Difference-in-Differences with Unequal Baseline Treatment Status*

Alisa Tazhitdinova Gonzalo Vazquez-Bare

May 14, 2025

Abstract

In DiD settings, many empirical studies compare a group of units that switch treatment status to a group that does not, where the two groups have different treatment statuses in the pre-switch period. We propose a framework that allows switchers and stayers to start from different treatment levels and show that these unequal-baseline DiD comparisons implicitly impose two restrictions on treatment effect heterogeneity, namely, that treatment effects are time invariant, and do not accumulate over time. When treatment effects vary or accumulate over time, DiD estimators are biased for the effect on the switchers, and pre-trends tests may detect significant differences even when the canonical parallel-trends assumption holds. We study the performance of event-study estimators in such settings and provide new results that formally characterize the commonly used event-study regression with a linear-trend adjustment. We then show that the bias from this adjustment may be larger than the bias of unadjusted event-study estimators when treatment effects exhibit nonlinearities. We discuss the implications of our results for empirical practice.

JEL Classification: C21, C23
Keywords: differences-in-differences, event study, parallel-trends assumption, linear trends

*Alisa Tazhitdinova: Department of Economics, UCSB and NBER (tazhitda@ucsb.edu); Gonzalo Vazquez-Bare: Department of Economics, UCSB (gvazquez@econ.ucsb.edu). We thank Youssef Benzarti, Clément de Chaisemartin, Olga Namen, Heather Royer, and Doug Steigerwald for helpful suggestions. Sofia Olguin and Xhulio Uruci provided excellent research assistance.

In difference-in-differences (DiD) settings, the effect of a treatment is estimated by comparing the outcome trend of a group that switches treatment status at some point in time to the outcome trend of a comparison group that remains at the same treatment status throughout the period of study. In the canonical framework, both groups start at the same baseline treatment status and then diverge. Many empirical studies, however, instead compare a group that experiences a change in treatment status to a group that does not, while allowing groups to have different status in the baseline (i.e., pre-switch) period. We refer to these settings as "unequal-baseline DiD."

The unequal-baseline DiD approach is ubiquitous in empirical studies. For instance, it is widely used in tax research to measure the causal effects of taxes on income, wealth, investments, etc. A typical study exploits differential changes in marginal tax rates (MTRs) across groups of taxpayers over time (e.g., Kleven and Schultz, 2014; Jakobsen et al., 2020; Yagan, 2015; Fuest et al., 2018). Similarly, minimum wage studies often compare workers residing in states with different minimum wages or workers residing in the same state but subject to different minimum wage regimes (e.g., Dube et al., 2010; Giuliano, 2013; Cengiz et al., 2019; Jardim et al., 2022). Unequal-baseline DiDs have been employed in economics of education to study effects of affirmative action bans (Bleemer, 2022); in health economics to study the consequences of lead exposure (Grönqvist et al., 2020); in labor economics to study the employment effects of working hour reductions (Chemin and Wasmer, 2009); in environmental economics to study the energy effects of daylight savings time (Kotchen and Grant, 2011); in development economics to study rebellions (Cao and Chen, 2022), malaria eradication programs (Bleakley, 2010; Cutler et al., 2010; Lucas, 2010; Rossi and Villar, 2020), and management institutions (Sawada et al., 2022), to name just a few.

In this paper, we analyze unequal-baseline DiD settings and provide conditions under which commonly used event-study and two-way fixed effects (TWFE) estimators recover causally interpretable parameters. We consider a setting where a group of units, the switchers, switch between two treatment levels, whereas the comparison group, the stayers, remains at the same treatment level throughout the study. We show that, in addition to the usual parallel-trends assumption, identification of treatment effects in unequal-baseline DiDs requires treatment effects to be time-invariant and to not accumulate over time. When the latter conditions fail, the usual DiD estimands recover a combination of the effect of the treatment for switchers and the change in the treatment effects over time for stayers. We also show that, in such settings, the commonly used pre-trends tests that compare the evolution of outcomes between groups in the pre-switch period may capture the change in the treatment effect for the stayers over time, and thus may pick up significant differences in trends even when the standard parallel-trends assumption holds.

We then consider a solution frequently used to account for differential trends in the baseline period – the inclusion of a linear trend term in the event-study specification. This

approach has been employed in unequal-baseline DiD frameworks (see e.g., Jakobsen et al., 2020), but also in canonical DiD settings (see e.g., Bilinski and Hatfield, 2018; Mora and Reggio, 2019; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2024) where trends appear to be non-parallel in a systematic way. To our knowledge, we are the first to provide a closed-form characterization of the linear-trend-adjusted estimators and their corresponding estimands. We show that (i) the linear-trend coefficient is a weighted average of outcome pre-trends differences with a quadratic structure that gives higher weights to periods around the middle of the period used to estimate the difference in trends and (ii) that the linear-trend-adjusted event-study estimators are DiD comparisons after adjusting the outcomes for a specific linear combination of pre-switch trends. Importantly, the weights used by these estimators are observable, so our results allow the researcher to directly calculate these weights. We then show that the linear-trend-adjusted estimators are consistent for the average effects on the switchers if the differences in outcome trends are constant over time (or zero). In contrast, if differences in trends are not constant over time, then both the standard event-study estimators and the linear-trend-adjusted estimators are inconsistent, and it is generally not possible to determine which asymptotic bias is larger.

Finally, we argue that unequal-baseline DiDs can exacerbate the problems in TWFE regressions with staggered designs pointed out by recent literature (see Steigerwald et al., 2021; de Chaisemartin and D'Haultfœuille, 2022; Roth et al., 2023, for surveys). We show, however, that the researcher may drop the stayers from the sample and compare the just-switched to the not-yet-switched, as long as these groups start from the same baseline treatment. Thus, by providing an alternative set of comparison units, a staggered adoption setting offers the opportunity to recover treatment effects under the canonical parallel-trends assumption without restricting treatment effect heterogeneity.

Our analysis suggests that the unequal-baseline DiD approach may successfully recover treatment effects in settings where treatment effects are nearly constant, are expected to be fairly immediate, or have stabilized over time. For example, this approach may work well in studies estimating labor supply responses to personal income taxes, since labor responses tend to stabilize quickly, typically after a brief adjustment period. In contrast, studying wealth responses to a wealth tax will lead to biased estimates if the switchers and stayers face different wealth taxes in the baseline period. This is because wealth taxes have a cumulative effect on one's wealth over time, leading to time-varying treatment effects. Having data that covers the pre-baseline period in which groups experience equal treatment status can be useful to establish the nature of treatment effects in the specific setting studied. Alternatively, economic reasoning and prior research on the topic can be used to form expectations about the nature of treatment effects or any frictions that may affect it.

The rest of the paper is organized as follows. Section 1 illustrates the intuition behind our results with a simple three-period setting and graphical examples. Section 2 generalizes

our results to a multi-period setting, and characterizes estimators for event-study designs and designs with linear-trends adjustments. Section 3 concludes with practical recommendations.

Related literature. DiD models have been the focus of a rapidly growing literature analyzing the performance of panel data methods under treatment effect heterogeneity (see Steigerwald et al., 2021; de Chaisemartin and D'Haultfœuille, 2022; Roth et al., 2023, for recent surveys). In particular, several studies (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021; Athey and Imbens, 2022; Borusyak et al., 2024) have pointed out the identification problems that result from the "forbidden comparisons" implicitly used by TWFE specifications in staggered designs, whereby late treatment adopters are compared to cohorts that became treated in earlier periods and are therefore invalid as a comparison group. This literature has proposed several alternative methods that recover causal effects by ensuring that only valid comparisons, that is, comparisons between treated and never-treated or not-yet-treated, are used (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024).

Our setup can be seen as an extreme case of this forbidden comparisons problem, where no valid comparisons are available because switchers and stayers never share the same treatment status, and thus the solutions proposed by the literature cannot be directly applied. For instance, in the scenario that we refer to as "universal adoption", one group is always treated, whereas the other group enters treatment in a later period. When relying on two-way fixed effects models, the always-treated group acts as a comparison group (Goodman-Bacon, 2021), which generally results in inconsistent estimation of causal effects. To avoid this issue, a common recommendation is to exclude the always-treated group from the analysis: for instance, Sun and Abraham (2021) state that "we need to exclude [the always-treated] from estimation" (page 186) and Borusyak et al. (2024) point out that this cohort is "not useful for causal identification" (footnote 7). Excluding the always-treated is not feasible in our setting because there are no other treatment cohorts. Therefore, identification in the universal adoption setting requires additional assumptions. This setting is most closely related to de Chaisemartin and D'Haultfœuille (2018) and Kim and Lee (2019), who also consider a universal adoption setup. Within a fuzzy DiD setup, de Chaisemartin and D'Haultfœuille (2018) provide identification conditions for causal effects under a parallel-trends assumption that is analogous to our Condition (1) and requires that the outcome trends for the switchers and stayers are the same under their respective pre-policy change status (see Section 3.4.2 and the supplemental appendix in their paper). Kim and Lee (2019) provide conditions to identify the effect of a binary treatment in the pre-policy change period using a "reverse DiD" strategy in a two-period DiD. Our analysis adds to these results in multiple dimensions. First, we consider a more general setting with both converging and diverging treatment status, of which universal policy adoption is a particular case, and discuss different versions (and im-

plications) of the parallel-trends assumption. We also allow potential outcomes to depend on past treatment status, and our results show that the reverse-DiD strategy in Kim and Lee (2019) requires the (often restrictive) assumption that potential outcomes only depend on current treatment status (see Section 1.5). Finally, our framework allows for multiple periods, different estimation strategies (including event-study designs, which are one of the most commonly-used empirical DiD approaches), a frequently-employed linear-trend adjustment, and we characterize the behavior of the estimators typically used to test for pre-trends in outcomes.

Another recent strand of the literature analyzes DiD models with continuous and multi-valued treatments. Continuously distributed treatments (such as pollution levels or trade tariffs) naturally give rise to settings where units start from different treatment levels. de Chaisemartin et al. (2025) argue that comparing units with different baseline treatment levels requires combining a parallel-trends assumption with a treatment effect homogeneity restriction, a point that we also discuss in Section 2. To avoid restricting treatment effect heterogeneity, they propose two-step nonparametric estimators that "match" switchers to stayers with the same treatment level in the pre-policy change period. Under an appropriate parallel-trends assumption, these estimators are consistent for weighted averages of outcome slopes for the switchers. In this paper, we focus instead on the parameters that can be estimated using standard two-way fixed effects and event-study designs commonly implemented in practice. On the other hand, Callaway et al. (2024) consider a continuous or multi-valued treatment setting where all units start from a no-treatment baseline, so their results do not directly apply to our case.

Finally, our paper is also related to the literature that focuses more directly on the parallel-trends assumption. In particular, our results show that DiD designs with unequal baseline treatment status may often suffer from non-parallel trends, a topic that has received increased attention in recent years (Manski and Pepper, 2018; Kahn-Lang and Lang, 2019; Bilinski and Hatfield, 2018; Rambachan and Roth, 2023; Roth and Sant'Anna, 2023; Roth, 2022). We also discuss the performance of the usual pre-trends test in the unequal-baseline setting and argue that it is typically less informative about the validity of the parallel-trends assumption compared to the canonical case.

1 Intuition-Building Example

1.1 Setup

We illustrate the main intuition behind our results by considering a simple setting with three periods, $t = 0, 1, 2$. The potential outcomes for a random unit from the population are given by $Y_0(d_0)$ in period 0, $Y_1(d_1, d_0)$ in period 1, and $Y_2(d_2, d_1, d_0)$ in period 2. Our

framework allows potential outcomes to depend on the treatment assignment in all previous periods, a property often known as *dynamic potential outcomes*. We consider a universal adoption policy setting where one group of units, *the switchers*, are untreated in periods $t = 0, 1$ but switch to treatment in period 2, whereas the other group, *the stayers*, are treated in all periods (see Appendix A for examples of this setting in the empirical literature). Thus $d_0 = d_1 = 0$ and $d_2 = 1$ for the switchers, and $d_0 = d_1 = d_2 = 1$ for the stayers. We let $\tau_2(d_2, d_1, d_0) = Y_2(d_2, d_1, d_0) - Y_2(0, 0, 0)$ be the treatment effect in period 2 of treatment path (d_0, d_1, d_2) compared to the always-untreated status $(0, 0, 0)$, and similarly for $\tau_1(d_1, d_0) = Y_1(d_1, d_0) - Y_1(0, 0)$ and $\tau_0(1) = Y_0(1) - Y_0(0)$. We let S be a switcher indicator, so that $S = 1$ for the switchers and $S = 0$ for the stayers.

1.2 Illustrative Example

We use a wealth tax reform as a hypothetical illustrative example. Consider a country where half of the adults are subject to a 1% wealth tax in periods $t = 0, 1, 2$. The remaining half is not subject to any wealth tax in periods $t = 0, 1$. In period $t = 2$, a wealth tax reform is introduced so that everyone is subject to the 1% tax. In this setup, individuals who were always subject to the wealth tax are the stayers, while individuals subject to the wealth tax only in period 2 are the switchers. We will discuss whether the typical DiD approach can recover the effect of the 1% tax increase on the switchers' wealth levels. Figures 1(a) and (b) illustrate two possible scenarios in our setting. Figure (a) assumes that the treatment effect of the wealth tax is immediate, while Figure (b) assumes that the treatment effect appears two periods after treatment.

1.3 DiD Estimand

The DiD estimand between periods 2 and 1 is $\Delta_{\mathsf{post}} = \mathbb{E}[Y_2 - Y_1|S = 1] - \mathbb{E}[Y_2 - Y_1|S = 0]$, which can be expressed as:

$$\begin{aligned}\Delta_{\mathsf{post}} &= \mathbb{E}[Y_2(1,0,0) - Y_1(0,0)|S = 1] - \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 0] \\ &= \mathbb{E}[\tau_2(1,0,0)|S = 1] + \mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S = 1] - \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 0].\end{aligned}$$

Suppose the following condition holds:

$$\mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S = 1] = \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 0]. \tag{1}$$

Then, $\Delta_{\mathsf{post}} = \mathbb{E}[\tau_2(1,0,0)|S = 1]$, which is the average effect of entering the treatment in period two on the switchers. Note that assumption (1) is not the standard parallel-trends assumption used in canonical DiD settings because switchers and stayers are in different

treatment statuses before period $t = 2$. To better understand the implied restrictions behind this condition, we can rewrite the right-hand side of (1) as follows:

$$\begin{aligned}\mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 0] =& \mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S = 0] \\ &+ \mathbb{E}[\tau_2(1,1,1) - \tau_1(1,1)|S = 0].\end{aligned}$$

It is now easy to see that the following two assumptions are sufficient for condition (1):

Assumption A (Canonical parallel-trends assumption) *The average potential outcomes of switchers and stayers under no treatment follow the same trajectory over time,*

$$\mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S = 1] = \mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S = 0].$$

Assumption B (Time-invariant non-cumulative effects) *The average effect of the treatment on the stayers is invariant to the number of periods under treatment and invariant to the initial period of treatment,*

$$\mathbb{E}[\tau_2(1,1,1)|S = 0] = \mathbb{E}[\tau_1(1,1)|S = 0].$$

Assumption B implicitly imposes two conditions. Notice that:

$$\tau_2(1,1,1) - \tau_1(1,1) = [\tau_2(1,1,1) - \tau_2(1,1,0)] + [\tau_2(1,1,0) - \tau_1(1,1)]\,.$$

This difference is zero when $\mathbb{E}[\tau_2(1,1,1)|S = 0] = \mathbb{E}[\tau_2(1,1,0)|S = 0]$, so the effect of being treated for three periods equals the effect of being treated for two periods (non-cumulative effects), and $\mathbb{E}[\tau_2(1,1,0)|S = 0] = \mathbb{E}[\tau_1(1,1)|S = 0]$ so the effect of being treated for exactly two periods is the same regardless of whether the treatment started in period 2 or 1 (time-invariant effects).

Under Assumptions A and B, condition (1) holds, and therefore the DiD estimand Δ_{post} identifies the average effect of entering the treatment in period 2 on the switchers. This demonstrates that the parallel-trends condition (1) is not equivalent to the canonical parallel-trends assumption (Assumption A) that requires that potential outcomes evolve similarly *under no treatment*, nor is Assumption A alone sufficient to identify the effect of the treatment on the switchers. Specifically, under Assumption A only,

$$\Delta_{\mathsf{post}} = \mathbb{E}[\tau_2(1,0,0)|S = 1] - \mathbb{E}[\tau_2(1,1,1) - \tau_1(1,1)|S = 0].$$

Thus, when treatment effects vary or accumulate over time, the DiD estimand will recover the difference between the average effect on the switchers and the change in the average effect over time for the stayers. In settings where the treatment effect is small in the beginning and

increases over time, this issue can lead to sign reversals (de Chaisemartin and D'Haultfœuille, 2022; Roth et al., 2023): the DiD estimand may be negative even when all the treatment effects are positive.

In our wealth tax example, condition (1) requires individuals subject to the wealth tax throughout the years to experience a similar wealth growth path as individuals who were not subject to the wealth tax before period $t = 2$. While this assumption appears to be innocuous, generally speaking, it cannot be satisfied: in the absence of perfectly offsetting behavioral responses, the wealth of taxed individuals will grow at (approximately) a 1% lower rate than that of the untaxed individuals, simply because wealth is a cumulative measure. In contrast, Assumption A requires stayers and switchers to experience similar wealth growth paths when there is no wealth tax – a plausible assumption if the two groups are sufficiently similar. Note that Assumption B also does not hold in our wealth tax example – while wealth tax treatment is likely to be invariant to the initial period of treatment, the treatment is cumulative. Individuals subject to the wealth tax for two years will have a 1% lower wealth level than individuals subject to a wealth tax for one year only, and a 2% lower wealth level than individuals not subject to the wealth tax at all (again, abstracting from behavioral responses).

1.4 Pre-Trends Tests

The parallel-trends assumption A involves unobservable counterfactual outcomes and therefore cannot be tested directly. In practice, researchers often strengthen this assumption by requiring that it holds in all pre-periods, thus adding to Assumption A the requirement that $\mathbb{E}[Y_1(0,0) - Y_0(0)|S = 1] = \mathbb{E}[Y_1(0,0) - Y_0(0)|S = 0]$ between periods $t = 0$ and $t = 1$ as well. In a canonical DiD setting, this second requirement is testable because both units are untreated in these periods. This is the so-called *pre-trends test*. Based on this common practice, in an unequal baseline DiD, a researcher can consider the difference in outcome trends between switchers and stayers in periods 1 and 0, $\Delta_{\mathsf{pre}} = \mathbb{E}[Y_1 - Y_0|S = 1] - \mathbb{E}[Y_1 - Y_0|S = 0]$. This difference can be expressed as:

$$\begin{aligned}\Delta_{\mathsf{pre}} &= \mathbb{E}[Y_1(0,0) - Y_0(0)|S = 1] - \mathbb{E}[Y_1(1,1) - Y_0(1)|S = 0] \\ &= \mathbb{E}[Y_1(0,0) - Y_0(0)|S = 1] - \mathbb{E}[Y_1(0,0) - Y_0(0)|S = 0] - \mathbb{E}[\tau_1(1,1) - \tau_0(1)|S = 0].\end{aligned}$$

Thus, we see that Δ_{pre} may be non-zero even under the canonical parallel-trends assumption whenever treatment effects vary or accumulate over time. Figure 1(a) illustrates this case: even though switchers and stayers exhibit the same trend under no treatment, they will not exhibit parallel trends in the pre-switch period if the treatment effect of wealth tax is immediate.

Figure 1(b) illustrates a case where the pre-trends test will (inadvertently) suggest that the groups are comparable, while in reality, the identification conditions fail. In Figure (b), we assume that the treatment effects on the stayers are not immediate. Consequently, the parallel-trends assumption appears to be satisfied in the pre-treatment periods, but the treatment effects are not identified because trends become non-parallel in the post-switch period. Note that the bias in Figure 1(b) implies a DiD estimand of the opposite sign from the true treatment effect.

Our analysis thus highlights that when the treatment effects are cumulative over time or depend on the timing of treatments, the DiD estimator is biased, and this bias can potentially result in a sign reversal. When the treatment effects on the stayers are immediate, as illustrated in Figure 1(a), this issue may be detected from a pre-trends analysis. However, in cases like Figure 1(b), pre-trends tests are unable to detect this issue. Importantly, our stylized illustrative examples abstract away from important practical issues related to inference. In practice, differentiating between cases 1(a) and 1(b) may not be straightforward because of noisy data and/or because researchers often have access to only a short pre-switch period, which results in low statistical power to detect existing differences.[1] This demonstrates why the validity of the parallel-trends assumption in unequal-baseline settings should not be assessed solely on statistical grounds, and why empirical evidence needs to be complemented with institutional knowledge and economic theory.

1.5 Remarks and Further Discussions

Treatment Renaming. A savvy reader may note that condition (1) can be converted into a seemingly innocuous canonical Assumption A by choosing an alternative definition of treatment. For example, in our wealth tax example, one could define treatment as "experiencing a wealth tax change of 1%". In this case, the stayers and the switchers are not treated in periods 0 and 1, while the switchers experience treatment in period 2. Such renaming simply masks, but does not solve, the problem. Under such a modified definition of treatment, the parallel-trends assumption now requires that treated individuals in the absence of treatment would have experienced the same wealth growth path as comparison individuals. Note that the comparison group consists of individuals who were always subject to the wealth tax (these individuals do not "experience a wealth tax change of 1%"), while the treatment group consists of individuals who in absence of treatment were not subject to wealth tax (and hence do "experience a wealth tax change of 1%"). Since those subject to a wealth tax will generally not have a wealth growth path as those not subject to a wealth tax, such renaming does not

[1] It is well known that the absence of differences in trends before the policy change does not imply that trends would have been the same in the post period in the absence of the policy change. Recent literature has also pointed out that pre-trends tests often have limited statistical power to detect violations of parallel trends (Bilinski and Hatfield, 2018; Freyaldenhoven et al., 2019; Kahn-Lang and Lang, 2019; Roth, 2022).

alleviate the problem discussed in this section.

Completed Treatments on the Stayers. The analysis in this section shows that if treatment effects on the stayers vary or accumulate over time, the DiD estimand will not recover the effect on the switchers. In practice, treatment effects typically feature one of three paths: they may be constant and immediate, constant in the long run but gradual in the short run, or they may vary/accumulate over time, both in the short and long run. As Assumption B makes it clear, to avoid bias, treatment effects on the stayers must either be constant and immediate, or, at the moment of evaluation (i.e., by period 1), have reached the long run state, thus effectively reaching the constant treatment effect phase.

Our wealth tax example represents a setting where treatment effects always accumulate over time: as long as the individual is subject to the wealth tax treatment, his wealth will grow more slowly than the wealth of a similar individual who is not subject to the wealth tax. However, for most other tax types – e.g., income taxes, corporate income taxes, etc – we would generally expect constant treatment effects, but perhaps after a brief adjustment period. Such DiD settings will produce unbiased estimates as long as sufficient time has passed from when the stayers were initially treated, so that the stayers have "completed" their treatments. Figure 2(a) illustrates such a scenario. In this example, the treatment on the stayers reaches the constant treatment phase before the period of study.

Consequently, unequal-baseline DiD analysis requires that (i) researchers have a prior belief about the nature of treatment effects that is consistent with Assumptions A and B, and (ii) the estimated treatment effects are consistent with this prior.

Temporary Treatments on the Stayers. In some settings, a treatment is introduced in a certain period and then removed in subsequent periods. Consider the case where the treatment is assigned in period 0 for the stayers and then removed in periods 1 and 2. As before, the switchers are treated in period 2. In this case,

$$\begin{aligned}\Delta_{\mathsf{post}} &= \mathbb{E}[Y_2(1,0,0) - Y_1(0,0)|S=1] - \mathbb{E}[Y_2(0,0,1) - Y_1(0,1)|S=0] \\ &= \mathbb{E}[Y_2(1,0,0) - Y_2(0,0,0)|S=1] \\ &+ \mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S=1] - \mathbb{E}[Y_2(0,0,1) - Y_1(0,1)|S=0],\end{aligned}$$

where

$$\mathbb{E}[Y_2(0,0,1) - Y_1(0,1)|S=0] = \mathbb{E}[Y_2(0,0,0) - Y_1(0,0)|S=0] + \mathbb{E}[\tau_2(0,0,1) - \tau_1(0,1)|S=0].$$

With such temporary treatments, the DiD parameter recovers the average effect on the switchers when Assumption A holds and when the effect of the treatment vanishes immedi-

ately after the treatment is removed, which is analogous to the time-invariant effects part of Assumption B. Under these conditions, since the treatment only lasts one period, the effects of the treatment do not accumulate over time and thus the condition of no cumulative effects plays no role.

In our wealth tax example, this scenario corresponds to a situation where some individuals are subject to the wealth tax in period 0 but not in periods 1 and 2, while other individuals are subject to tax in period 2 but not periods 0 and 1. In this case, the usual DiD approach will recover the effect of wealth taxation on income level if the wealth accumulation process is such that the cancellation of a wealth tax immediately restores one's wealth growth path to that of untaxed individuals. This is a plausible assumption in the case of wealth taxes. This case is illustrated in Figure 2(b).

Figure 2(c) illustrates the opposite scenario, where treatment effects are not immediate. In the case of wealth tax, this may happen if previously taxed individuals engage in costly tax evasion schemes that cannot be reversed immediately. In that case, their wealth growth path may not immediately return to that of untaxed individuals. In such circumstances, the DiD estimator is likely to be biased, and this bias can be large, even implying an estimated effect of the opposite sign than the true treatment effect, as in Figure 2(c).

Reverse DiD. Kim and Lee (2019) consider a DiD setting with universal adoption and show that under a parallel-trends assumption that involves the treated potential outcomes, it is possible to identify the average effect on the switchers in the pre-switch period. A crucial implicit assumption in their result is that potential outcomes are static and do not depend on past treatments, so that $Y_2(d_2, d_1, d_0) = Y_2(d_2)$ and $Y_1(d_1, d_0) = Y_1(d_1)$. Specifically, when potential outcomes are static,

$$\begin{aligned}\Delta_{\mathsf{post}} &= \mathbb{E}[Y_2(1) - Y_1(0)|S=1] - \mathbb{E}[Y_2(1) - Y_1(1)|S=0] \\ &= \mathbb{E}[Y_1(1) - Y_1(0)|S=1] + \mathbb{E}[Y_2(1) - Y_1(1)|S=1] - \mathbb{E}[Y_2(1) - Y_1(1)|S=0],\end{aligned}$$

and thus if $\mathbb{E}[Y_2(1) - Y_1(1)|S=1] = \mathbb{E}[Y_2(1) - Y_1(1)|S=0]$, the DiD estimand recovers the effect of the treatment in the pre-switch period.

When potential outcomes are dynamic, however, the DiD estimand can be written as:

$$\begin{aligned}\Delta_{\mathsf{post}} &= \mathbb{E}[\tau_1(1,0)|S=1] \\ &\quad + \mathbb{E}[Y_2(1,0,0) - Y_1(1,0)|S=1] - \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S=0] \\ &= \mathbb{E}[\tau_1(1,0)|S=1] \\ &\quad + \mathbb{E}[\tau_2(1,0,0) - \tau_2(1,1,1)|S=1] + \mathbb{E}[\tau_1(1,1) - \tau_1(1,0)] \\ &\quad + \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S=1] - \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S=0].\end{aligned}$$

From this expression it is clear that, in addition to the parallel-trends assumption under treatment $\mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 1] = \mathbb{E}[Y_2(1,1,1) - Y_1(1,1)|S = 0]$, identifying $\mathbb{E}[\tau_1(1,0)|S = 1]$ also requires restricting how treatment effects vary over time. Thus, our framework generalizes the results of Kim and Lee (2019) by allowing for dynamic treatment effects and by not requiring treatment convergence.

Note that in our wealth tax example, potential outcomes are not static, since one's wealth level today depends not only on the current level of wealth taxation but also on the past levels of wealth taxation. Consequently, the results of Kim and Lee (2019) would not apply.

2 General Setting and Estimation

2.1 Setup

We now generalize the previous setup to multiple periods and different treatment switching regimes. We consider balanced panel data with units $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$, with a period $1 < t^* \leq T$ in which the switchers change from treatment status d^{pre} to d^{post}. We refer to the periods $t < t^*$ as the "pre-switch" periods and $t \geq t^*$ as the "post-switch periods". As before, $S = 1$ denotes the switchers and $S = 0$ denotes the stayers, who remain at treatment status d^0 throughout the observed period. In addition to the canonical DiD, where $d^0 = d^{\mathsf{pre}} = 0$ and $d^{\mathsf{post}} = 1$, this setup encompasses universal adoption settings where $d^0 = d^{\mathsf{post}} = 1$ and $d^{\mathsf{pre}} = 0$, policy reversal settings illustrated in Figure 3(a) where $d^0 = d^{\mathsf{post}} = 0$ and $d^{\mathsf{pre}} = 1$, and settings where the treatment statuses do not necessarily converge, $d^0 \neq d^{\mathsf{post}}$, as illustrated in Figure 3(b). Appendix A lists several examples of each setting in the empirical literature.

For any $t' < t$, we let $\mathbf{d}_{t:t'} = (d_t, d_{t-1}, \ldots, d_{t'})$ denote the $(t-t'+1)$-dimensional vector of treatments up to period t starting from period t', with support $\mathcal{D}_{t:t'}$. The potential outcome at time t can depend on all the treatment values up to time t. The vector of treatment statuses is $\mathbf{d}^0_{t:1} = (d^0, d^0, \ldots, d^0)$ for stayers and $\mathbf{d}^{\mathsf{pre}}_{t:1} = (d^{\mathsf{pre}}, d^{\mathsf{pre}}, \ldots, d^{\mathsf{pre}})$ for $t < t^*$ and $(\mathbf{d}^{\mathsf{post}}_{t:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) = (d^{\mathsf{post}}, \ldots, d^{\mathsf{post}}, d^{\mathsf{pre}}, \ldots, d^{\mathsf{pre}})$ for $t \geq t^*$ for switchers, where the switch occurs at time t^*. The observed outcome and treatment status in each period are denoted by Y_t and D_t, respectively. We assume the observed data obeys the following sampling scheme, which is standard in panel data settings.

Assumption 1 (Sampling and moments)

1. *Observations* $(Y_{i1}, Y_{i2}, \ldots, Y_{iT}, D_{i1}, D_{i2}, \ldots, D_{iT})_{i=1}^n$ *are iid draws from an infinite superpopulation of units.*

2. $0 < \mathbb{P}[S = 1] < 1$ *and for all* $\mathbf{d}_{t:1} \in \mathcal{D}_{t:1}$, $\mathbb{E}[Y_t(\mathbf{d}_{t:1})^2] < \infty$.

The goal is to identify the effect of the treatment switch from d^{pre} to d^{post} on the switchers,

$$\mathbb{E}[Y_t(\mathbf{d}^{\mathsf{post}}_{t:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) - Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1})|S=1],$$

which may vary over time. Thus, the relevant counterfactual for switchers is $\mathbb{E}[Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1})|S=1]$, that is, the outcome that would have been observed had the switchers remained at their initial treatment status. Ideally, the researcher would compare the outcome evolution of switchers to stayers that remain at treatment status d^{pre}. This comparison relies on the canonical parallel-trends assumption, stated as follows.

Assumption 2 (Canonical parallel-trends assumption) *For all t and t',*

$$\mathbb{E}[Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1}) - Y_{t'}(\mathbf{d}^{\mathsf{pre}}_{t':1})|S=1] = \mathbb{E}[Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1}) - Y_{t'}(\mathbf{d}^{\mathsf{pre}}_{t':1})|S=0].$$

Assumption 2 generalizes Assumption A and states that the outcomes would exhibit the same trajectory across groups at the pre-switch status of the switchers. In our wealth tax example, this requires stayers and switchers to experience similar wealth growth paths when there is no wealth tax.

Because there are no stayers at d^{pre}, the switchers have to be compared instead to stayers that remain at d^0. Thus, the canonical parallel-trends assumption is not sufficient to identify the effect on the switchers. We introduce the following assumption to restore identification.

Assumption 3 (Time-invariant non-cumulative effects) *For all t and t',*

$$\mathbb{E}[Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1}) - Y_t(\mathbf{d}^{0}_{t:1})|S=0] = \mathbb{E}[Y_{t'}(\mathbf{d}^{\mathsf{pre}}_{t':1}) - Y_{t'}(\mathbf{d}^{0}_{t':1})|S=0].$$

Assumption 3 generalizes Assumption B and implies that, for stayers, the average effect of receiving treatment level d^{pre} for t periods compared to receiving treatment d^0 for t periods is the same, regardless of the period in which this comparison is made. As discussed previously, this means that the treatment effects do not vary and do not accumulate over time. Sufficient conditions for Assumption 3 are that (i) potential outcomes are static, that is, they do not depend on past treatments, $Y_t(d_t, d_{t-1}, \ldots, d_1) = Y_t(d_t)$, and (ii) that the average treatment effects on the stayers comparing d^{pre} to d^0 is time invariant, $\mathbb{E}[Y_t(d^{\mathsf{pre}}) - Y_t(d^0)|S=0] = \mathbb{E}[Y_{t'}(d^{\mathsf{pre}}) - Y_{t'}(d^0)|S=0]$ for any t and t'.

When Assumptions 2 and 3 hold simultaneously,

$$\mathbb{E}[Y_t(\mathbf{d}^{\mathsf{pre}}_{t:1}) - Y_{t'}(\mathbf{d}^{\mathsf{pre}}_{t':1})|S=1] = \mathbb{E}[Y_t(\mathbf{d}^{0}_{t:1}) - Y_{t'}(\mathbf{d}^{0}_{t':1})|S=0], \tag{2}$$

a condition that generalizes condition (1).[2] As a result, the DiD estimands recover the effects

[2] Alternatively, to guarantee this condition, one may assume that Assumption 2 holds under the stayers'

on the switchers, as we discuss next.

2.2 Event-Study Designs

Consider the standard event-study specification:

$$Y_{it} = \alpha_i + \delta_t + \sum_{\ell \neq \tilde{t}} \beta_\ell D_{it}^\ell + \varepsilon_{it}, \quad D_{it}^\ell = S_i \mathbb{1}(t = \ell), \tag{3}$$

where α_i are unit fixed effects, δ_t are time effects, and $\tilde{t} < t^*$ is the baseline or reference period. We refer to the linear projection coefficients β_ℓ for all $\ell \geq t^*$ as the "post-switch coefficients" and for $\ell < t^*$ as the "pre-switch coefficients".

Proposition 1 *Under Assumption 1,*

$$\hat{\beta}_\ell = \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}})\, S_i}{\sum_{i=1}^n S_i} - \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}})\,(1 - S_i)}{\sum_{i=1}^n (1 - S_i)} \to_{\mathbb{P}} \beta_\ell$$

where for $\ell \geq t^*$,

$$\begin{aligned}
\beta_\ell = {} & \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{post}}_{\ell:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) - Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) | S = 1] \\
& + \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) | S = 1] - \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) | S = 0] \\
& + \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_\ell(\mathbf{d}^{0}_{\ell:1}) | S = 0] - \mathbb{E}[Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) - Y_{\tilde{t}}(\mathbf{d}^{0}_{\tilde{t}:1}) | S = 0],
\end{aligned}$$

and for $\ell < \tilde{t}$,

$$\begin{aligned}
\beta_\ell = {} & \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) | S = 1] - \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) | S = 0] \\
& + \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_\ell(\mathbf{d}^{0}_{\ell:1}) | S = 0] - \mathbb{E}[Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) - Y_{\tilde{t}}(\mathbf{d}^{0}_{\tilde{t}:1}) | S = 0].
\end{aligned}$$

This result shows that the post-switch coefficients recover the sum of the effect on the switchers, the difference in trends under d^{pre}, and the change over time of the treatment effect on the stayers. The pre-switch coefficients capture the latter two terms in pre-periods.

Proposition 1 demonstrates that when Assumptions 2 and 3 hold, the post-switch coefficients recover the average effect on the switchers in period ℓ, while pre-switch coefficients β_ℓ are equal to zero. However, when the treatment effect on the stayers varies over time, i.e. when Assumption 3 fails, the post-switch coefficients recover a combination of the effect of the treatment on the switchers and the change of the effect on the stayers over time, while the pre-switch coefficients capture the change in the effect on the switchers in the pre-switch period.

treatment status d_0. All our results hold under this alternative assumption, with the only difference that Assumption 3 has to be imposed on the effects on the switchers instead of the stayers.

Jakobsen et al. (2020) provide a compelling illustration of these issues in an empirical setting. This study explores how wealthy individuals respond to wealth taxes in Denmark. Their first approach exploits a 1989 reform that increased the wealth tax exemption threshold for couples relative to singles. As a result of the reform, some married individuals who were previously subject to wealth taxation became exempt from it. The authors estimate causal effects of this change by comparing couples in the affected wealth range to two plausible comparison groups: (a) single individuals in the same wealth range who were always subject to wealth tax (their preferred specification) and (b) couples in a lower wealth range who were always exempt from wealth tax.[3]

Figure 4 reproduces Jakobsen et al. (2020) findings. Note that comparing treated couples to untreated singles in Figure 4(a) follows the canonical DiD framework: both groups experience the same treatment status (a positive wealth tax) in the baseline period. Consequently, the DiD approach recovers the average effect on the switchers as long as the parallel-trends Assumption 2 holds. However, the comparison of treated couples to couples who were not subject to wealth tax in the first period in Figure 4(b) constitutes an unequal-baseline DiD comparison. Thus, this approach identifies the effect on the switchers under the additional assumption that the wealth tax results in a constant treatment effect. This is unlikely in this setting because wealth taxes lead to mechanical changes in wealth: even if individuals choose not to respond to tax incentives, their wealth accumulates slower (faster) in the presence of higher (lower) wealth tax rates. Indeed, as Figure 4 shows, the first comparison group appears to provide a better comparison than the second, with the raw data showing parallel trends in (a) but divergent trends in (b). The divergent trends observed in Figure 4(b) could either be due to a cumulative treatment effect of wealth taxation (similar to Figure 1(a)) or because groups are not comparable, with no way of telling these apart. The authors assume the former is true and account for the differential trends by including a linear trend term in their specification. We now evaluate the validity of such linear-trend adjustment.

2.3 Linear-Trend Adjustments

A frequently used solution to account for differential trends in the pre-switch period is to include a linear trend in the event-study specification. For example, Jakobsen et al. (2020) account for differential trends in Figures 4(b) and 5 by including "a linear differential pretrend identified based on [...] prereform years (i.e., the omitted years in the first term on the right-hand side)." In an unequal-baseline DiD, time-varying treatment effects will generally result in differential pre-switch trends, and thus, the linear adjustment may help address this issue.

[3]For simplicity and lack of relevance, our discussion abstracts away from other practical issues that may affect individuals' treatment status, e.g. fluctuations of wealth and changes of marital status. The authors address these separately.

To analyze this approach, we introduce some additional notation. Let $\mathcal{T}_{\mathsf{LA}} \subseteq \{1, \ldots, t^* - 1\}$ be the subset of periods used to estimate the linear-trend adjustment, and let $T_{\mathsf{LA}} \geq 2$ be the number of elements in $\mathcal{T}_{\mathsf{LA}}$. We assume that $\mathcal{T}_{\mathsf{LA}}$ consists of consecutive periods, $\mathcal{T}_{\mathsf{LA}} = \{t_m, t_m + 1, \ldots, t_M - 1, t_M\}$, where $1 \leq t_m < t_M \leq t^* - 1$. Define the average and the variance of time periods in set $\mathcal{T}_{\mathsf{LA}}$ as $\bar{t}_{\mathsf{LA}} = \sum_{t \in \mathcal{T}_{\mathsf{LA}}} t/T_{\mathsf{LA}}$, $V_{\mathsf{LA}} = \sum_{t \in \mathcal{T}_{\mathsf{LA}}} (t - \bar{t}_{\mathsf{LA}})^2/T_{\mathsf{LA}}$, and let $\Delta Y_{it} = Y_{it} - Y_{it-1}$ denote one-period differences. We consider the following specification:

$$Y_{it} = \alpha_i + \delta_t + \gamma S_i t + \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \beta_\ell^{\mathsf{LA}} S_i \mathbb{1}(t = \ell) + \eta_{it}, \tag{4}$$

and we refer to γ as the "linear-adjustment coefficient".

Proposition 2 *Under Assumption 1,*

$$\hat{\gamma} = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma \left(\frac{\sum_{i=1}^n \Delta Y_{it} S_i}{\sum_{i=1}^n S_i} - \frac{\sum_{i=1}^n \Delta Y_{it}(1 - S_i)}{\sum_{i=1}^n (1 - S_i)} \right) \to_{\mathbb{P}} \gamma,$$

$$\hat{\beta}_\ell^{\mathsf{LA}} = \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1 - \bar{S})} \left(Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M} \Delta Y_{it} \omega_t^\ell \right) \to_{\mathbb{P}} \beta_\ell^{\mathsf{LA}},$$

where the weights ω_t^γ and ω_t^ℓ are given by:

$$\omega_t^\gamma = \frac{(t - t_m)(t_M + 1 - t)}{2T_{\mathsf{LA}} V_{\mathsf{LA}}}, \quad \omega_t^\ell = \left(\frac{t_M + 1 - t}{T_{\mathsf{LA}}} \right) + (\ell - \bar{t}_{\mathsf{LA}}) \frac{(t - t_m)(t_M + 1 - t)}{2T_{\mathsf{LA}} V_{\mathsf{LA}}}.$$

Furthermore,

$$\gamma = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}}) | S = 1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0) | S = 0] \right),$$

for $\ell \geq t^*$,

$$\begin{aligned} \beta_\ell^{\mathsf{LA}} &= \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:t^*}^{\mathsf{post}}, \mathbf{d}_{t^*-1:1}^{\mathsf{pre}}) - Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) \middle| S = 1 \right] \\ &\quad + \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) - Y_{t_m}(\mathbf{d}_{t_m:1}^{\mathsf{pre}}) \middle| S = 1 \right] - \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^0) - Y_{t_m}(\mathbf{d}_{t_m:1}^0) \middle| S = 0 \right] \\ &\quad - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}}) | S = 1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0) | S = 0] \right) \end{aligned}$$

and for $\ell < t^*$, $\ell \notin \mathcal{T}_{\mathsf{LA}}$,

$$\beta_\ell^{\mathsf{LA}} = \mathbb{E}[Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) - Y_{t_m}(\mathbf{d}_{t_m:1}^{\mathsf{pre}}) | S = 1] - \mathbb{E}[Y_\ell(\mathbf{d}_{\ell:1}^0) - Y_{t_m}(\mathbf{d}_{t_m:1}^0) | S = 0]$$

$$- \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}})|S=1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^{0}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{0})|S=0] \right).$$

Proposition 2 shows that the inclusion of a linear-adjustment coefficient γ adjusts the DiD estimators by subtracting a weighted average of pre-switch differences in trends between switchers and stayers. The proof in Appendix B shows that $\omega_t^\ell \geq 0$ and $\sum_{t=t_m+1}^{t_M} \omega_t^\ell = \ell - t_m$. The weights ω_t^γ are non-negative, sum up to one, and are quadratic in t, with the largest weight given to the middle of the estimation period $\mathcal{T}_{\mathsf{LA}}$ (specifically, period $\bar{t}+0.5$), and are decreasing towards early and late periods t_m and t_M, respectively. Note that all the weights are non-random and observable, since they are functions of the time periods only, so they can be easily calculated in any application.

The coefficient γ is equal to zero under Assumptions 2 and 3. Therefore, the inclusion of a linear trend helps evaluate the validity of the combined assumptions. However, the test may fail in two circumstances. First, if differences in trends are volatile over time, the negative differences may cancel out with positive differences, resulting in an approximately zero estimate of $\hat{\gamma}$ even though pre-trends are not parallel. Similarly, the quadratic nature of the weights ω_l^γ implies that the test may fail to detect differences in trends if these differences are small in the vicinity of period $\bar{t}+0.5$ but large otherwise. Visual examination of the data is useful to rule out such possibilities.

The linear-adjustment specification allows the researcher to replace the requirement of constant, non-cumulative treatment effects with the assumption that treatment effects are linear over time.

Assumption 4 (Linear treatment effects on the stayers) *For all t, there is a constant κ independent of t such that:*

$$\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{it}(\mathbf{d}_{t:1}^{0})|S=0] - \mathbb{E}[Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{0})|S=0] = \kappa.$$

Under this assumption, the following result holds.

Corollary 1 *Under Assumptions 2 and 4, $\gamma = \kappa$ and for $\ell \geq t^*$,*

$$\beta_\ell^{\mathsf{LA}} = \mathbb{E}[Y_\ell(\mathbf{d}_{\ell:t^*}^{\mathsf{post}}, \mathbf{d}_{t^*-1:1}^{\mathsf{pre}}) - Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}})|S=1],$$

while for $\ell < t^*$, $\ell \notin \mathcal{T}_{\mathsf{LA}}$, $\beta_\ell^{\mathsf{LA}} = 0$.

The above result shows that under the canonical parallel-trends assumption and under linearity of the treatment effects, the post-switch coefficients recover the average effects on the switchers, the pre-switch coefficients are equal to zero, and the linear-adjustment coefficient equals the change in the treatment effect on the stayers between consecutive periods.

While Assumption 4 can be thought of as a generalization of Assumption 3 because it allows for time-varying treatment effects, it does so in a parametric way. If treatment effects vary nonlinearly over time, both the event-study estimators and the estimators with a linear adjustment are inconsistent, and it is not possible in general to determine which asymptotic bias is larger. We illustrate this point with a simple three-period example. Suppose there are two pre-switch periods, and one post-switch period, i.e., $t = 1, 2, 3$ and $t^* = T = 3$. Suppose that the canonical parallel-trends Assumption 2 holds. The event-study estimator uses period $t = 2$ as the reference period $\tilde{t}$. The linear-adjustment estimator uses the first two periods to estimate the linear-trend component, so that $\mathcal{T}_{\mathsf{LA}} = \{1, 2\}$. The effect of the switch is then estimated in period 3. By Propositions 1 and 2,

$$\begin{aligned}\hat{\beta}_3 \to_{\mathbb{P}} \beta_3 = \mathbb{E}\left[Y_3(d^{\mathsf{post}}, \mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) \middle| S = 1\right] \\ + \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^{0}_{3:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right]\end{aligned}$$

and

$$\begin{aligned}\hat{\beta}^{\mathsf{LA}}_3 \to_{\mathbb{P}} \beta^{\mathsf{LA}}_3 = \mathbb{E}\left[Y_3(d^{\mathsf{post}}, \mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) \middle| S = 1\right] \\ + \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^{0}_{3:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right] \\ - \left(\mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_1(d^{\mathsf{pre}}) - Y_1(d^{0}) \middle| S = 0\right]\right).\end{aligned}$$

Then the bias of the event-study estimator equals the change in the effect for the stayers between periods 2 and 3, whereas the bias of the linear-adjusted estimator is the difference in changes of the effect on the stayers between periods 2 and 3 and periods 1 and 2. We consider the following four cases, each resulting in a different asymptotic bias summarized in Table 1.

Case 1: constant treatment effects.

$$\begin{aligned}0 = \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_1(d_{\mathsf{pre}}) - Y_1(d_0) \middle| S = 0\right] \\ 0 = \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^{0}_{3:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right].\end{aligned}$$

Case 2: linear treatment effects. For some constant κ,

$$\begin{aligned}\kappa = \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_1(d_{\mathsf{pre}}) - Y_1(d_0) \middle| S = 0\right] \\ \kappa = \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^{0}_{3:1}) \middle| S = 0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^{0}_{2:1}) \middle| S = 0\right].\end{aligned}$$

Case 3: concave treatment effects. For some constant κ,

$$3\kappa = \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^0_{2:1})\middle|\, S=0\right] - \mathbb{E}\left[Y_1(d_{\mathsf{pre}}) - Y_1(d_0)\middle|\, S=0\right]$$
$$\kappa = \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^0_{3:1})\middle|\, S=0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^0_{2:1})\middle|\, S=0\right].$$

Case 4: convex treatment effects. For some constant κ,

$$\kappa = \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^0_{2:1})\middle|\, S=0\right] - \mathbb{E}\left[Y_1(d_{\mathsf{pre}}) - Y_1(d_0)\middle|\, S=0\right]$$
$$3\kappa = \mathbb{E}\left[Y_3(\mathbf{d}^{\mathsf{pre}}_{3:1}) - Y_3(\mathbf{d}^0_{3:1})\middle|\, S=0\right] - \mathbb{E}\left[Y_2(\mathbf{d}^{\mathsf{pre}}_{2:1}) - Y_2(\mathbf{d}^0_{2:1})\middle|\, S=0\right].$$

As shown in Table 1, when the trajectory of treatment effects is nonlinear, the asymptotic bias of the linear-trend-adjusted estimators may be larger or smaller in magnitude than the bias from the unadjusted estimators.

Finally, we point out that the choice of $\mathcal{T}_{\mathsf{LA}}$, the set of periods used to estimate the linear trend component, is important for this specification. This choice is subject to a bias-variance trade-off: on the one hand, the more periods are used to estimate the linear component, the more the strategy relies on a parametric functional form assumption, which is susceptible to misspecification. On the other hand, estimating the linear adjustment using fewer periods is less sensitive to functional form assumptions, but is more sensitive to period-specific idiosyncratic shocks in the trends, possibly resulting in imprecise estimators.

2.4 Settings with Staggered Treatments

In the standard DiD setting with equal baselines, a large literature has pointed out that the standard TWFE and event-study specifications do not generally recover causally interpretable parameters in staggered designs where different cohorts switch to treatment in different periods (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). In such settings, these estimation approaches involve valid comparisons of never-treated or not-yet-treated to just-treated units alongside forbidden comparisons of just-treated units to previously-treated units – in other words, the unequal-baseline comparisons discussed in this paper. The literature has proposed an array of solutions, all of which, generally speaking, solve the problem by explicitly excluding the forbidden comparisons (that is, the comparisons of units with unequal baselines) from the computation of the estimators.

Based on these existing results and the ones in this paper, it is easy to see that in an unequal-baseline setting with staggered adoption, the problems with the TWFE specification are amplified because even the comparisons between switchers and stayers, which are valid in a canonical staggered design, constitute forbidden comparisons in our setting.

A staggered-unequal-baseline design, however, offers a simple solution that allows the

researcher to recover average treatment effects on switchers under a canonical parallel-trends assumption without restricting treatment effect heterogeneity. For example, in a universal adoption setting, the researcher can drop the always-treated cohort as well as the periods after all switcher cohorts have entered treatment. In this smaller data set, the researcher can then compare units that enter treatment in each period to the units that are not yet treated, as suggested by Callaway and Sant'Anna (2021) and Sun and Abraham (2021).

3 Discussion and Practical Recommendations

Our results show that the ability of the unequal-baseline DiD approach to recover treatment effects depends on the nature of treatment effects in the setting under study. The unequal-baseline DiD is best applied to settings where (i) treatment effects are fairly immediate and/or constant in nature, (ii) where treatment effects on the stayers have reached the steady-state phase because a sufficiently long time has passed since the stayers were originally treated, or (iii) where treatment effects are immediate and the treatment on the stayers applied only temporarily.

Two approaches can be taken to establish the nature of treatment effects. The most straightforward, although often infeasible, solution is to obtain more data on earlier periods where both groups had a common treatment status, as in the periods before $t = 0$ in Figure 1. Because groups start from the same treatment status in this pre-baseline periods, this can be seen as a canonical DiD. In this case, it is possible to split the time periods into three groups: one initial period in which both groups share the same treatment status, one intermediate period in which they diverge, and a final period in which they converge or further diverge. This allows the researcher to estimate both the effects of the initial policy change and its reversal, and also test whether outcomes revert back to their initial trend in converging settings.

Alternatively, economic reasoning and prior research on the topic can be used to form expectations about treatment effect dynamics, and any frictions that may delay effects. In some settings, researchers would naturally expect a constant treatment effect, perhaps, once allowed for a short adjustment period. For example, an increase in capital gains tax rate is expected to permanently decrease individual's capital gains realizations. With the exception of a plausibly short adjustment period (e.g. due to information frictions), the treatment effect should be constant in nature (Agersnap and Zidar, 2021; Lavecchia and Tazhitdinova, 2024). In these cases, the approach will correctly estimate the average effect on the switchers provided sufficient time has passed from the initial treatment to allow for the treatment effect to apply. On the other hand, non-constant treatment effects are expected in settings where treatment effects tend to accumulate or dissipate over time, or where adjustment frictions are particularly large and lead to delayed responses. For example, non-constant treatment

effects have been observed 10 years after policy changes in studies of knowledge production and invention (e.g. Furman and Stern, 2011; Moser and Voena, 2012; Akcigit et al., 2022), international migration (Kleven et al., 2013), labor outcomes (e.g. Walker, 2013), and health outcomes (e.g. Greenstone and Hanna, 2014; Alsan and Goldin, 2019).

Another option is to impose assumptions on treatment effect heterogeneity and account for it explicitly, such as the linear-treatment-effects Assumption 4 and the trends adjustment considered in Section 2.3. It should be kept in mind that such assumption imposes parametric restrictions that may result in misspecification, as discussed in Section 2.3.

On the other hand, in staggered DiD settings, the problem of unequal baselines can be solved by dropping the stayers and relying on valid comparisons between switchers and not-yet-switched, exploiting the fact that all switchers start from the same baseline treatment, as discussed in Section 2.4.

Finally, in the context of canonical DiDs, Manski and Pepper (2018) and Rambachan and Roth (2023) have proposed that researchers evaluate the sensitivity of treatment effect estimates to deviations from the parallel-trends assumption, instead of imposing such an assumption outright. This type of sensitivity analysis may be particularly useful in unequal-baseline DiD settings because of the likely violations of the parallel-trends assumption. For example, researchers may wish to assess the sensitivity of the estimates to different treatment effect dynamics, instead of assuming that treatment effects are constant or linear over time. We refer the reader to the aforementioned papers for details on this sensitivity analysis.

References

Agersnap, Ole and Owen Zidar, "The tax elasticity of capital gains and revenue-maximizing rates," *American Economic Review: Insights*, 2021, *3* (4), 399–416.

Akcigit, Ufuk, John Grigsby, Tom Nicholas, and Stefanie Stantcheva, "Taxation and Innovation in the Twentieth Century," *The Quarterly Journal of Economics*, February 2022, *137* (1), 329–385.

Alsan, Marcella and Claudia Goldin, "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880–1920," *Journal of Political Economy*, April 2019, *127* (2), 586–638.

Athey, Susan and Guido W. Imbens, "Design-based analysis in Difference-In-Differences settings with staggered adoption," *Journal of Econometrics*, 2022, *226* (1), 62–79.

Bilinski, Alyssa and Laura A. Hatfield, "Nothing to See Here? Non-inferiority Approaches to Parallel Trends and Other Model Assumptions," *arXiv:1805.03273*, 2018.

Bleakley, Hoyt, "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 1–45.

Bleemer, Zachary, "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209," *The Quarterly Journal of Economics*, February 2022, *137* (1), 115–160.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, December 2021, *225* (2), 200–230.

_ , Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna, "Difference-in-Differences with a Continuous Treatment," 2024.

Cao, Yiming and Shuo Chen, "Rebel on the Canal: Disrupted Trade Access and Social Conflict in China, 1650–1911," *American Economic Review*, 2022, *112* (5), 1555–1590.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chemin, Matthieu and Etienne Wasmer, "Using Alsace-Moselle Local Laws to Build a Difference-in-Differences Estimation Strategy of the Employment Effects of the 35-Hour Workweek Regulation in France," *Journal of Labor Economics*, October 2009, *27* (4), 487–524.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl, "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 72–94.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Fuzzy Differences-in-Differences," *The Review of Economic Studies*, April 2018, *85* (2), 999–1028.

_ and _, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110*, 2964–2996.

_ and _, "Two-way fixed effects and Differences-in-Differences with Heterogeneous Treatment Effects: a Survey," *The Econometrics Journal*, 06 2022.

_, _, Félix Pasquier, Doulo Sow, and Gonzalo Vazquez-Bare, "Difference-in-Differences for Continuous Treatments and Instruments with Stayers," SSRN working paper 4011782 2025.

Dube, Arindrajit, T William Lester, and Michael Reich, "Minimum wage effects across state borders: Estimates using contiguous counties," *The review of economics and statistics*, 2010, *92* (4), 945–964.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro, "Pre-event Trends in the Panel Event-Study Design," *American Economic Review*, September 2019, *109* (9), 3307–3338.

Fuest, Clemens, Andreas Peichl, and Sebastian Siegloch, "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany," *American Economic Review*, February 2018, *108* (2), 393–418.

Furman, Jeffrey L. and Scott Stern, "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research," *American Economic Review*, August 2011, *101* (5), 1933–1963.

Giuliano, Laura, "Minimum wage effects on employment, substitution, and the teenage labor supply: Evidence from personnel data," *Journal of Labor Economics*, 2013, *31* (1), 155–194.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, December 2021, *225* (2), 254–277.

Greenstone, Michael and Rema Hanna, "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India," *American Economic Review*, October 2014, *104* (10), 3038–3072.

Grönqvist, Hans, J. Peter Nilsson, and Per-Olof Robling, "Understanding How Low Levels of Early Lead Exposure Affect Children's Life Trajectories," *Journal of Political Economy*, September 2020, *128* (9), 3376–3433.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman, "Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark," *The Quarterly Journal of Economics*, February 2020, *135* (1), 329–388.

Jardim, Ekaterina, Mark C Long, Robert Plotnick, Emma Van Inwegen, Jacob Vigdor, and Hilary Wething, "Minimum-wage increases and low-wage employment: Evidence from Seattle," *American Economic Journal: Economic Policy*, 2022, *14* (2), 263–314.

Kahn-Lang, Ariella and Kevin Lang, "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business & Economic Statistics*, 2019, *38* (3), 613–620.

Kim, Kimin and Myoung jae Lee, "Difference-in-Differences in Reverse," *Empirical Economics*, 2019, *57* (3), 705–725.

Kleven, Henrik Jacobsen and Esben Anton Schultz, "Estimating Taxable Income Responses Using Danish Tax Reforms," *American Economic Journal: Economic Policy*, November 2014, *6* (4), 271–301.

_ , Camille Landais, and Emmanuel Saez, "Taxation and International Migration of Superstars: Evidence from the European Football Market," *American Economic Review*, August 2013, *103* (5), 1892–1924.

Kotchen, Matthew J. and Laura E. Grant, "Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana," *The Review of Economics and Statistics*, November 2011, *93* (4), 1172–1185.

Lavecchia, Adam M and Alisa Tazhitdinova, "Permanent and transitory responses to capital gains taxes: Evidence from a lifetime exemption in Canada," *Review of Economics and Statistics*, 2024, pp. 1–45.

Lucas, Adrienne M., "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 46–71.

Manski, Charles F. and John V. Pepper, "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *The Review of Economics and Statistics*, 2018, *100* (2), 232–244.

Mora, Ricardo and Iliana Reggio, "Alternative diff-in-diffs estimators with several pre-treatment periods," *Econometric Reviews*, 2019, *38* (5), 465–486.

Moser, Petra and Alessandra Voena, "Compulsory Licensing: Evidence from the Trading with the Enemy Act," *American Economic Review*, February 2012, *102* (1), 396–427.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, 02 2023, *90* (5), 2555–2591.

Rossi, Pauline and Paola Villar, "Private Health Investments under Competing risks: Evidence from Malaria Control in Senegal," *Journal of Health Economics*, September 2020, *73*, 102330.

Roth, Jonathan, "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, September 2022, *4* (3), 305–22.

_ **and Pedro HC Sant'Anna**, "When is Parallel Trends Sensitive to Functional Form?," *Econometrica*, 2023, *91*, 737–747.

_ **, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe**, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Sawada, Yasuyuki, Takeshi Aida, Andrew S. Griffen, Eiji Kozuka, Haruko Noguchi, and Yasuyuki Todo, "Democratic Institutions and Social Capital: Experimental Evidence on School-Based Management from a Developing Country," *Journal of Economic Behavior & Organization*, June 2022, *198*, 267–279.

Slattery, Cailin, Alisa Tazhitdinova, and Sarah Robinson, "Corporate Political Spending and State Tax Policy: Evidence from Citizens United," *Journal of Public Economics*, 2023, *221*, 104859.

Steigerwald, Douglas G., Gonzalo Vazquez-Bare, and Jason Maier, "Measuring Heterogeneous Effects of Environmental Policies Using Panel Data," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (2), 277–313.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, December 2021, *225* (2), 175–199.

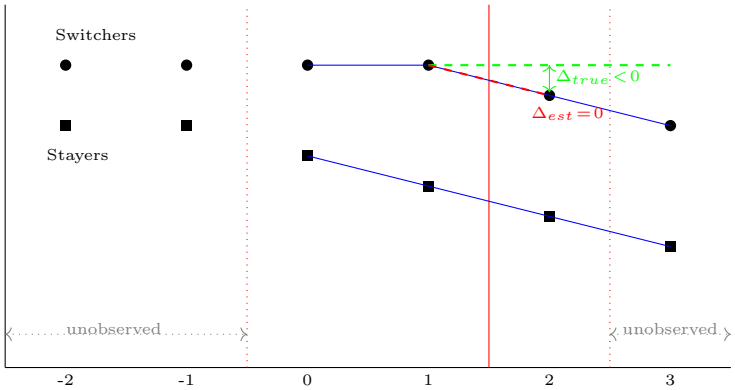
Walker, W. Reed, "The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce," *The Quarterly Journal of Economics*, November 2013, *128* (4), 1787–1835.

Yagan, Danny, "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut," *American Economic Review*, December 2015, *105* (12), 3531–3563.

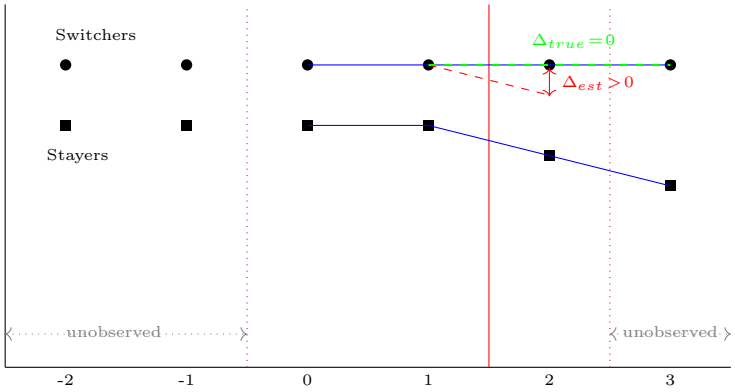
Zwick, Eric and James Mahon, "Tax Policy and Heterogeneous Investment Behavior," *American Economic Review*, 2017, *107*.

Figure 1: Universal Adoption Settings

(a) Immediate Treatment Effects: Negative Treatment Effect Appears Immediately.



(b) Delayed Treatment Effects: Negative Treatment Effect Appears Two Periods After Treatment.

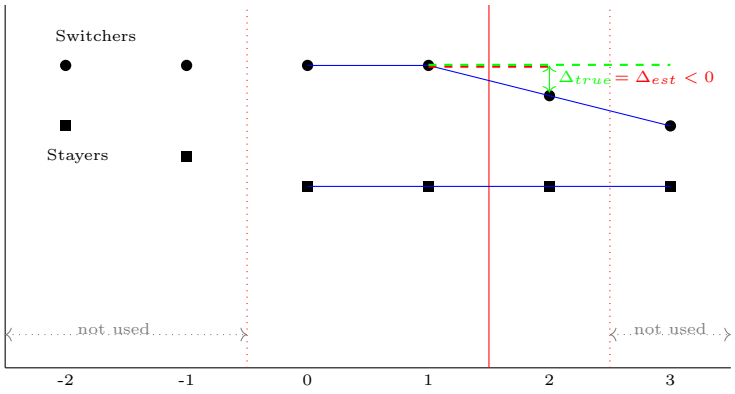


Notes: This figure illustrates the possible evolution of outcomes for switchers (circles) and stayers (squares) in a setting with a cumulative negative treatment effect. The treatment is imposed on stayers in periods $t = 0, 1, 2, 3$, and on switchers in periods $t = 2, 3$. No treatment is imposed in periods $t = -1, -2$. Figure (a) assumes that the negative treatment effect appears immediately, while Figure (b) assumes that the negative treatment effect appears two periods after the onset of treatment. The Δ_{true} identifies the true ATT, while the Δ_{est} identifies the estimated ATT.

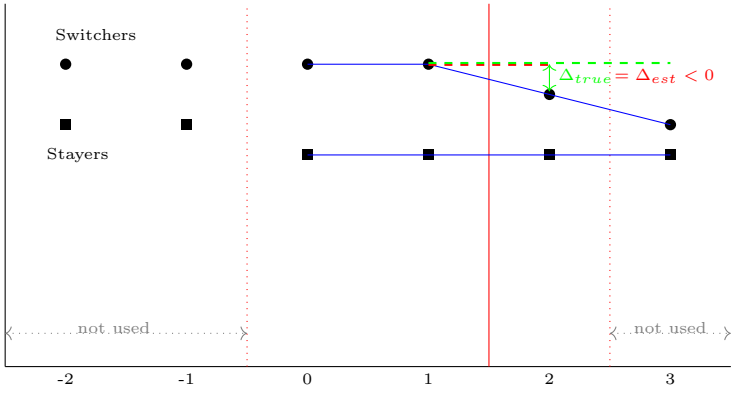
The x-axis identifies time periods, and the y-axis identifies outcome values. It is assumed that the researcher is only able to observe and use in his analysis data from periods $t = 0, 1, 2$, while the data from earlier and later periods are unobservable. The solid blue lines identify the observed outcome paths for the stayers and the switchers, while the dashed green line identifies the true counterfactual outcome path for the switchers in the absence of treatment. The red dashed line identifies the predicted counterfactual path for the switchers that a DiD would employ.

Figure 2: Completed and Temporary Treatments

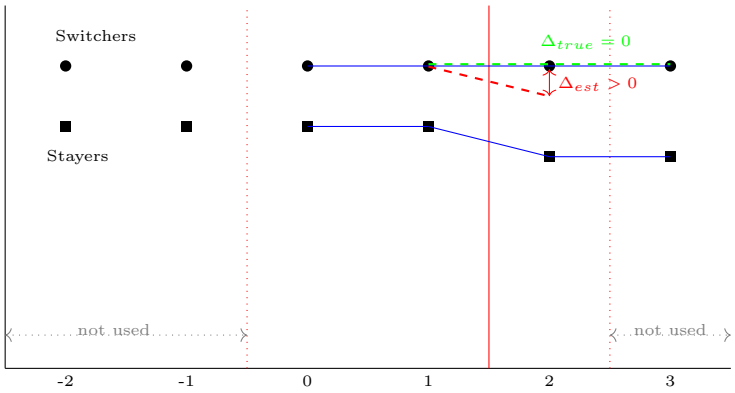
(a) Treatment on the Stayers Completes Prior to $t = 0$



(b) Temporary but Immediate Treatments



(c) Temporary but Non-Immediate Treatments



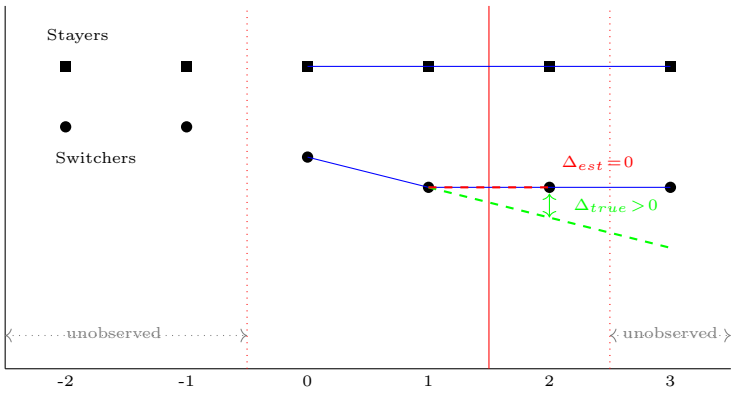
Notes: This figure illustrates the possible evolution of outcomes for switchers (circles) and stayers (squares) in a setting with a negative treatment effect. In figure (a), the treatment is imposed on stayers in periods $t = 0$ and on switchers in periods $t = 2, 3$. It is assumed that the treatment effect is negative and immediate. In Figure (b), the treatment is imposed on stayers in periods $t = 0$, and on switchers in periods $t = 2, 3$. However, it is assumed that the treatment effect appears two days after the onset of treatment.

The Δ_{true} identifies the true ATT, while the Δ_{est} identifies the estimated ATT.

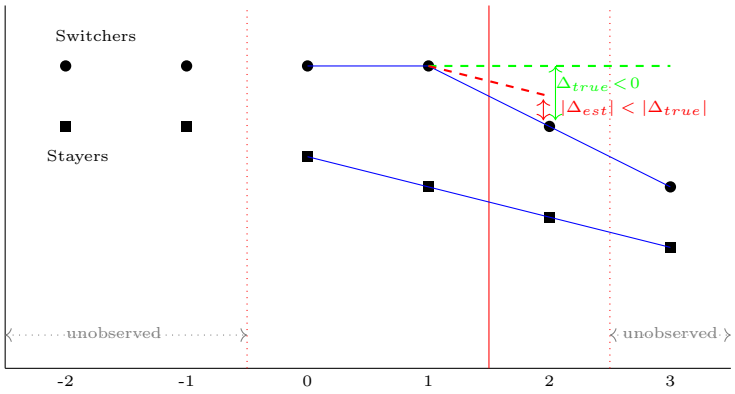
The x-axis identifies time periods, and the y-axis identifies outcome values. It is assumed that the researcher is only able to observe and use in his analysis data from periods $t = 0, 1, 2$, while the data from earlier and later periods are unobservable. The solid blue lines identify the observed outcome paths for the stayers and the switchers, while the dashed green line identifies the true counterfactual outcome path for the switchers in the absence of treatment. The red dashed line identifies the predicted counterfactual path for the switchers that a DiD would employ.

Figure 3: Policy Reversals and Nonconvergent Settings

(a) Policy Reversal Case with Immediate Treatment Effects



(b) Nonconvergent Treatments with Immediate Treatment Effects



Notes: This figure illustrates the possible evolution of outcomes for switchers (circles) and stayers (squares) in a setting with a cumulative negative treatment effect. In Figure (a), the treatment is never imposed on the stayers, and only imposed on the switchers in periods $t = 0, 1$. No treatment is imposed in periods $t = -1, -2$. In Figure (b), the treatment is imposed on stayers in periods $t = 0, 1, 2, 3$, and a stronger treatment is imposed on switchers in periods $t = 2, 3$. No treatment is imposed in periods $t = -1, -2$. The Δ_{true} identifies the true ATT, while the Δ_{est} identifies the estimated ATT.

The x-axis identifies time periods, and the y-axis identifies outcome values. It is assumed that the researcher is only able to observe and use in his analysis data from periods $t = 0, 1, 2$, while the data from earlier and later periods are unobservable. The solid blue lines identify the observed outcome paths for the stayers and the switchers, while the dashed green line identifies the true counterfactual outcome path for the switchers in the absence of treatment. The red dashed line identifies the predicted counterfactual path for the switchers that a DiD would employ.

Figure 4: Empirical Illustration – Convergent Treatment Status: Jakobsen et al. (2020)

(a) Equal baseline status:
both control and treated are taxed

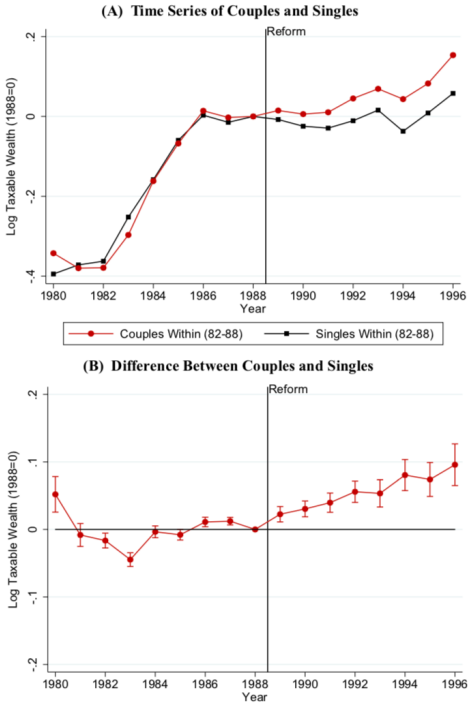
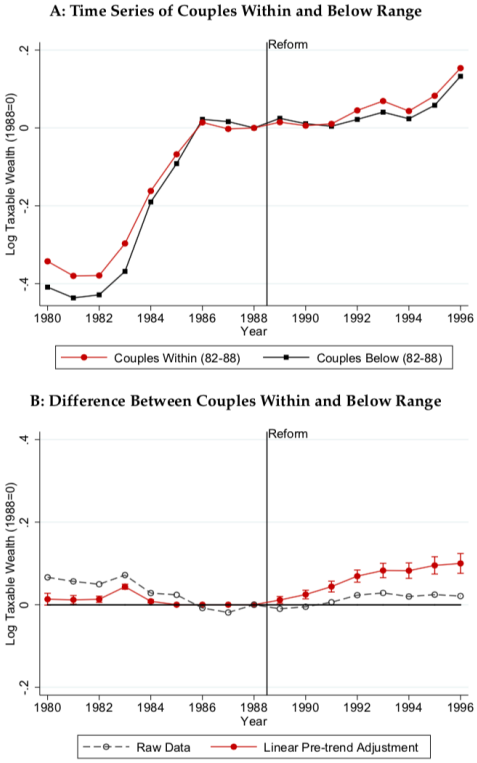


FIGURE IV

Difference-in-Differences Comparing Couples and Singles within Exempted Range

(b) Unequal baseline status:
treated are taxed but control are not



Notes: This figure reproduces (a) Figure IV and (b) Appendix Figure A.VII from Jakobsen et al. (2020). These figures show the evolution of taxable wealth and the difference between treatment and control groups before and after the 1989 reform that increased the exemption threshold for couples but not for single individuals. Loosely speaking, in both figures, the treatment group consists of couples who were subject to wealth tax before 1989 but not after 1989. In (a), the control group consists of single individuals who were subject to wealth tax before and after 1989 – the "Singles" group, while in (b), the control group consists of couples who were exempt from wealth tax both before and after 1989 – the "Below Range" group. In all figures, the treatment group is shown in red dots, while the control group is shown in black squares.

Figure 5: Empirical Illustration – Non-Convergent Treatment Status: Jakobsen et al. (2020)

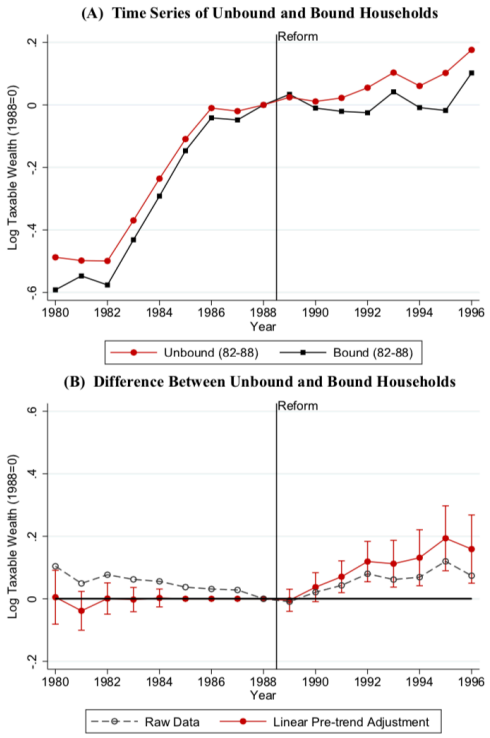


FIGURE VI

Difference-in-Differences Comparing Households Unbound and Bound by Tax Ceiling

Notes: This figure reproduces Figure VI from Jakobsen et al. (2020). These figures show the evolution of taxable wealth and the difference between treatment and control groups before and after the 1989 reform that reduced the wealth tax rate from 2.2% to 1% on the very wealthy – "the Unbound" group. The authors compare wealthy individuals to a control group subject to a 0% marginal tax rate because of a tax ceiling provision ("Det Vandrette Skatteloft"), both before and after 1989 – the "Bound" group. The treatment group is shown in red dots, while the control group is shown in black squares.

Table 1: Asymptotic bias - Event-Study vs Linear-Adjustment Estimators

	Event Study Estimator	Linear Adjustment Estimator
Case 1: constant treatment effects	0	0
Case 2: linear treatment effects	κ	0
Case 3: concave treatment effects	κ	-2κ
Case 4: convex treatment effects	3κ	2κ

Notes: This table illustrates how the asymptotic bias relates to the nature of treatment effects and the estimator used. The asymptotic bias is calculated for four cases of treatment effects described in Section 2.3.

APPENDIX

A Examples of Unequal-Baseline DiDs

Universal adoption. This setting has been explored in environmental economics to study the energy effects of daylight saving time (DST) by Kotchen and Grant (2011), who exploit a 2006 reform that resulted in a universal adoption of DST in Indiana. Prior to the studied policy change, 77 counties did not practice DST while 15 did. In the post-treatment period, all counties switched to DST. Sawada et al. (2022) study introduction of a formal, democratic local school-based management (SBM) institution in rural Burkina Faso. This study compares schools that implemented SBM in the first period to schools that implemented SBM in the second period. In labor economics, Chemin and Wasmer (2009) study the employment effects of working hour reductions. They exploit a change in working hours laws in France that resulted in a switch from 39 hours per week to 35 in all areas of France except for Alsace-Moselle where working hours decreased from a lower starting point.

Policy reversal. Bleemer (2022) studies the effects of race-based affirmative action bans on student outcomes. This study compares outcomes of the underrepresented minority (URM) applicants to the outcomes of non-URM students with similar prior academic opportunity and preparation, before and after the 1998 affirmative action ban at California public universities, which removed preferential treatment for URMs. Slattery et al. (2023) study how the cancellation of independent campaign contribution bans due to *Citizens United* ruling affected state tax policies, by comparing tax policies in states that had bans to states that did not. Cao and Chen (2022) study the abandonment of the Grand Canal in China on rebellions. Their treatment group consists of counties through which the canal ran before it was abandoned, while the control group includes distant counties. Grönqvist et al. (2020) study the effects of leaded gasoline phaseout by comparing life outcomes of children born in Swedish neighborhoods with high vs low lead exposures, before and after 1981, when lead levels per liter of gasoline were rapidly reduced in all areas of Sweden. Relatedly, Bleakley (2010), Cutler et al. (2010), Lucas (2010), and Rossi and Villar (2020) study malaria eradication campaigns and exploit variation in malaria prevalence prior to anti-malaria programs as a measure of intensity of treatment.

Non-convergent settings. This version of DiD is ubiquitously used by empirical tax economists to measure the causal effects of taxes. For example, Kleven and Schultz (2014) measure the causal effects of income taxes while Jakobsen et al. (2020) measure the effects of wealth taxes by comparing individuals in different income tax brackets who experience differential changes in relative MTRs. Yagan (2015) studies the effect of dividend taxes

on investment by comparing firms that are subject to dividend taxes (C-corporations) to firms that are not (S-corporations). Zwick and Mahon (2017) study the effects of the bonus depreciation scheme on investment by exploiting differences in the relative magnitude of incentives, while Fuest et al. (2018) study the effect of corporate taxes on wages using variation in local corporate rates. In all these studies, treatment and control groups experience different levels of taxes in the baseline period and the post-period. Similarly, minimum wage studies often compare workers residing in states with different minimum wages or workers residing in the same state but subject to different minimum wage regimes (e.g., Dube et al., 2010; Giuliano, 2013; Cengiz et al., 2019; Jardim et al., 2022). In such comparisons, workers in the treatment and control groups are typically subject to different minimum wage rules before and after the reform.

B Proofs

Notation. For any variable A_{it} and any subset $\mathcal{T} \subseteq \{1, 2, \ldots, T\}$ of cardinality $|\mathcal{T}|$, define $\bar{A}_i(\mathcal{T}) = \sum_{t\in\mathcal{T}} A_{it}/|\mathcal{T}|$, $\tilde{A}_t = \sum_i A_{it}/n$, $\bar{A}(\mathcal{T}) = \sum_i \sum_{t\in\mathcal{T}} A_{it}/(n|\mathcal{T}|)$ and $\ddot{A}_{it}(\mathcal{T}) = A_{it} - \bar{A}_i(\mathcal{T}) - \tilde{A}_t + \bar{A}(\mathcal{T})$. We also denote the population counterpart of the double-demeaned $\ddot{A}_{it}(\mathcal{T})$ as $\check{A}_{it}(\mathcal{T}) = A_{it} - \bar{A}_i(\mathcal{T}) - \mathbb{E}[A_{it}] + \mathbb{E}[\bar{A}_i(\mathcal{T})]$. Also let $X^\ell_{it} = \mathbb{1}(t = \ell)S_i$ and $Z_{it} = tS_i$. To reduce notation we will not distinguish between lead and lag coefficients. As discussed in the paper, we consider the event-study regression:

$$Y_{it} = \alpha_i + \delta_t + \sum_{\ell\neq\tilde{t}} \beta_\ell X^\ell_{it} + \varepsilon_{it}$$

where

$$\ddot{X}^\ell_{it}(\mathcal{T}) = (S_i - \bar{S})\left(\mathbb{1}(t = \ell) - \frac{1}{|\mathcal{T}|}\right)$$

and the event-study regression with a linear adjustment:

$$Y_{it} = \alpha_i + \delta_t + \sum_{\ell\notin\mathcal{T}_{\mathsf{LA}}} \beta^{\mathsf{LA}}_\ell X^\ell_{it} + \gamma Z_{it} + u_{it}$$

where

$$\ddot{Z}_{it}(\mathcal{T}) = (S_i - \bar{S})\left(t - \sum_{t\in\mathcal{T}} \frac{t}{|\mathcal{T}|}\right)$$

and where $\mathcal{T}_{\mathsf{LA}}$ is the subset of periods used to estimate the linear trend and $T_{\mathsf{LA}} = |\mathcal{T}_{\mathsf{LA}}|$ (see Section 2.3 for details).

B.1 Proof of Proposition 1

The population coefficients $\{\beta_\ell\}_\ell$ are characterized by the system of equations:

$$0 = \sum_t \mathbb{E}\left[(\check{Y}_{it} - \sum_{\ell \neq \tilde{t}} \beta_\ell \check{X}^\ell_{it}) X^\ell_{it}\right]$$

whereas the estimators $\{\hat{\beta}_\ell\}_{\ell \neq \tilde{t}}$ are characterized by the sample analog:

$$0 = \sum_i \sum_t (\ddot{Y}_{it} - \sum_{\ell \neq \tilde{t}} \hat{\beta}_\ell \ddot{X}^\ell_{it}) X^\ell_{it} = \sum_i (\ddot{Y}_{i\ell} - \hat{\beta}_\ell \ddot{X}^\ell_{i\ell} - \sum_{m \neq \ell, \tilde{t}} \hat{\beta}_m \ddot{X}^m_{i\ell}) S_i$$

We show the derivation for the OLS estimators, as the one for the population coefficients follows an identical reasoning replacing sample averages by expectations. We have that:

$$0 = \sum_i \ddot{Y}_{i\ell} S_i - \hat{\beta}_\ell n\bar{S}(1-\bar{S})\left(1 - \frac{1}{T}\right) + \sum_{m \neq \ell, \tilde{t}} \hat{\beta}_m n\bar{S}(1-\bar{S})\frac{1}{T}$$

from which

$$\hat{\beta}_\ell = \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} + \frac{1}{T}\sum_{\ell \neq \tilde{t}} \hat{\beta}_\ell. \tag{5}$$

Summing over $\ell \neq \tilde{t}$,

$$\sum_{\ell \neq \tilde{t}} \hat{\beta}_\ell = \sum_{\ell \neq \tilde{t}} \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} + \frac{T-1}{T}\sum_{\ell \neq \tilde{t}} \hat{\beta}_\ell$$

and thus

$$\frac{1}{T}\sum_{\ell \neq \tilde{t}} \hat{\beta}_\ell = -\frac{\sum_i \ddot{Y}_{i\tilde{t}} S_i}{n\bar{S}(1-\bar{S})}$$

using that $\sum_{\ell \neq \tilde{t}} \ddot{Y}_{i\ell} = \sum_\ell \ddot{Y}_{i\ell} - \ddot{Y}_{i\tilde{t}} = -\ddot{Y}_{i\tilde{t}}$. Plugging back into (5),

$$\hat{\beta}_\ell = \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} - \frac{\sum_i \ddot{Y}_{i\tilde{t}} S_i}{n\bar{S}(1-\bar{S})} = \frac{\sum_i (\ddot{Y}_{i\ell} - \ddot{Y}_{i\tilde{t}}) S_i}{n\bar{S}(1-\bar{S})}.$$

Finally, use that $\ddot{Y}_{i\ell} - \ddot{Y}_{i\tilde{t}} = Y_{i\ell} - Y_{i\tilde{t}} - \sum_i (Y_{i\ell} - Y_{i\tilde{t}})/n$ to get

$$\hat{\beta}_\ell = \frac{\sum_i (Y_{i\ell} - Y_{i\tilde{t}})(S_i - \bar{S})}{n\bar{S}(1-\bar{S})} = \frac{\sum_i (Y_{i\ell} - Y_{i\tilde{t}}) S_i}{\sum_i S_i} - \frac{\sum_i (Y_{i\ell} - Y_{i\tilde{t}})(1 - S_i)}{\sum_i (1 - S_i)}$$

and by the law of large numbers, as $n \to \infty$, $\hat{\beta}_\ell \to_{\mathbb{P}} \beta_\ell = \mathbb{E}[Y_\ell - Y_{\tilde{t}}|S=1] - \mathbb{E}[Y_\ell - Y_{\tilde{t}}|S=0]$. Next, if $\ell \geq t^*$,

$$\begin{aligned}
\beta_\ell &= \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{post}}_{\ell,t^*} : \mathbf{d}^{\mathsf{pre}}_{t^*-1,1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1})|S=1] - \mathbb{E}[Y_\ell(\mathbf{d}^0_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^0_{\tilde{t}:1})|S=0] \\
&= \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{post}}_{\ell:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) - Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1})|S=1] \\
&+ \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1})|S=1] - \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1})|S=0] \\
&+ \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_\ell(\mathbf{d}^0_{\ell:1})|S=0] - \mathbb{E}[Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) - Y_{\tilde{t}}(\mathbf{d}^0_{\tilde{t}:1})|S=0]
\end{aligned}$$

and when $\ell < t^*$,

$$\begin{aligned}
\beta_\ell &= \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1})|S=1] - \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1})|S=0] \\
&+ \mathbb{E}[Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1}) - Y_\ell(\mathbf{d}^0_{\ell:1})|S=0] - \mathbb{E}[Y_{\tilde{t}}(\mathbf{d}^{\mathsf{pre}}_{\tilde{t}:1}) - Y_{\tilde{t}}(\mathbf{d}^0_{\tilde{t}:1})|S=0].
\end{aligned}$$

which completes the proof. □

B.2 Proof of Proposition 2

As before, we only show the proof for the OLS estimators, since the proof for the linear projection coefficients is analogous after replacing sample averages by expectations. The estimators $\{\hat{\beta}^{\mathsf{LA}}_\ell\}_{\ell \notin \mathcal{T}_{\mathsf{LA}}}$ and $\hat{\gamma}$ are characterized by the system of equations:

$$\begin{aligned}
0 &= \sum_i \sum_t (\ddot{Y}_{it} - \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_\ell \ddot{X}^\ell_{it} - \hat{\gamma}\ddot{Z}_{it}) X^\ell_{it} = \sum_i \sum_t (\ddot{Y}_{it} - \hat{\beta}^{\mathsf{LA}}_\ell \ddot{X}^\ell_{it} - \sum_{m \neq \ell, m \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_m \ddot{X}^m_{it} - \hat{\gamma}\ddot{Z}_{it}) X^\ell_{it} \\
0 &= \sum_i \sum_t (Y_{it} - \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_\ell X^\ell_{it} - \hat{\gamma} Z_{it}) \ddot{Z}_{it}
\end{aligned}$$

From the first equation,

$$\begin{aligned}
0 &= \sum_i (\ddot{Y}_{i\ell} - \hat{\beta}^{\mathsf{LA}}_\ell \ddot{X}^\ell_{i\ell} - \sum_{m \neq \ell, m \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_m \ddot{X}^m_{i\ell} - \hat{\gamma}\ddot{Z}_{i\ell}) X^\ell_{i\ell} \\
&= \sum_i \ddot{Y}_{i\ell} S_i - \hat{\beta}^{\mathsf{LA}}_\ell \sum_i \ddot{X}^\ell_{i\ell} X^\ell_{i\ell} - \sum_{m \neq \ell, m \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_m \sum_i \ddot{X}^m_{i\ell} X^\ell_{i\ell} - \hat{\gamma} \sum_i \ddot{Z}_{i\ell} X^\ell_{i\ell} \\
&= \sum_i \ddot{Y}_{i\ell} S_i - n\bar{S}(1-\bar{S})\hat{\beta}^{\mathsf{LA}}_\ell \left(1 - \frac{1}{T}\right) + \frac{n\bar{S}(1-\bar{S})}{T} \sum_{m \neq \ell, m \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_m - \hat{\gamma}(\ell - \bar{t}) n\bar{S}(1-\bar{S})
\end{aligned}$$

where $\bar{t} = \sum_t t/T = (T+1)/2$, from which

$$\hat{\beta}^{\mathsf{LA}}_\ell = \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} + \frac{1}{T} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}}_\ell - \hat{\gamma}\,(\ell - \bar{t})\,.$$

Summing over $\ell \notin \mathcal{T}_{\mathsf{LA}}$,

$$\sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}_\ell^{\mathsf{LA}} = \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} + \frac{T - T_{\mathsf{LA}}}{T} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}_\ell^{\mathsf{LA}} - \hat{\gamma} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} (\ell - \bar{t})$$

so that

$$\frac{1}{T} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}_\ell^{\mathsf{LA}} = -\frac{\sum_i S_i \left(\sum_{\ell \in \mathcal{T}_{\mathsf{LA}}} \ddot{Y}_{i\ell} / T_{\mathsf{LA}} \right)}{n\bar{S}(1-\bar{S})} - \frac{\hat{\gamma}}{T_{\mathsf{LA}}} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} (\ell - \bar{t})$$

and thus letting $\bar{\ddot{Y}}_i^{\mathsf{LA}} = \sum_{\ell \in \mathcal{T}_{\mathsf{LA}}} \ddot{Y}_{i\ell} / T_{\mathsf{LA}}$,

$$\begin{aligned}
\hat{\beta}_\ell^{\mathsf{LA}} &= \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} - \frac{\sum_i S_i \left(\sum_{\ell \in \mathcal{T}_{\mathsf{LA}}} \ddot{Y}_{i\ell} / T_{\mathsf{LA}} \right)}{n\bar{S}(1-\bar{S})} - \frac{\hat{\gamma}}{T_{\mathsf{LA}}} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} (\ell - \bar{t}) - \hat{\gamma} (\ell - \bar{t}) \\
&= \frac{\sum_i (\ddot{Y}_{i\ell} - \bar{\ddot{Y}}_i^{\mathsf{LA}}) S_i}{n\bar{S}(1-\bar{S})} - \hat{\gamma} (\ell - \bar{t}_{\mathsf{LA}}) \\
&= \frac{\sum_i (Y_{i\ell} - \bar{Y}_i^{\mathsf{LA}})(S_i - \bar{S})}{n\bar{S}(1-\bar{S})} - \hat{\gamma} (\ell - \bar{t}_{\mathsf{LA}})
\end{aligned}$$

where $\bar{t}_{\mathsf{LA}} = \sum_{t \in \mathcal{T}_{\mathsf{LA}}} t / T_{\mathsf{LA}}$ and $\bar{Y}_i^{\mathsf{LA}} = \sum_{\ell \in \mathcal{T}_{\mathsf{LA}}} Y_{i\ell} / T_{\mathsf{LA}}$. Next, plugging this expression into the first order condition for $\hat{\gamma}$,

$$\begin{aligned}
0 &= \frac{\sum_i \sum_t Y_{it}(S_i - \bar{S})(t - \bar{t})}{n\bar{S}(1-\bar{S})} - \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \hat{\beta}^{\mathsf{LA}} (\ell - \bar{t}) - \hat{\gamma} \sum_t (t - \bar{t}) t \\
&= \frac{\sum_i \sum_t Y_{it}(S_i - \bar{S})(t - \bar{t})}{n\bar{S}(1-\bar{S})} - \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} \frac{\sum_i (Y_{i\ell} - \bar{Y}_i^{\mathsf{LA}})(S_i - \bar{S})(\ell - \bar{t})}{n\bar{S}(1-\bar{S})} \\
&+ \hat{\gamma} \sum_{\ell \notin \mathcal{T}_{\mathsf{LA}}} (\ell - \bar{t}_{\mathsf{LA}}) (\ell - \bar{t}) - \hat{\gamma} \sum_t (t - \bar{t}) t \\
&= \sum_i \frac{S_i - \bar{S}}{n\bar{S}(1-\bar{S})} \left\{ \sum_t Y_{it}(t - \bar{t}) - \sum_{t \notin \mathcal{T}_{\mathsf{LA}}} (Y_{it} - \bar{Y}_i^{\mathsf{LA}})(t - \bar{t}) \right\} \\
&+ \hat{\gamma} \left\{ \sum_{t \notin \mathcal{T}_{\mathsf{LA}}} (t - \bar{t}_{\mathsf{LA}}) (t - \bar{t}) - \sum_t (t - \bar{t}) t \right\} \\
&= \frac{\sum_i (S_i - \bar{S}) \sum_{t \in \mathcal{T}_{\mathsf{LA}}} Y_{it}(t - \bar{t}_{\mathsf{LA}})}{n\bar{S}(1-\bar{S})} - \hat{\gamma} T_{\mathsf{LA}} V_{\mathsf{LA}}
\end{aligned}$$

where $V_{\mathsf{LA}} = \sum_{t\in\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}_{\mathsf{LA}})^2/T_{\mathsf{LA}}$ and where the last equality uses that:

$$\begin{aligned}\sum_{t\notin\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}_{\mathsf{LA}})\,(t-\bar{t}) - \sum_t (t-\bar{t})t &= \sum_{t\notin\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}_{\mathsf{LA}})\,(t-\bar{t}) - \sum_t (t-\bar{t})(t-\bar{t}_{\mathsf{LA}}) \\ &= -\sum_{t\in\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}_{\mathsf{LA}})\,(t-\bar{t}) \\ &= -\sum_{t\in\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}_{\mathsf{LA}})\,t = -T_{\mathsf{LA}}V_{\mathsf{LA}}\end{aligned}$$

and

$$\begin{aligned}\sum_t Y_{it}(t-\bar{t}) - \sum_{t\notin\mathcal{T}_{\mathsf{LA}}}(Y_{it}-\bar{Y}_i^{\mathsf{LA}})(t-\bar{t}) &= \sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}(t-\bar{t}) + \bar{Y}_i^{\mathsf{LA}}\sum_{t\notin\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}) \\ &= \sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}(t-\bar{t}) - \bar{Y}_i^{\mathsf{LA}}\sum_{t\in\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}) \\ &= \sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}(t-\bar{t}) - \frac{1}{T_{\mathsf{LA}}}\sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}\sum_{t\in\mathcal{T}_{\mathsf{LA}}}(t-\bar{t}) \\ &= \sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}(t-\bar{t}_{\mathsf{LA}}).\end{aligned}$$

Finally, recall that $\mathcal{T}_{\mathsf{LA}} = \{t_m, t_m+1, \ldots, t_M-1, t_M\}$ and note that for $t_m \le t \le t_M$, $Y_{it} = Y_{it_m} + \mathbb{1}(t>t_m)\sum_{s=t_m+1}^{t_M}\Delta Y_{is}$ where $\Delta Y_{is} = Y_{is} - Y_{is-1}$, and thus

$$\begin{aligned}\sum_{t\in\mathcal{T}_{\mathsf{LA}}} Y_{it}(t-\bar{t}_{\mathsf{LA}}) &= \sum_{t=t_m}^{t_M}\left(Y_{it_m} + \mathbb{1}(t>t_m)\sum_{s=t_m+1}^{t_M}\Delta Y_{is}\right)(t-\bar{t}_{\mathsf{LA}}) \\ &= \sum_{t=t_m+1}^{t_M}\sum_{s=t_m+1}^{t_M}\Delta Y_{is}(t-\bar{t}_{\mathsf{LA}}) \\ &= \sum_{s=t_m+1}^{t_M}\sum_{t=s}^{t_M}\Delta Y_{is}(t-\bar{t}_{\mathsf{LA}}) \\ &= \sum_{s=t_m+1}^{t_M}\Delta Y_{is}\sum_{t=s}^{t_M}(t-\bar{t}_{\mathsf{LA}})\end{aligned}$$

where the order of summation is reversed using that $\{t_m+1 \le t \le t_M, t_m+1 \le s \le t\}$ is equivalent to $\{t_m+1 \le s \le t_M, s \le t \le t_M\}$. Now,

$$\begin{aligned}\sum_{t=s}^{t_M}(t-\bar{t}_{\mathsf{LA}}) &= \frac{t_M(t_M+1)}{2} - \frac{s(s-1)}{2} - \bar{t}_{\mathsf{LA}}(t_M-s+1) \\ &= \frac{t_M(t_M+1)}{2} - \frac{s(s-1)}{2} - \frac{t_M+t_m}{2}(t_M+1-s)\end{aligned}$$

$$= \frac{t_M(t_M+1)}{2} - \frac{s(s-1)}{2} - \frac{t_M(t_M+1)}{2} + s\frac{t_M}{2} - \frac{t_m}{2}(t_M+1-s)$$
$$= \frac{1}{2}(s-t_m)(t_M+1-s).$$

Collecting these results,

$$\hat{\gamma} = \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1-\bar{S})} \sum_{t=t_m+1}^{t_M} \Delta Y_{it}\omega_t^\gamma = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma \left(\frac{\sum_i \Delta Y_{it} S_i}{\sum_i S_i} - \frac{\sum_i \Delta Y_{it}(1-S_i)}{\sum_i (1-S_i)} \right)$$

where

$$\omega_t^\gamma = \frac{(t-t_m)(t_M+1-t)}{2T_{\mathsf{LA}}V_{\mathsf{LA}}}.$$

which is non-negative for $t_m \leq t \leq t_M$. To see that the weights sum to one, first consider $\sum_{t=t_m+1}^{t_M}(t-t_m)(t_M+1-t)$. Define $u = t - t_m$ so that $t_m + 1 \leq t \leq t_M \Leftrightarrow 1 \leq u \leq t_M - t_m$ or equivalently $1 \leq u \leq T_{\mathsf{LA}} - 1$ where $T_{\mathsf{LA}} = t_M - t_m + 1$. Next, noting that $u = t - t_m$ implies $t_M + 1 - t = T_{\mathsf{LA}} - u$,

$$\begin{aligned}
\sum_{t=t_m+1}^{t_M} (t-t_m)(t_M+1-t) &= \sum_{u=1}^{T_{\mathsf{LA}}-1} u(T_{\mathsf{LA}}-u) \\
&= \frac{T_{\mathsf{LA}}^2(T_{\mathsf{LA}}-1)}{2} - \frac{(T_{\mathsf{LA}}-1)T_{\mathsf{LA}}(2T_{\mathsf{LA}}-1)}{6} \\
&= \frac{(T_{\mathsf{LA}}-1)T_{\mathsf{LA}}}{2}\left(T_{\mathsf{LA}} - \frac{2T_{\mathsf{LA}}-1}{3}\right) \\
&= \frac{(T_{\mathsf{LA}}-1)T_{\mathsf{LA}}(T_{\mathsf{LA}}+1)}{6} \\
&= \binom{T_{\mathsf{LA}}+1}{3}.
\end{aligned}$$

By a similar argument,

$$\begin{aligned}
\sum_{t\in\mathcal{T}_{\mathsf{LA}}} (t-\bar{t}_{\mathsf{LA}})^2 = \sum_{t\in\mathcal{T}} t(t-\bar{t}) &= \sum_{t=t_m}^{t_M} (t-t_m)(t-t_m-(\bar{t}_{\mathsf{LA}}-t_m)) \\
&= \sum_{u=0}^{T_{\mathsf{LA}}-1} u\left(u - \frac{T_{\mathsf{LA}}-1}{2}\right) \\
&= \sum_{u=1}^{T_{\mathsf{LA}}-1} u\left(u - \frac{T_{\mathsf{LA}}-1}{2}\right) \\
&= \frac{(T_{\mathsf{LA}}-1)T_{\mathsf{LA}}(T_{\mathsf{LA}}+1)}{6} - \frac{T_{\mathsf{LA}}(T_{\mathsf{LA}}-1)^2}{4} \\
&= \frac{(T_{\mathsf{LA}}-1)T_{\mathsf{LA}}(T_{\mathsf{LA}}+1)}{12}
\end{aligned}$$

$$= \frac{1}{2}\binom{T_{\mathsf{LA}}+1}{3}$$

which gives the first result. Next, for $\hat{\beta}_\ell^{\mathsf{LA}}$,

$$\begin{aligned}\hat{\beta}_\ell^{\mathsf{LA}} &= \frac{\sum_i (Y_{i\ell} - \bar{Y}_i^{\mathsf{LA}})(S_i - \bar{S})}{n\bar{S}(1-\bar{S})} - \hat{\gamma}\,(\ell - \bar{t}_{\mathsf{LA}}) \\ &= \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1-\bar{S})}\left(Y_{i\ell} - \bar{Y}_i^{\mathsf{LA}} - (\ell - \bar{t}_{\mathsf{LA}}) \sum_{t=t_m+1}^{t_M} \Delta Y_{it}\omega_t^\gamma\right)\end{aligned}$$

but

$$\bar{Y}_i^{\mathsf{LA}} = \frac{1}{T_{\mathsf{LA}}}\sum_{t=t_m}^{t_M} Y_{it} = \frac{1}{T_{\mathsf{LA}}}\sum_{t=t_m}^{t_M}\left(Y_{it_m} + \mathbb{1}(t > t_m)\sum_{s=t_m+1}^{t}\Delta Y_{is}\right) = Y_{it_m} + \sum_{s=t_m+1}^{t_M}\Delta Y_{is}\frac{(t_M+1-s)}{T_{\mathsf{LA}}}$$

so

$$\begin{aligned}\hat{\beta}_\ell^{\mathsf{LA}} &= \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1-\bar{S})}\left(Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M}\Delta Y_{it}\left(\frac{(t_M+1-t)}{T_{\mathsf{LA}}} + (\ell - \bar{t}_{\mathsf{LA}})\omega_t^\gamma\right)\right) \\ &= \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1-\bar{S})}\left(Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M}\Delta Y_{it}\omega_t^\ell\right)\end{aligned}$$

where

$$\omega_t^\ell = \frac{(t_M+1-t)}{T_{\mathsf{LA}}} + (\ell - \bar{t}_{\mathsf{LA}})\omega_t^\gamma$$

which is non-negative for $t_m \le t \le t_M$ and where

$$\begin{aligned}\sum_{t_{m+1}}^{t_M}\omega_t^\ell &= \sum_{t_{m+1}}^{t_M}\frac{(t_M+1-t)}{T_{\mathsf{LA}}} + \ell - \bar{t}_{\mathsf{LA}} = (t_M - t_m)\frac{(t_M+1)}{T_{\mathsf{LA}}} - \bar{t}_{\mathsf{LA}} + \frac{t_m}{T_{\mathsf{LA}}} + \ell - \bar{t}_{\mathsf{LA}} \\ &= \frac{T_{\mathsf{LA}}-1}{T_{\mathsf{LA}}}(t_M+1) - 2\bar{t}_{\mathsf{LA}} + \frac{t_m}{T_{\mathsf{LA}}} + \ell = t_M + 1 - \frac{t_M+1-t_m}{T_{\mathsf{LA}}} - t_M - t_m + \ell \\ &= \ell - t_m\end{aligned}$$

and the result follows by a standard application of the law or large numbers. Next, plugging in the potential outcomes,

$$\gamma = \sum_{t=t_m+1}^{t_M}\omega_t^\gamma\left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}})|S=1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0)|S=0]\right).$$

Similarly,

$$\beta_\ell^{\mathsf{LA}} = \mathbb{E}[Y_{i\ell} - Y_{it_m}|S_i = 1] - \mathbb{E}[Y_{i\ell} - Y_{it_m}|S_i = 0] - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[\Delta Y_{it}|S_i = 1] - \mathbb{E}[\Delta Y_{it}|S_i = 0]\right).$$

Then for $\ell < t^*, \ell \notin \mathcal{T}_{\mathsf{LA}}$,

$$\begin{aligned}\beta_\ell^{\mathsf{LA}} &= \mathbb{E}[Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) - Y_{t_m}(\mathbf{d}_{t_m:1}^{\mathsf{pre}})|S = 1] - \mathbb{E}[Y_\ell(\mathbf{d}_{\ell:1}^0) - Y_{t_m}(\mathbf{d}_{t_m:1}^0)|S = 0] \\ &\quad - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}})|S = 1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0)|S = 0]\right)\end{aligned}$$

and for $\ell \geq t^*$,

$$\begin{aligned}\beta_\ell^{\mathsf{LA}} &= \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:t^*}^{\mathsf{post}}, \mathbf{d}_{t^*-1:1}^{\mathsf{pre}}) - Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}})\middle| S = 1\right] \\ &\quad + \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) - Y_{t_m}(\mathbf{d}_{t_m:1}^{\mathsf{pre}})\middle| S = 1\right] - \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^0) - Y_{t_m}(\mathbf{d}_{t_m:1}^0)\middle| S = 0\right] \\ &\quad - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}})|S = 1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0)|S = 0]\right).\end{aligned}$$

Finally note that when the difference in trends is constant,

$$\begin{aligned}\beta_\ell^{\mathsf{LA}} &= \mathbb{E}[Y_{i\ell}(d_{\mathsf{post}}, d_{\mathsf{pre}}) - Y_{i\ell}(d_{\mathsf{pre}}, d_{\mathsf{pre}})|S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\ell}(d_{\mathsf{pre}}, d_{\mathsf{pre}}) - Y_{it_m}(d_{\mathsf{pre}})|S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{it_m}(d_0)|S_i = 0] \\ &\quad - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[\Delta Y_{it}(d_{\mathsf{pre}})|S_i = 1] - \mathbb{E}[\Delta Y_{it}(d_0)|S_i = 0]\right) \\ &= \mathbb{E}[Y_{i\ell}(d_{\mathsf{post}}, d_{\mathsf{pre}}) - Y_{i\ell}(d_{\mathsf{pre}}, d_{\mathsf{pre}})|S_i = 1] + \kappa(\ell - t_m) - \kappa \sum_{t=t_m+1}^{t_M} \omega_t^\ell \\ &= \mathbb{E}[Y_{i\ell}(d_{\mathsf{post}}, d_{\mathsf{pre}}) - Y_{i\ell}(d_{\mathsf{pre}}, d_{\mathsf{pre}})|S_i = 1]\end{aligned}$$

using that $\sum_{t=t_m+1}^{t_M} \omega_t^\ell = \ell - t_m$, which completes the proof. □

B.3 Proof of Corollary 1

Under Assumption 4, for $\ell \geq t^*$,

$$\begin{aligned}\beta_\ell^{\mathsf{LA}} &= \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:t^*}^{\mathsf{post}}, \mathbf{d}_{t^*-1:1}^{\mathsf{pre}}) - Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}})\middle| S = 1\right] \\ &\quad + \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^{\mathsf{pre}}) - Y_{t_m}(\mathbf{d}_{t_m:1}^{\mathsf{pre}})\middle| S = 1\right] - \mathbb{E}\left[Y_\ell(\mathbf{d}_{\ell:1}^0) - Y_{t_m}(\mathbf{d}_{t_m:1}^0)\middle| S = 0\right] \\ &\quad - \sum_{t=t_m+1}^{t_M} \omega_t^\ell \left(\mathbb{E}[Y_t(\mathbf{d}_{t:1}^{\mathsf{pre}}) - Y_{t-1}(\mathbf{d}_{t-1:1}^{\mathsf{pre}})|S = 1] - \mathbb{E}[Y_t(\mathbf{d}_{t:1}^0) - Y_{t-1}(\mathbf{d}_{t-1:1}^0)|S = 0]\right)\end{aligned}$$

$$= \mathbb{E}\left[Y_\ell(\mathbf{d}^{\mathsf{post}}_{\ell:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) - Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1})\middle| S=1\right] + \kappa(\ell - t_m) - \kappa \sum_{t=t_m+1}^{t_M} \omega_t^\ell$$

$$= \mathbb{E}\left[Y_\ell(\mathbf{d}^{\mathsf{post}}_{\ell:t^*}, \mathbf{d}^{\mathsf{pre}}_{t^*-1:1}) - Y_\ell(\mathbf{d}^{\mathsf{pre}}_{\ell:1})\middle| S=1\right]$$

and similarly for $\ell < t^*$, $\ell \notin \mathcal{T}_{\mathsf{LA}}$, $\beta_\ell^{\mathsf{LA}} = 0$. $\square$