and the last period at which a group has had treatment d all along. For $\ell \in \{0,...,L_{+,d}\}$, we let

$$
\mathrm{DID}_{+,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t \mathrm{DID}_{+,d,t,\ell}}{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t}
$$

$$
\mathrm{DID}^D_{+,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t \mathrm{DID}^D_{+,d,t,\ell}}{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t},
$$

and we let

$$
\widehat{\delta}^{\mathrm{tru}}_{+,d} = \frac{\sum_{\ell=0}^{L_{+,d}} w_{+,d,\ell}\mathrm{DID}_{+,d,\ell}}{\sum_{\ell=0}^{L_{+,d}} w_{+,d,\ell}\mathrm{DID}^D_{+,d,\ell}},
$$

where

$$
w_{+,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t}{\sum_{\ell=0}^{L_{+,d}} \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+}\beta^t}. \tag{13}
$$

Similarly, for $\ell \in \{0,...,L_{-,d}\}$, we let

$$
\mathrm{DID}_{-,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t \mathrm{DID}_{-,d,t,\ell}}{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t}
$$

$$
\mathrm{DID}^D_{-,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t \mathrm{DID}^D_{-,d,t,\ell}}{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t},
$$

and we let

$$
\widehat{\delta}^{\mathrm{tru}}_{-,d} = \frac{\sum_{\ell=0}^{L_{-,d}} w_{-,d,\ell}\mathrm{DID}_{-,d,\ell}}{\sum_{\ell=0}^{L_{-,d}} w_{-,d,\ell}\mathrm{DID}^D_{-,d,\ell}},
$$

where

$$
w_{-,d,\ell} = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t}{\sum_{\ell=0}^{L_{-,d}} \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-}\beta^t}.
$$

Finally, we let

$$\widehat{\delta}_+^{\text{tru}} = \sum_{d=0}^{\overline{d}} v_{+,d} \widehat{\delta}_{+,d}^{\text{tru}}$$

$$\widehat{\delta}_-^{\text{tru}} = \sum_{d=0}^{\overline{d}} v_{-,d} \widehat{\delta}_{-,d}^{\text{tru}}.$$

Theorem 6 *Suppose that Assumptions 1-2 and 14-15 hold.*

1. *If Point 1 of Assumption 16 also holds, $E\left[\widehat{\delta}_+^{\text{tru}}\right] = \delta_+^{\text{tru}}$.*

2. *If Point 2 of Assumption 16 also holds, $E\left[\widehat{\delta}_-^{\text{tru}}\right] = \delta_-^{\text{tru}}$.*

Finally, one may be interested in computing estimators of the effect of having switched treatment for the first time ℓ periods ago, irrespective of the baseline value one switched from. For any $\ell \in \{0, ..., \max_{d \in \{0,...,\overline{d}\}}(\max(L_{+,d}, L_{-,d}))\}$, we define such estimators as

$$\text{DID}_\ell = \sum_{d=0}^{\overline{d}} v_{+,d}^\ell \text{DID}_{+,d,\ell} + v_{-,d}^\ell \text{DID}_{-,d,\ell}, \tag{14}$$

where

$$v_{+,d}^\ell = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+} \beta^t}{\sum_{d=0}^{\overline{d}} \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+} \beta^t + \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-} \beta^t}$$

$$v_{-,d}^\ell = \frac{\sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-} \beta^t}{\sum_{d=0}^{\overline{d}} \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,+} \beta^t + \sum_{t=\ell+2}^{AT_d} N_{t,\ell}^{\neq d,-} \beta^t}.$$

1.4   Ruling out lagged treatment effects

Up to now, we have made no restriction on the effect of past treatment. 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), all t, and all $(d_1, ..., d_t) \in \{0,1\}^t$, $Y_{g,t}((d_{t'})_{t' \leq t}) = Y_{g,t}(d_{t-k}, ..., d_t)$.

Assumption 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 and Jaravel (2017) and in Schmidheiny and Siegloch (2020) as a possible way to identify these regressions.

In our context, imposing Assumption 17 can serve two purposes. First, it provides a solution to the "initial conditions" problem. So far, we have assumed that the treatments prior to the

start of the panel $(D_{g,t})_{t \leq 0}$ do not affect potential outcomes. There are at least two situations where this assumption is innocuous. First, such treatments may not exist. Assume for instance that one seeks to estimate the effect of being unionized on earnings, using the NLSY panel. In this data set, $t = 1$ corresponds to the first year on the labor market, so $D_{g,t}$ is not defined for $t \leq 0$. Second, in staggered adoption designs, our main results still hold if potential outcomes depend on $(D_{g,t})_{t \leq 0}$. In groups g untreated at period 1, $D_{g,t} = 0$ for all $t \leq 0$, so the first points of Theorems 1, 2 and 4 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 $k+1$.

Second, under Assumption 17-k one may be able to estimate δ_+ (resp. δ_-) even when there is no never-treated (resp. always-treated) group. Let $\psi_t = E[Y_{g,t}(\mathbf{0}) - Y_{g,t-1}(\mathbf{0})]$ for all $t > 1$. In the estimator $\mathrm{DID}_{+,t,\ell}$ defined above, we use the set of groups untreated from period 1 to t to unbiasedly estimate $\sum_{j=0}^{\ell} \psi_{t-j}$, the evolution of the never treated outcome from period 1 to t. Under 17-k, it is not necessary to restrict ourselves to groups untreated from period 1 to t. Instead, assume that for each t, there are groups untreated from period t to $t-k-1$: $N_t^{nt,k} = \sum_{g:D_{g,t}=...=D_{g,\max(1,t-k-1)}=0} N_{g,t} > 0$. Then, we can unbiasedly estimate ψ_t by

$$\frac{1}{N_t^{nt,k}} \sum_{g:D_{g,t}=...=D_{g,\max(1,t-k-1)}=0} N_{g,t} \left(Y_{g,t} - Y_{g,t-1}\right).$$

In turn, this allows us to estimate $\sum_{j=0}^{\ell} \psi_{t-j}$ and, in the end, δ_+. If $N_t^{nt,k} = 0$ for some t, we can at least unbiasedly estimate $\delta_+^{\mathrm{tru}\prime}$, defined exactly as δ_+^{tru} but with NT replaced by

$$NT' = \max\{t : \exists g : D_{g,t} = ... = D_{g,\max(1,t-k-1)} = 0\}.$$

Because $NT' \geq NT$, we have $\lambda_+^{\mathrm{tru}\prime} \geq \lambda_+^{\mathrm{tru}}$, implying that $\delta_+^{\mathrm{tru}\prime}$ is likely closer to δ_+ than δ_+^{tru}.

1.5 Fuzzy designs

In this subsection, we briefly discuss some fuzzy designs, where the treatment varies within (g,t) cells, and where our approach is still applicable. Assume that all groups are fully untreated at period 1. Then, one can redefine $F_{g,1}$ as the first period at which the proportion of units treated in group g is strictly positive, and redefine $\mathrm{DID}_{+,\ell}$ and $\widehat{\delta}_+^{\mathrm{tru}}$ accordingly. One can also redefine the actual-versus-status-quo parameter δ_+^{tru}, replacing $Y_{g,t}(\mathbf{D})$ by $\frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}(\mathbf{D}_{i,g})$, where $\mathbf{D}_{i,g} = (D_{i,g,1}, ..., D_{i,g,T})$ is a vector stacking the treatments of observation i in group g. Then,

one can show that under Assumptions 2-5, $E\left(\widehat{\delta}_+^{\text{tru}}\right) = \delta_+^{\text{tru}}$, as in Point 1 of Theorem 1. In fuzzy designs where some but not all groups are fully untreated at period 1, a similar result holds, in the subset of groups fully untreated at period 1. In fuzzy designs where some or all groups are fully treated at period 1, a similar result holds: one has to redefine $F_{g,0}$ as the first period at which the proportion of units untreated in group g is strictly positive, and the result one finally obtains involves $\widehat{\delta}_-^{\text{tru}}$ and δ_-^{tru}.

In fuzzy designs where all groups are partly treated at period 1, the estimators proposed in this paper are not applicable, even to a subset of groups. First, only groups where the proportion of treated units does not change over time can be used as controls, and such groups may not exist. Second, even when such groups exist, the outcome evolution of treated and untreated units has to be estimated separately in those groups, to estimate the outcome evolution that groups where the proportion of treated units changes would have experienced without that change (see de Chaisemartin and D'Haultfœuille, 2018). The algebra involved in those "treatment-adjusted" DID differs from the algebra underlying the estimators in this paper.

2   First-difference placebo estimators

For any $k \in \{1, ..., T-2\}$ and t such that $t - k \geq 2$, let

$$\text{DID}_{+,t,k}^{\text{fpl}} = \sum_{g:F_{g,1}=t} \frac{N_{g,t}}{N_{t,0}^1}(Y_{g,t-k} - Y_{g,t-k-1}) - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}}(Y_{g,t-k} - Y_{g,t-k-1})$$

if $N_{t,0}^1 > 0$ and $N_t^{nt} > 0$, and let $\text{DID}_{+,t,k}^{\text{fpl}} = 0$ otherwise. "fpl" stands for first-difference placebo. Let also $K_{nt}^{\text{fpl}} = \max\{t : 2 \leq t \leq NT, \exists g : F_{g,1} = t\}$ denote the last time period at which a group switches from untreated to treated for the first time while there is still a group that has always been untreated. We let

$$\text{DID}_{+,k}^{\text{fpl}} = \frac{\sum_{t=k+2}^{K_{nt}^{\text{fpl}}} N_{t,0}^1 \beta^t \text{DID}_{t,k}^{\text{fpl}}}{\sum_{t=k+2}^{K_{nt}^{\text{fpl}}} \sum_{g:F_{g,1}=t} N_{g,t}\beta^t D_{g,t}}$$

if $k \leq K_{nt}^{\text{fpl}} - 2$, and we let $\text{DID}_{+,k}^{\text{fpl}} = 0$ otherwise.

Similarly, for any $k \in \{1, ..., T-2\}$ and t such that $t - k \geq 2$, let

$$\text{DID}_{-,t,k}^{\text{fpl}} = \sum_{g:F_{g,0}=t} \frac{N_{g,t}}{N_{t,0}^0}(Y_{g,t-k} - Y_{g,t-k-1}) - \sum_{g:F_{g,0}>t} \frac{N_{g,t}}{N_t^{at}}(Y_{g,t-k} - Y_{g,t-k-1})$$

if $N_{t,0}^0 > 0$ and $N_t^{at} > 0$, and let $\text{DID}_{-,t,k}^{\text{fpl}} = 0$ otherwise. Let also $K_{at}^{\text{fpl}} = \max\{t : 2 \leq t \leq AT, \exists g : F_{g,0} = t\}$ denote the last time period at which a group switches from treated to untreated for

the first time while there is still a group that has always been treated. We let

$$\text{DID}^{\text{fpl}}_{-,k} = \frac{\sum_{t=k+2}^{K^{\text{fpl}}_{at}} N^0_{t,0} \beta^t \text{DID}^{\text{fpl}}_{t,k}}{\sum_{t=k+2}^{K^{\text{fpl}}_{at}} \sum_{g:F_{g,0}=t} N_{g,t} \beta^t D_{g,t}}$$

if $k \leq K^{\text{fpl}}_{at} - 2$, and we let $\text{DID}^{\text{fpl}}_k = 0$ otherwise.

As with the long-difference placebos, one can show that if Assumptions 1-2, Point 1 of Assumption 3, and Assumptions 4-5 hold,

$$E\left[\text{DID}^{\text{fpl}}_{+,k}\right] = 0 \quad \forall k \in \{1, ..., T-2\}.$$

Similarly, if Assumptions 1-2, Point 2 of Assumption 3, and Assumptions 6-7 hold,

$$E\left[\text{DID}^{\text{fpl}}_{-,k}\right] = 0 \quad \forall k \in \{1, ..., T-2\}.$$

References

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

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

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

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 Proof of Theorem 4

We only prove Point 1, as the proof of Point 2 is similar. First, let us introduce additional notation. Let $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_G)$, $\varepsilon_{g,t}(\boldsymbol{d}_t) = Y_{g,t}(\boldsymbol{d}_t) - X'_{g,t}\theta_0$ and $\varepsilon_{g,t} = \varepsilon_{g,t}(\boldsymbol{D}_g)$. Let also $\widetilde{\delta}_+^{\text{tru},X}$ be as $\widehat{\delta}_+^{\text{tru},X}$ but replacing $\widehat{\theta}_0$ by θ_0. Similarly, let $\widetilde{\text{DID}}_{+,t,\ell}^X$ be as $\text{DID}_{+,t,\ell}^X$ but with $\widehat{\theta}_0$ replaced by θ_0. For any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$ such that $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$, we have

$$E\left(\widetilde{\text{DID}}_{+,t,\ell}^X | \boldsymbol{D}, \boldsymbol{X}\right) = \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t} - Y_{g,t-\ell-1} - (X_{g,t} - X_{g,t-\ell-1})'\theta_0 | \boldsymbol{D}, \boldsymbol{X}\right)$$

$$- \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} E\left(\varepsilon_{g,t} - \varepsilon_{g,t-\ell-1} | \boldsymbol{D}, \boldsymbol{X}\right)$$

$$= \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}) | \boldsymbol{D}, \boldsymbol{X}\right)$$

$$+ \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(\varepsilon_{g,t}(\boldsymbol{0}) - \varepsilon_{g,t-\ell-1}(\boldsymbol{0}) | \boldsymbol{D}, \boldsymbol{X}\right)$$

$$- \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} E\left(\varepsilon_{g,t}(\boldsymbol{0}) - \varepsilon_{g,t-\ell-1}(\boldsymbol{0}) | \boldsymbol{D}, \boldsymbol{X}\right)$$

$$= \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}) | \boldsymbol{D}, \boldsymbol{X}\right).$$

The first equality follows from the definition of $\widetilde{\text{DID}}_{+,t,\ell}^X$, $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$. The second equality follows from Assumption 2. The third equality follows from Assumption 9. Then, following the exact same steps as in the proof of Theorem 1, we obtain

$$E\left[\widetilde{\delta}_+^{\text{tru},X}\right] = \delta_+^{\text{tru}}. \tag{15}$$

Now, note that

$$E\left(\text{DID}_{+,t,\ell}^X - \widetilde{\text{DID}}_{+,t,\ell}^X | \boldsymbol{D}, \boldsymbol{X}\right) = \beta^t \left\{ \sum_{g:F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^1} (X_{g,t} - X_{g,t-\ell-1})' \left[\theta_0 - E\left(\widehat{\theta}_0 | \boldsymbol{D}, \boldsymbol{X}\right)\right] \right.$$

$$\left. - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N_t^{nt}} (X_{g,t} - X_{g,t-\ell-1})' \left[\theta_0 - E\left(\widehat{\theta}_0 | \boldsymbol{D}, \boldsymbol{X}\right)\right] \right\}$$

$$= 0. \tag{16}$$

The first equality follows from the definition of $\text{DID}_{+,t,\ell}^X$, $\widetilde{\text{DID}}_{+,t,\ell}^X$, $N_{t,\ell}^1 > 0$ and $N_t^{nt} > 0$. The second equality stems from $E\left(\widehat{\theta}_0 | \boldsymbol{D}, \boldsymbol{X}\right) = \theta_0$, which is due to Assumption 9. The result follows by combining (15) with (16).

15

A.2 Proof of Theorem 5

We only prove Point 1, as Point 2 can be obtained in a similar way. Reasoning as for obtaining (10), we have, for all $s \in \mathcal{S}$ and any $\ell \in \{0, ..., T-2\}$ and $t \in \{\ell+2, ..., T\}$ such that $N_{t,\ell}^{1,s} \times N_t^{nt,s} > 0$,

$$E\left(\mathrm{DID}_{+,t,\ell}^s \big| \boldsymbol{D}\right) = \sum_{g \in s: F_{g,1}=t-\ell} \frac{N_{g,t}}{N_{t,\ell}^{1,s}} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}) \big| \boldsymbol{D}\right). \tag{17}$$

Point 1 of Assumption 11 ensures that $\mathcal{S}_+$ is not empty. By definition of NT^s and because for all $s \in \mathcal{S}_+$ there exists $(g, g') \in s$ such that $1 < F_{g,1} < F_{g',1}$, we have $N_t^{nt,s} > 0$ for all $2 \le t \le NT^s$. Consider an arbitrary $\ell \in \{0, ..., L_{nt}^s\}$. With the convention that a sum over an empty set is equal to 0, we get, for all $s \in \mathcal{S}_+$ and $\ell + 2 \le t \le NT^s$,

$$N_{t,\ell}^{1,s} E\left(\mathrm{DID}_{+,t,\ell}^s \big| \boldsymbol{D}\right) = \sum_{g \in s: F_{g,1}=t-\ell} N_{g,t} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}) \big| \boldsymbol{D}\right). \tag{18}$$

Moreover, by definition of L_{nt}^s, we have $\sum_{t=\ell+2}^{NT^s} \beta^t N_{t,\ell}^1 > 0$ for all $\ell \in \{0, ..., L_{nt}^s\}$. Then, summing over t in (18), we obtain

$$E\left(\mathrm{DID}_{+,\ell}^s \big| \boldsymbol{D}\right) = \frac{1}{\sum_{t=\ell+2}^{NT^s} \beta^t N_{t,\ell}^1} \sum_{t=\ell+2}^{NT^s} \sum_{g: F_{g,1}=t-\ell} \beta^t N_{g,t} E[Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{0}) \big| \boldsymbol{D}].$$

This equation is the equivalent of (11) in the proof of Theorem 1. The result then follows exactly as the end of the proof of Theorem 1.

A.3 Proof of Theorem 6

We only prove Point 1, as Point 2 can be obtained in a similar way. Using the same reasoning as to obtain (10), we obtain, for all (d, t, ℓ) such that $N_{t,\ell}^{\neq d,+} > 0$ and $N_t^{=d} > 0$,

$$E\left(\mathrm{DID}_{+,d,t,\ell} \big| \boldsymbol{D}\right) = \sum_{g: F_{g,\neq d}=t-\ell, I_{g,d}=1} \frac{N_{g,t}}{N_{t,\ell}^{\neq d,+}} E\left(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(d) \big| \boldsymbol{D}\right).$$

Also, given that Point 1 of Assumption 16 holds, we have

$$\sum_{\ell=0}^{L_{+,d}} \sum_{t=\ell+2}^{AT_d} \sum_{g: F_{g,\neq d}=t-\ell, I_{g,d}=1} N_{g,t} \beta^t (D_{g,t} - d)$$

$$= \sum_{g: 2 \le F_{g,\neq d} \le AT_d, I_{g,d}=1} \sum_{t=F_{g,\neq d}}^{AT_d} N_{g,t} \beta^t (D_{g,t} - d) > 0.$$

Then, reasoning as to obtain (12), we get

$$\frac{\sum_{\ell=0}^{L_{+,d}} w_{+,d,\ell} E\left(\mathrm{DID}_{+,d,\ell}|\boldsymbol{D}\right)}{\sum_{\ell=0}^{L_{+,d}} w_{+,d,\ell} \mathrm{DID}_{+,d,\ell}^{D}} = \frac{\sum_{\ell=0}^{L_{+,d}} \sum_{t=\ell+2}^{AT_d} \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=1} N_{g,t}\beta^t E(Y_{g,t}(\boldsymbol{D}_g) - Y_{g,t}(\boldsymbol{d})|\boldsymbol{D})}{\sum_{\ell=0}^{L_{+,d}} \sum_{t=\ell+2}^{AT_d} \sum_{g:F_{g,\neq d}=t-\ell,I_{g,d}=1} N_{g,t}\beta^t (D_{g,t} - d)}$$

$$= E[\Delta_{+,d}^{tru}|\boldsymbol{D}].$$

Hence,

$$E[\widehat{\delta}_{+}^{\mathrm{tru}}|\boldsymbol{D}] = \sum_{d=0}^{\bar{d}} v_{+,d} E[\Delta_{+,d}^{tru}|\boldsymbol{D}].$$

The result follows from the law of iterated expectations and the previous display.