

# Difference-in-Differences with Unequal Baseline Treatment Status\*

Alisa Tazhitdinova      Gonzalo Vazquez-Bare

March 20, 2023

## Abstract

We study a difference-in-differences (DiD) framework where groups experience unequal treatment statuses in the pre-policy change period. This approach is commonly employed in empirical studies but it contradicts the canonical model's assumptions. We show that in such settings, the standard DiD approach fails to recover the average treatment effect (ATT), unless the treatment effect is immediate and constant over time. Furthermore, the usual parallel trends test is invalid, meaning one may find pre-trends when the parallel trends assumption holds, and vice versa. We discuss two solutions. First, we show that including a linear term trend will recover the ATT if the differences in trends are constant over time (both in unequal baseline and canonical DiD settings) but not otherwise. Second, estimation in reverse also recovers the ATT if the potential outcomes do not depend on past treatments and post-policy statuses are converging.

**JEL Classification:** C21, C23

**Keywords:** differences-in-differences, event study, parallel trends assumption, panel data, policy evaluation

---

\*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. Xhulio Uruci provided excellent research assistance.

In the canonical difference-in-differences (DiD) framework, the effect of a treatment is estimated by comparing the outcome trend of a group that is treated at some point in time to the outcome trend of a comparison group that remains untreated throughout the period of study. In such settings, both groups start at the same baseline treatment status and then diverge. Many empirical studies, however, employ a modified version of this setup, instead comparing a group that experiences a change to a group that does not, while allowing groups to have different status in the baseline (i.e. pre-policy change) period. We will refer to these types of frameworks as “unequal baseline DiD.” The goal of this paper is to evaluate the validity of this unequal baseline DiD approach. We show that while this estimation approach does not recover the average treatment effect on the treated (ATT) in general, it results in consistent estimates when the treatment effect is immediate and constant over time. Furthermore, we show that in unequal baseline DiD settings, truly comparable groups, i.e. groups whose outcome trends would have been the same in the absence of the policy, may exhibit divergent observed trends, implying that the usual parallel trends test may lead to inappropriate conclusions.

The unequal baseline DiD approach is ubiquitously used in empirical studies. First, it is widely used in empirical tax research to measure the causal effects of taxes on income, wealth, investments, and other outcomes of interest. 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](#)). Most of such studies fail the canonical assumption of the DiD framework because taxpayers experience different MTRs in the period before the studied tax change.

Beyond tax studies, unequal baseline DiD frameworks have been employed in other settings: 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.

Despite the ubiquity of unequal baseline DiD, they have not been formally analyzed. In this paper, we provide the first formal evaluation of this approach. First, we

theoretically characterize the estimators and parameters that can be recovered using an unequal baseline DiD and discuss the conditions under which they have a causal interpretation. Our analysis allows potential outcomes to depend on past treatment status, and applies to settings where treatment statuses eventually converge as well as settings where they do not. We illustrate the implications of our theoretical results in empirical settings, and highlight the importance of institutional knowledge and economic theory for empirical analysis. Finally, we evaluate the validity of two solutions – estimation in reverse (i.e. designating a reference period to be some period after the policy change) and the inclusion of a linear trend (to correct for differences in pre-trends), and provide practical recommendations for empirical researchers.

Our theoretical framework considers a population with two groups. The first group, the stayers, starts and remains at the baseline treatment level  $d_0$ , while the second group, the switchers, starts at treatment level  $d_{\text{pre}}$  and then switches to treatment level  $d_{\text{post}}$  in period  $t^*$ . Our framework thus diverges from the canonical model (that requires  $d_{\text{pre}} = d_0$ ) in one dimension: whenever  $d_{\text{pre}} \neq d_0$ , switchers and stayers experience unequal baseline statuses. Our goal is to characterize estimators from the commonly used event-study specification that includes interaction terms of treatment status and time period indicators.<sup>1</sup> In such specifications, the lag coefficients are expected to estimate ATT effects over time, while the lead coefficients are expected to be approximately zero thus providing evidence in favor of the parallel trends assumption.

We first show that whenever  $d_{\text{pre}} \neq d_0$ , the ATT is identified under a parallel trends assumption that requires that the evolution of outcomes be similar for the two groups under *each group's own baseline treatment status*  $d_0$  or  $d_{\text{pre}}$ . However, we formally show that this assumption is highly restrictive, in the sense that it effectively rules out time-varying treatment effects. To see this, note that if the policy in question has lasting treatment effects, then individuals with different policy statuses cannot have similar outcome paths. Therefore, researchers cannot show evidence of parallel trends, yet allow for time-varying treatment effects. In other words, for the framework to have internal validity, non-significant and near-zero estimates of lead coefficients must be coupled with approximately constant estimates of lag coefficients. Significant and/or non-zero estimates of lead coefficients suggest that the above parallel trends assumption is unlikely to hold, while non-constant lag coefficients imply time-varying

---

<sup>1</sup>Formally,  $Y_{it} = \alpha_i + \delta_t + \sum_{\ell < t^*, \ell \neq t^*-1} \theta_\ell S_i \mathbb{1}(t = \ell) + \sum_{\ell \geq t^*} \beta_\ell S_i \mathbb{1}(t = \ell) + \varepsilon_{it}$ , where  $Y_i$  measures outcome of interest,  $S_i$  identifies whether individual  $i$  is a switcher,  $\alpha_i$  are individual fixed effects and  $\delta_t$  are time period fixed effects. We refer to  $\theta_\ell$  as lead coefficients and  $\beta_\ell$  as lag coefficients.

treatment effects, which again contradict the above parallel trends assumption.

Next, we consider two alternative parallel trends assumptions that do not result in restrictions on the nature of treatment effects: that the evolution of outcomes is similar for the two groups at the (1) *stayers' baseline treatment status*  $d_0$  or at the (2) *switchers' baseline treatment status*  $d_{\text{pre}}$ . We find that under either of these assumptions, the ATT is only identified when the treatment effect is immediate and constant over time. Importantly, the usual “parallel trends test” – i.e. verifying that lead coefficients are near-zero and non-significant – does not provide evidence in support of the chosen (either (1) or (2)) parallel trends assumption, unless the treatment effect is constant over time. We now describe the intuition for these results.

Consider a policy that was initially introduced in period  $t_{-1}$  for a subgroup of the population, after which it is cancelled in period  $t^*$ .<sup>2</sup> For example, [Jakobsen et al. \(2020\)](#) study how wealthy individuals respond to wealth taxes in Denmark. One of their approaches exploits a 1989 reform that increased the wealth tax exemption threshold for couples. As a result of the reform, some married individuals who were previously subject to wealth taxation became exempt from it. The authors estimate the 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 and whose tax liability did not change (their preferred specification) and (b) couples in a lower wealth range who were always exempt from the wealth tax. Both groups represent plausible controls: group (a) has similar wealth levels while group (b) has similar marital status. Therefore, each group is an appealing comparison that may satisfy parallel trends assumption under the switchers' baseline status (i.e. if the wealth tax applied to each group). However, group (a) satisfies the canonical DiD assumption as both groups face similar tax liability in the pre-reform period, while group (b) does not because, in the pre-reform period, group (b) was not subject to the wealth tax, while the treatment group was. Since wealth taxes lead to mechanical reductions of wealth, the control group (b) should experience faster wealth growth than the treatment group in the pre-reform period, resulting in non-parallel paths of wealth prior to the reform. This example illustrates why in settings with unequal baseline status and continuous treatment effects, one may estimate significant pre-trends even if the treatment and control groups are truly comparable. We show

---

<sup>2</sup>Such settings have been studied by [Bleemer \(2022\)](#), [Cao and Chen \(2022\)](#), [Jakobsen et al. \(2020\)](#), [Rossi and Villar \(2020\)](#), [Bleakley \(2010\)](#), [Cutler et al. \(2010\)](#), and [Lucas \(2010\)](#) among others. For a visual illustration, See Figure 1.

that in such settings, the lead coefficients estimate not only the difference in trends between the two groups if the policy never applied but also the effect of the policy over time. This also means that one may fail to detect differences in pre-trends if these differences cancel out with the average treatment effect, thus incorrectly concluding that groups are similar. [Jakobsen et al. \(2020\)](#) indeed find non-parallel trends when using control group (b), which could either be due to continuous treatment effect of wealth taxation or because group (b) is not a good comparison for the treatment group, with no econometric way of telling these apart (see Figure 2).

We show that in such settings, the lag coefficients estimate the sum of the ATT plus the average difference between the trend that would have been observed had the switchers remained at their initial treatment level  $d_{\text{pre}}$ , and the trend that is observed for the stayers at treatment level  $d_0$ . When the policy leads to non-constant treatment effects, the second difference is typically not zero, and the ATT estimates are inconsistent.

Our results demonstrate that when the policy treatment effect is constant over time and immediate, unequal baseline DiD approach recovers the ATT, and the usual pre-trends test provides suggestive evidence in support of the chosen parallel trends assumption. However, unequal baseline frameworks will lead to inconsistent ATT estimates and confusing pre-trends whenever the treatment effects vary over time. Therefore, without explicit knowledge of the nature of the treatment effects – either from data on earlier periods, from previous empirical work in related settings, or from theoretical models – one cannot interpret the observed pre-trends and establish consistency of the estimates.

Thus, in cases where constant treatment effects are unlikely, alternative solutions are needed. We consider the validity of two intuitive solutions: (1) estimation in reverse and (2) the inclusion of a linear trend. First, in settings where the treatment status is different in the baseline period but converges after the policy change, an appealing solution is to estimate event studies in reverse, by designating a reference period to be some period after the policy change. We show that this estimation approach will only recover the ATT if the trends are parallel under the stayers’ treatment status *and* if the potential outcomes do not depend on the previous status, i.e. outcomes are static. Therefore, the reverse estimation strategy is hard to justify, unless researchers have prior knowledge on the nature of potential outcome dynamics.

Second, we consider a solution frequently used to account for differential trends

in the baseline period – the inclusion of a linear trend term. This approach has been employed in unequal baseline DiD frameworks, but has also been used in canonical DiD settings where trends appear to be non-parallel in a plausibly systematic way. We provide the first formal characterization of this popular estimator and its corresponding estimand, and show that this approach indeed can work well if differences in baseline period trends are constant over time (or zero). In such settings, the linear trend term serves both as a “test” of the parallel trends assumption, and as a correction, thus allowing to recover the ATT. We demonstrate how the estimates change with the choice of the linear trend estimation window, thus providing practical guidance on its implementation.

However, we also show that inclusion of a linear trend is not a cure-all: if differences in trends are not constant over time, then estimates are inconsistent, and it is generally not possible to determine which asymptotic bias is larger – when using a simple event study design or one with a linear trend. We show that in some cases, simple event studies perform better than event studies with a linear trend, and sometimes vice versa. Therefore, the choice of specification should be guided by economic theory, prior empirical evidence, or direct observation of long pre-reform trends, if the data series are sufficiently long.

Our paper contributes to a growing literature analyzing identification and estimation of treatment effects using DiD methods under different scenarios (de Chaisemartin and D’Haultfoeulle, 2020; Sant’Anna and Zhao, 2020; Callaway and Sant’Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2021; Roth and Sant’Anna, 2021; de Chaisemartin et al., 2022). These papers consider the canonical case in which all units start from the same treatment level and then diverge in later period(s) (see Steigerwald et al., 2021; de Chaisemartin and D’Haultfoeulle, 2022; Roth et al., 2022, for recent surveys). In contrast, our paper analyzes settings where units start at *different baseline* levels and then the treatment changes discretely and by the same magnitude for all switchers. We allow for continuous treatment and characterize estimators in a two-way fixed effects specification.<sup>3</sup> Our framework is most closely related to de Chaisemartin and D’Haultfoeulle (2018) and Kim and Lee (2019), who consider a DiD setting where the stayers remain treated while the switchers move from

---

<sup>3</sup>In recent work, de Chaisemartin et al. (2022) and Callaway et al. (2021) consider DiD with continuous treatments where units can start from different treatment levels. However, these papers focus on other estimands, such as average derivatives or response functions that require more variation in the treatment, different estimation methods and different parallel trends assumptions.

untreated to treated, a scenario that we refer to as “universal adoption” (see Section 2.2). Within a fuzzy DiD setup, [de Chaisemartin and D’Haultfoeuille \(2018\)](#) provide identification conditions for causal effects under a parallel trends assumption that is analogous to our Assumption 5 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 implications) of the parallel trends assumption. We also allow potential outcomes to depend on past treatment status, and our results show that the strategy in [Kim and Lee \(2019\)](#) requires the (often restrictive) assumption that potential outcomes only depend on current treatment status (see Section 3.2). Finally, we consider multiple periods and different estimation strategies including event study designs, which are one of the most commonly used empirical tools for estimation in DiD, and a linear trend adjustment, and characterize the behavior of the lead coefficient estimators typically used to test for pre-trends in outcomes.

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](#)). To the best of our knowledge, while other studies have analyzed similar forms of adjustments ([Bilinski and Hatfield, 2018](#); [Borusyak et al., 2021](#)), to our knowledge we are the first to derive a closed-form characterization of the OLS estimators and estimands from an event-study specification with a linear adjustment for the differences in trends. We formally show that while this approach leads to consistent estimates of causal effects when differences in trends are constant (including zero), the inclusion of a linear trend could result in more biased estimates when the differences in trends are non-constant.

# 1 Setup and Main Results

We consider balanced panel data with units  $i = 1, \dots, n$  and time periods  $t = 1, \dots, T$ . There is a treatment of interest  $D_{it}$ , not necessarily binary,<sup>4</sup> taking values in a set  $\mathcal{D}$ . The population consists of two groups. The first group, the stayers, starts at treatment level  $d_0$ , which we refer to as the “baseline treatment status”, and remains at that treatment level throughout the observed period. The second group in the population, the switchers, starts at a treatment level  $d_{\text{pre}}$  and switches to treatment level  $d_{\text{post}} \neq d_{\text{pre}}$  in period  $t^*$ , where  $1 < t^* \leq T$ , and stays in treatment status  $d_{\text{post}}$  for the remaining periods. We refer to  $t^*$  as the policy change period. We define the variable  $S_i$  which is equal to one if unit  $i$  switches treatment status in period  $t^*$  (the switchers) and zero if unit  $i$  remains at treatment level  $d_0$  for the whole period (the stayers). The observed outcome of interest of unit  $i$  in period  $t$  is denoted by  $Y_{it}$ . We will assume our observed data follows the following sampling scheme, which is standard in panel data settings.

## Assumption 1 (Sampling and moments)

1. *Observations  $(Y_{i1}, Y_{i2}, \dots, Y_{iT}, D_{i1}, D_{i2}, \dots, D_{iT})_{i=1}^n$  are iid draws from an infinite superpopulation of units.*
2.  $0 < \mathbb{P}[S_i = 1] < 1$  and for all  $d \in \mathcal{D}$  and  $t = 1, \dots, T$ ,  $\mathbb{E}[Y_{it}(d)^2] < \infty$ .

To define causal parameters, we now introduce the potential outcomes which we view as random. In general, the potential outcome for unit  $i$  in period  $t$  can depend on the path of treatments from period 1 up to period  $t$ ,  $(d_1, d_2, \dots, d_t)$ . Because we consider a simultaneous adoption design in which stayers remain at the same treatment level and all the switchers change status in the same time period, we can represent potential outcomes as follows. For any period  $t < t^*$ , we write the potential outcome as  $Y_{it}(d_{\text{pre}})$ , which is shorthand notation for a path of treatments where treatment status remains at  $d_{\text{pre}}$ ,  $(d_1, d_2, \dots, d_t) = (d_{\text{pre}}, d_{\text{pre}}, \dots, d_{\text{pre}})$ . For any period  $t \geq t^*$ , we write the potential outcome as  $Y_{it}(d_{\text{post}}, d_{\text{pre}})$ , which is shorthand notation for a path of treatment that starts at  $d_{\text{pre}}$  over the pre-policy change period and then switches to  $d_{\text{post}}$  in period

---

<sup>4</sup>While our results allow for a multivalued or even continuous treatment (e.g. marginal tax rates), we only consider cases in which the treatment switches discretely from one value to another one. We do not consider identification and estimation of average derivatives or other more general estimands. See [de Chaisemartin et al. \(2022\)](#) and [Callaway et al. \(2021\)](#) for studies focusing specifically on DiD with continuous treatments.



$t^*$ , that is,  $(d_1, d_2, \dots, d_{t^*-1}, d_{t^*}, d_{t^*+1}, \dots, d_t) = (d_{\text{pre}}, d_{\text{pre}}, \dots, d_{\text{pre}}, d_{\text{post}}, d_{\text{post}}, \dots, d_{\text{post}})$ . In our notation, the first argument indicates current treatment status and the second argument indicates past treatment status.

In this setup, the potential outcomes for a stayer are  $Y_{it}(d_0)$  for  $t < t^*$  and  $Y_{it}(d_0, d_0)$  for  $t \geq t^*$ . On the other hand, the potential outcomes for a switcher are  $Y_{it}(d_{\text{pre}})$  for  $t < t^*$  and  $Y_{it}(d_{\text{post}}, d_{\text{pre}})$  for  $t \geq t^*$ . The individual-specific treatment effect of treatment status  $d$  over treatment status  $d'$  in the pre-policy change period is  $Y_{it}(d) - Y_{it}(d')$ , and the treatment effect of treatment status  $d$  over treatment status  $d'$  in the post-policy change period given pre-policy change treatment status  $d_{\text{pre}}$  is  $Y_{it}(d, d_{\text{pre}}) - Y_{it}(d', d_{\text{pre}})$ . Notice that this allows for effect heterogeneity both across units over time. That is, we allow, for example, for  $Y_{it}(d, d_{\text{pre}}) - Y_{it}(d', d_{\text{pre}}) \neq Y_{it+1}(d, d_{\text{pre}}) - Y_{it+1}(d', d_{\text{pre}})$  and  $Y_{it}(d, d_{\text{pre}}) - Y_{it}(d', d_{\text{pre}}) \neq Y_{jt}(d, d_{\text{pre}}) - Y_{jt}(d', d_{\text{pre}})$ .

Researchers relying on DiD strategies commonly assume, often implicitly, that potential outcomes do not depend on previous treatments. We refer to this assumption as the static potential outcomes assumption.

**Assumption 2 (Static potential outcomes)** *For all  $i$ ,  $t \geq t^*$  and  $d_{\text{post}}, d_{\text{pre}} \in \mathcal{D}$ ,  $Y_{it}(d_{\text{post}}, d_{\text{pre}}) = Y_{it}(d_{\text{post}})$ .*

This assumption states that potential outcomes in the post-policy periods do not depend on their second argument, which is pre-policy change treatment status. This assumption is strong, as it implies, for example, that if a unit is subject to the policy in periods  $t < t^*$  and the policy is removed in period  $t^*$ , the outcomes would switch immediately to the value that would have been observed had the policy never been in place (see Figure 1 for examples). Our results in upcoming sections do not rely on this assumption, but we discuss how identified parameters change if one is willing to impose it.

Empirical researchers commonly estimate the effect of the policy change by comparing the evolution of outcomes between switchers and stayers before and after the policy change. We consider the following event study specification:

$$Y_{it} = \alpha_i + \delta_t + \sum_{\ell < t^*, \ell \neq \tilde{t}} \theta_\ell S_i \mathbb{1}(t = \ell) + \sum_{\ell \geq t^*} \beta_\ell S_i \mathbb{1}(t = \ell) + \varepsilon_{it}, \quad (1)$$

We set period  $\tilde{t} < t^*$  as the baseline (omitted) period. A common choice for  $\tilde{t}$  in empirical studies is  $\tilde{t} = t^* - 1$ , the last pre-policy change period. We refer to the

estimated coefficients  $\hat{\beta}_\ell$  as the “lag coefficients”, and the estimated coefficients  $\hat{\theta}_\ell$  as the “lead coefficients”. We view the regression coefficients in Equation (1) as linear projection coefficients, and we do not make any assumptions on the shape of the conditional expectation of the observed outcome.

The following results characterizes the coefficients and OLS estimators from the event study design in Equation (1). Our results are based on an asymptotic panel data setting where  $n \rightarrow \infty$  with  $T$  fixed and we let “ $\rightarrow_{\mathbb{P}}$ ” represent convergence in probability.

**Lemma 1** *Let  $\{\beta_\ell\}_{\{\ell \geq t^*\}}$  and  $\{\theta_\ell\}_{\{\ell \neq t^*, \ell < t^*\}}$  be the linear projection coefficients from Equation (1) and  $\{\hat{\beta}_\ell\}_{\{\ell \geq t^*\}}$  and  $\{\hat{\theta}_\ell\}_{\{\ell \neq t^*, \ell < t^*\}}$  be the corresponding OLS estimators. Under Assumption 1, as  $n \rightarrow \infty$  with  $T$  fixed, for any  $\ell \geq t^*$ ,*

$$\begin{aligned}\beta_\ell &= \mathbb{E}[Y_{it} - Y_{i\tilde{t}} | S_i = 1] - \mathbb{E}[Y_{it} - Y_{i\tilde{t}} | S_i = 0], \\ \hat{\beta}_\ell &= \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}}) S_i}{\sum_{i=1}^n S_i} - \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}}) (1 - S_i)}{\sum_{i=1}^n (1 - S_i)} \rightarrow_{\mathbb{P}} \beta_\ell\end{aligned}$$

and for any  $\ell < t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\begin{aligned}\theta_\ell &= \mathbb{E}[Y_{it} - Y_{i\tilde{t}} | S_i = 1] - \mathbb{E}[Y_{it} - Y_{i\tilde{t}} | S_i = 0], \\ \hat{\theta}_\ell &= \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}}) S_i}{\sum_{i=1}^n S_i} - \frac{\sum_{i=1}^n (Y_{i\ell} - Y_{i\tilde{t}}) (1 - S_i)}{\sum_{i=1}^n (1 - S_i)} \rightarrow_{\mathbb{P}} \theta_\ell.\end{aligned}$$

Lemma 1 shows that the coefficient estimators from the event-study specification in Equation (1) are DiD estimators between periods  $\ell$  and the baseline period  $\tilde{t}$  and converge in probability to their population counterparts. Our next result links the event study coefficients to potential outcomes.

**Proposition 1** *Let  $\{\beta_\ell\}_{\{\ell \geq t^*\}}$  and  $\{\theta_\ell\}_{\{\ell \neq t^*, \ell < t^*\}}$  be the linear projection coefficients from Equation (1). For  $\ell \geq t^*$ ,*

$$\begin{aligned}\beta_\ell &= \mathbb{E}[Y_{i\ell}(d_{\text{post}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0]\end{aligned}$$

and for  $\ell < t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\theta_\ell = \mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0].$$

If, in addition, Assumption 2 holds,

$$\begin{aligned}\beta_\ell &= \mathbb{E}[Y_{i\ell}(d_{\text{post}}) - Y_{i\ell}(d_{\text{pre}}) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0].\end{aligned}$$

Proposition 1 shows that the lead coefficients  $\theta_\ell$  recover the pre-policy change difference in trends had the switchers remained at treatment status  $d_{\text{pre}}$  and stayers remained at treatment level  $d_0$ . The lag coefficients  $\beta_\ell$  equal the ATT in period  $\ell$ , plus the difference in trends that would be observed between switchers and stayers, again keeping their corresponding baseline statuses fixed.

## 1.1 Parallel Trends Assumptions and Pre-Trends Tests

Proposition 1 shows that the lag coefficients converge to the sum of the ATT plus the difference in trends between switchers and stayers. A parallel trends assumption is typically invoked to “erase” differences in trends and thus recover a consistent estimate of the ATT. A traditional DiD parallel trends assumption states that the evolution of outcomes is expected to be similar for the two groups at the *stayers’ treatment status*  $d_0$ , that is:

**Assumption 3 (Parallel trends - PT( $d_0$ ))** For any  $\ell \geq t^*$  and  $\tilde{t} < t^*$ ,

$$\mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0) | S_i = 1] = \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0].$$

In canonical DiD settings, the baseline status of treated and comparison group is the same, and therefore the parallel trends Assumption 3 is natural. However, in settings with unequal baseline status, the parallel trends assumption is less obvious. One may assume that outcomes would exhibit the same trajectory across groups at the pre-policy status of the *stayers* (as in Assumption 3) or at the pre-policy status of the *switchers*:

**Assumption 4 (Parallel trends - PT( $d_{\text{pre}}$ ))** For any  $\ell \geq t^*$  and  $\tilde{t} < t^*$ ,

$$\mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] = \mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 0].$$

An alternative assumption and one that appears most intuitive in unequal baseline settings is to require that outcomes would exhibit the same trajectory at each group’s pre-policy status, that is:

**Assumption 5 (Parallel trends - PT( $d_0, d_{\text{pre}}$ ))** For any  $\ell \geq t^*$  and  $\tilde{t} < t^*$ ,

$$\mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}})|S_i = 1] = \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0)|S_i = 0].$$

Which assumption is appropriate depends on the studied setting, and economic reasoning. For example, if researchers study the roll out of some new policy, then requiring parallel trends in absence of the policy may seem more reasonable than in presence of the policy. The reverse appears more appropriate if one studies changes to a policy that has been in place for a long time.

It can be readily seen, that besides being intuitive, Assumption 5 also ensures that the lag coefficients  $\hat{\beta}_\ell$  of Proposition 1 recover an ATT. However, while Assumption 5 is appealing at first sight, it is actually highly restrictive. To see this, note that we can rewrite Assumption 5 as follows:

$$\begin{aligned} & \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0)|S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0)|S_i = 0] \\ & + \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}})|S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}})|S_i = 1] = 0. \end{aligned}$$

In other words, Assumption 5 holds if Assumption 3 holds *and* the treatment effect of switching from status  $d_{\text{pre}}$  to status  $d_0$  for the switchers is constant over time.<sup>5</sup> Alternatively, Assumption 5 holds if Assumption 4 holds *and* the effect of treatment status  $d_0$  compared to  $d_{\text{pre}}$  for the stayers is constant over time:

$$\begin{aligned} & \mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}})|S_i = 1] - \mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}})|S_i = 0] \\ & + \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}})|S_i = 0] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}})|S_i = 0] = 0. \end{aligned}$$

Therefore, researchers cannot simultaneously presume that Assumption 5 holds and expect to recover (possibly) nonconstant treatment effects of the policy in question. For this reason, we consider Assumption 5 as stronger requirement that adds a treatment effect homogeneity restriction on top of Assumptions 3 or 4. This combination of parallel trends and constant treatment effects assumptions is also considered in [de Chaisemartin and D'Haultfoeulle \(2018\)](#) for identification in a fuzzy DiD setup (see their assumptions 4 and 5).

Assessing the validity of parallel trends assumptions based on pre-trends differ-

---

<sup>5</sup>Alternatively, the treatment effect could be such that it offsets the difference in trends between stayers and switchers precisely – intuitively, a knife-edge scenario that is hard to defend in practice.

ences, both visually – by plotting the evolution of outcomes for switchers and stayers, or formally with t- or F-tests for lead coefficients, has become standard practice in empirical work. It is well known, however, 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. Furthermore, recent literature has 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](#)).

These issues are particularly important in an unequal baseline DiD, where different pre-policy statuses may lead to different dynamics for switchers and stayers, making interpretation of the pre-trends test difficult. For this reason, it is important to distinguish between testing for pre-trends, which is based on observable data, and assessing the validity of the parallel trends assumption, which involves counterfactual unobserved magnitudes.

Consider the following example. Suppose the data contains four time periods,  $t = 1, 2, 3, 4$  and the policy change occurs in period  $t^* = 3$ . Suppose the stayers and switchers are truly comparable in that their outcomes would have exhibited the same trend in the absence of the treatment, but stayers are always treated whereas the switchers only receive treatment in period  $t = 3$ . Finally, suppose the treatment leads to doubling of the outcome in each subsequent period, but may start with a delay.<sup>6</sup> If the treatment effect is not delayed, then the event study lead coefficients will show significant differences in trends even though the groups are comparable (because the stayers’ outputs are doubling while switchers’ outputs are not). On the other hand, if the treatment effect is delayed and does not appear until the unit is treated for three or more years, then the lead coefficients will show no significant differences between switchers and stayers in periods 1 and 2. Yet, the lag coefficients will show a non-zero effect even though the true treatment effect is zero (because the stayers start to experience the effect of the treatment while for switchers the effect does not kick in until the third year of treatment). This example demonstrates why the validity of the parallel trends assumption cannot be assessed solely on statistical grounds, and why empirical evidence needs to be complemented with institutional knowledge and economic theory.

Proposition 1 applies to the most general case where groups start at different

---

<sup>6</sup>These patterns are common, for example, in agricultural or environmental programs (see e.g. [Maffioli et al., 2011](#); [Barreca et al., 2021](#)).

treatment statuses in the pre-period and never converge, i.e.  $d_{\text{pre}} \neq d_0 \neq d_{\text{post}}$ . We now apply Proposition 1 to two special cases. First, to settings with equal baseline status, i.e. where  $d_{\text{pre}} = d_0$ . And then to settings with converging post-status, i.e. where  $d_{\text{post}} = d_0$ .

## 1.2 Canonical DiD: Equal Baseline Status

In the canonical DiD setting, both groups start at the same treatment level and then diverge, i.e.  $d_{\text{pre}} = d_0$  and  $d_{\text{post}} \neq d_0$ . In such settings, it is common to assume that, had all units remained at the pre-policy change treatment status  $d_0$ , the average outcomes would have followed the same evolution over time, i.e. Assumption 3 holds. The next result follows directly from Proposition 1.

**Corollary 1 (Equal baseline treatment status)** *When  $d_{\text{pre}} = d_0$  and  $d_{\text{post}} \neq d_0$ , under Assumption 3, for  $\ell \geq t^*$*

$$\beta_\ell = \mathbb{E} [Y_{i\ell}(d_{\text{post}}, d_0) - Y_{i\ell}(d_0, d_0) | S_i = 1],$$

and for  $\ell < t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\theta_\ell = \mathbb{E} [Y_{i\ell}(d_0) - Y_{i\tilde{t}}(d_0) | S_i = 1] - \mathbb{E} [Y_{i\ell}(d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0].$$

Corollary 1 shows that if the parallel trends Assumption 3 is satisfied, then the lag coefficients equal the average effect on the switchers of moving from treatment status  $d_0$  to  $d_{\text{post}}$  in each post-policy change period  $\ell \geq t^*$ . At the same time, the lead coefficients  $\hat{\theta}_\ell$  equal the difference in pre-policy change trends between groups for the same treatment level. Notice that in general, the parallel trends Assumption 3 does not necessitate that the lead coefficients  $\theta_\ell$  equal zero. However, because lag coefficients mimic the parallel trends equation (which cannot be observed directly), researchers typically provide evidence in support of Assumption 3 by showing that  $\hat{\theta}_\ell$  are close to zero and statistically insignificant.

## 1.3 DiD with Converging Treatment Status

Now consider DiD settings in which groups start from different treatment levels and then converge to the same treatment. In other words, suppose  $d_{\text{pre}} \neq d_0$  and  $d_{\text{post}} = d_0$ . Applying Assumption 3 to Proposition 1 generates the following result.

**Corollary 2 (Convergent treatment status under  $\text{PT}(d_0)$ )** When  $d_{\text{pre}} \neq d_0$  and  $d_{\text{post}} = d_0$ , under Assumption 3, for  $\ell \geq t^*$ ,

$$\beta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] + \mathbb{E}[Y_{i\ell}(d_0, d_{\text{pre}}) - Y_{i\ell}(d_0, d_0) | S_i = 1],$$

and for  $\ell < t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\theta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1].$$

If, in addition, Assumption 2 holds,

$$\beta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1].$$

Corollary 2 demonstrates that the lead coefficients  $\theta_\ell$  equal the difference in treatment effect on the switchers from switching from treatment status  $d_{\text{pre}}$  to  $d_0$  in period  $\ell$  versus in period  $\tilde{t}$ . Therefore, even if Assumption 3 holds, lead coefficients will not converge to zero unless the treatment effect on the switchers is constant over time. At the same time, the lag coefficients do not recover an ATT unless such treatment effect is constant over time.<sup>7</sup> Under the additional assumption of static potential outcomes, which can be seen as a best-case scenario, all the lag coefficients  $\{\hat{\beta}_\ell\}_\ell$  estimate the same parameter, namely, the average effect of status  $d_0$  versus  $d_{\text{pre}}$  at the baseline period  $\tilde{t}$ , and are not able to recover the dynamic effect of policy change.

Applying Assumption 4 to Proposition 1 generates the following result:

**Corollary 3 (Converging treatment status under  $\text{PT}(d_{\text{pre}})$ )** When  $d_{\text{pre}} \neq d_0$  and  $d_{\text{post}} = d_0$ , under Assumption 4, for  $\ell \geq t^*$ ,

$$\begin{aligned} \beta_\ell = & \mathbb{E}[Y_{i\ell}(d_0, d_{\text{pre}}) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] \\ & + \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 0] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 0], \end{aligned}$$

and for  $\ell < t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\theta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 0] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 0].$$

---

<sup>7</sup>To see this note that Corollary 2 implies that  $\hat{\beta}_\ell \rightarrow_{\mathbb{P}} \mathbb{E}[Y_{i\ell}(d_0, d_{\text{pre}}) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] + \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1]$ . Here the first term measures an ATT while the second two terms measure the difference in treatment effects on the switchers, comparing status  $d_{\text{pre}}$  to status  $d_0$  between periods  $\ell$  and  $\tilde{t}$ .

If, in addition, Assumption 2 holds,

$$\begin{aligned} \beta_\ell = & \mathbb{E}[Y_{i\ell}(d_0) - Y_{i\ell}(d_{\text{pre}}) | S_i = 1] \\ & - \{\mathbb{E}[Y_{i\ell}(d_0) - Y_{i\ell}(d_{\text{pre}}) | S_i = 0] - \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 0]\}. \end{aligned}$$

Corollary 3 demonstrates that the lead coefficients equal the difference in treatment effect of switching from status  $d_{\text{pre}}$  to status  $d_0$  on the stayers in period  $\ell$  versus in period  $\tilde{t}$ . Therefore, even if Assumption 4 holds, lead coefficients will not equal zero unless the treatment effect on the stayers is constant over time. At the same time, the lag coefficients do not recover the ATT unless such treatment effect is constant over time.

## 2 Empirical Applications

In this section we apply our results to two common empirical settings: policy reversals and universal policy adoptions. Finally, we discuss settings with non-convergent treatment status, where individuals experience different treatment status both in pre- and post periods.

### 2.1 Policy Reversal

In policy reversal studies, in the baseline period, the policy applies to the treatment group but not to the comparison group, before being cancelled in the post period.<sup>8</sup> This type of research design is common. For example, [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

---

<sup>8</sup>Settings where the policy is first applied to all individuals but then is cancelled for a subgroup of people fall under the canonical DiD framework, summarized in Section 1.2, since in the baseline period both treatment and control group are subject to the same status.



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.

In the policy reversal settings,  $d_{\text{pre}}$  identifies episodes when the policy applies and  $d_0$  when the policy does not apply. In this setting, it seems natural to assume that the parallel trends assumption holds at the comparison group’s treatment status  $d_0$  (i.e. Assumption 3), since all groups eventually revert to this status. Figure 1 illustrates the potential empirical challenges under a variety of assumptions, for a policy that is initially introduced in period  $t_{-1}$  and then cancelled in period  $t^*$ . Panel A considers truly comparable treatment and control groups, i.e. groups that show parallel trends prior to  $t_{-1}$ , while Panel B assumes groups are not comparable.

Figure 1 illustrates why the parallel trend test over the baseline period may fail: in the baseline period, i.e. after  $t_{-1}$  but before  $t^*$ , one may find pre-trends even if the groups are truly comparable – e.g. in (b) and (c), or may fail to detect differences in pre-trends because these differences cancel out with the policy treatment effect – e.g. in (e) and approximately (f). Figure 1 also illustrates why the lag coefficients  $\hat{\beta}_l$  do not identify the ATT when the policy leads to non-constant treatment effects as in (b)-(c) and (e)-(f). Note that in the hypothetical scenarios of Figure 1, the potential outcomes do not depend on previous treatments, i.e. Assumption 2 holds. In such cases Corollary 2 shows that lag coefficients converge to  $\mathbb{E}[Y_{it}(d_0) - Y_{it}(d_{\text{pre}}) | S_i = 1]$ . In other words, all lag coefficients estimate the same parameter – the effect of removing the policy in period  $\tilde{t}$ . Hence, even under a stronger assumption, the lag coefficients do not allow the researcher to estimate how the effect of interest varies over time and thus recover dynamic ATT.

[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 using two approaches. 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.<sup>9</sup>

Figure 2 reproduces Jakobsen et al. (2020) findings. Note that comparing treated couples to untreated singles in Figure 2(a) follows the canonical DiD framework: both groups experience the same treatment status (a positive wealth tax) in the baseline period. Therefore, Corollary 1 applies and the DiD approach recovers the ATT as long as the parallel trends assumption holds. However, comparing treated couples to couples who were not subject to wealth tax in the first period in Figure 2(b) violates the canonical assumptions because treatment and control experience unequal status in the first period, necessitating the application of Corollary 2 (or Corollary 3). Thus this approach identifies the ATT on the switchers if and only if 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 presence of higher (lower) wealth tax rates. Indeed, as Figure 2 shows, the first control 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 2(b) could either be due to continuous treatment effect of wealth taxation (similar to Figures 1(b)-(c)), or because groups are actually not comparable (similar to Figure 1(d)), 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 evaluate the validity of this approach in Section 3.3.

## 2.2 Universal Adoption

Corollaries 2 and 3 can also be applied to studies that explore universal adoptions of policies that were previously applied only to a subset of the population. In these settings, the comparison group consists of individuals that are subject to some policy in both periods, while the treatment group adopts the policy in the post period only.

---

<sup>9</sup>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. Authors address these separately.

This setting has been explored in environmental economics to study energy effects of daylight savings 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 implement 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 a 39 hours per week to 35 in all areas of France except for Alsace-Moselle where working hours decreased from a lower starting point.

In settings with universal policy adoptions,  $d_{\text{pre}}$  represents episodes when the policy does not apply, while status  $d_0$  when the policy does apply. In these settings, Assumption 3 implies that the trends would be parallel if the policy applied to both groups. One can interpret Assumption 3 as restricting the evolution of the treatment effect over time, since it requires the group of switchers and stayers to evolve in the same way under the policy. In fact, a sufficient condition for this version of the parallel trends assumption is that (i) the outcomes of both groups in the absence of treatment evolves in the same way over time and (ii) the average treatment effects for both groups exhibit the same trend. In this sense, Assumption 4 is more in line with the usual parallel trends assumption employed in the DiD literature. Nonetheless, Assumption 3 may be an intuitive choice in some settings, e.g. in the case of [Chemin and Wasmer \(2009\)](#), while Assumption 4 may be better suited for other settings, e.g. [Kotchen and Grant \(2011\)](#) and [Sawada et al. \(2022\)](#).

Figure 3 illustrates the issues for a policy that is adopted by the stayers in period  $t_{-1}$  and by the switchers in period  $t^*$ . Panel A considers truly comparable treatment and control groups under Assumption 4, i.e. groups that show parallel trends prior to  $t_{-1}$ , while Panel B assumes groups are not comparable. As Figures 3 (b) and (c) demonstrate, even if groups are truly comparable, the lead coefficients will not be zero unless the policy effect on the stayers/switchers is constant over time. Alternatively, the change in treatment for the stayers/switchers may offset differences in trends, resulting in incorrect conclusion that groups are indeed similar, as in 3(e).

Under Assumption 4, Corollary 3 shows that the lag coefficients converge to the

sum of the average effect on the switchers of adopting the policy after being untreated, and the difference in treatment effect of switching from status  $d_{\text{pre}}$  to status  $d_0$  on the stayers in period  $\ell$  versus in period  $\tilde{\ell}$ . This latter term is non-zero when treatment effects are non-constant, as in Figures 3 (b) and (c), irrespective of whether Assumption 2 holds in addition or not.

If Assumption 3 is more appropriate in the studied setting, then the results of Section 2.1 apply: the lag coefficients are biased unless the policy results in constant treatment effects, while the lead coefficients do not provide suggestive evidence for or against Assumption 3.

## 2.3 DiD with Non-converging Treatment Status

Finally, Proposition 1 applies to the most general case, where treatment and control groups experience differential treatment statuses in both periods. This version of DiD is ubiquitously used by empirical tax economists to measure causal effects of taxes. For example, Kleven and Schultz (2014) measure causal effects of income taxes while Jakobsen et al. (2020) measure 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 bonus depreciation scheme on investment by exploiting differences in 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 in the post-period.

Proposition 1 applies to these settings. As discussed in Section 1, the lead coefficients estimate difference in trends between switchers and stayers keeping their corresponding different pre-policy change status fixed, and therefore not providing an evaluation of Assumption 3 nor Assumption 4. The lag coefficients estimate the ATT plus the difference in trends that would be observed between switchers and stayers, had the switchers remained at treatment status  $d_{\text{pre}}$ . Again, neither Assumption 3 nor Assumption 4 alone can guarantee that this latter term is zero. In addition to one of the parallel trend assumptions, one must assume that the treatment effect is constant over time in order to recover the unbiased ATT.

Jakobsen et al. (2020) provide an empirical illustration of this case. Their second

set of analysis exploits a 1989 reform that reduced the wealth tax rate from 2.2% to 1% on the very wealthy. To estimate the causal effect of this change, the authors compare wealthy individuals to a group subject to zero marginal tax rate because of a tax ceiling provision. This approach does not conform to the classical DiD setup because the treatment and control group experience differential wealth tax treatment in the pre-period, and in contrast to the analysis discussed in Section 2.1, groups' statuses do not converge in a post-policy period. Again, unless we expect treatment effect of wealth taxes to be constant over time, this approach would not recover the ATT and the parallel trend test would fail. Figure 4 confirms our expectations: again, the raw wealth series show differential trends which authors adjust for with a linear trend. We discuss the validity of such linear trend adjustment in Section 3.3.

### 3 Practical Recommendations

The inability of researchers to verify the validity of parallel trends assumption and thus establish to what extent treatment and control groups are actually comparable is a well-known shortcoming of the DiD approach is. A growing literature has pointed out the limitations of pre-trends tests in canonical DiD settings (e.g., Roth, 2022; Rambachan and Roth, 2023; Kahn-Lang and Lang, 2019; Bilinski and Hatfield, 2018; Freyaldenhoven et al., 2019), but it is particularly troublesome in settings with unequal baseline DiD.

Sections 1 and 2 demonstrate that the ability of DiD approach to recover the true ATT in such settings depends on the nature of treatment effects. The unequal baseline DiD is best applied to settings where treatment effects are expected to be fairly immediate and constant in nature. Having data that covers 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 treatment effects, 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 marginal income tax is expected to permanently decrease individual's labor supply. With the exception of a plausibly short adjustment period (e.g. due to information and labor market frictions), the treatment effect should be constant in nature. In

these cases, the approach will correctly estimate the ATT as long as sufficient time has passed from the initial treatment to allow for the treatment effect to apply.

On the other hand, in other settings we would expect non-constant treatment effects. For example, a wealth tax should plausibly lead to a continuous treatment effect through tax’s mechanical effect on wealth accumulation. Therefore, even in absence of individuals’ responses, one would expect a divergence of wealth outcomes among individuals subject to different levels of wealth tax. Generally speaking, 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](#)).

In what follows, we discuss a few alternatives for analyzing unequal baseline DiD.

### 3.1 Additional Data on Pre-Baseline Periods

The most straightforward, although often infeasible, solution to the limitations of unequal baseline DiD is to obtain more data on earlier periods where both groups had a common treatment status, as in the periods before  $t_{-1}$  in Figure 1. Because groups start from the same treatment status in this pre-baseline periods, this can be seen as a canonical DiD, and all the standard results apply. 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. This allows the researcher to estimate both the effects of the policy change and its reversion, and also test whether outcomes immediately revert back to their initial trend.

For example, [Namen \(2021\)](#) studies an education reform in Colombia that capped schools’ grade retention rates at 5 percent. The policy was implemented in 2002 and later removed in 2009. By exploiting the extent to which schools were affected by this policy, this study compares the periods 2002-2008 and 2009-2012 to the period 1998-2001 to estimate the effects of implementing and removing the grade retention cap on students’ outcomes.

### 3.2 Estimation in Reverse for DiD with Converging Treatment Status

In settings where treatment status converges after policy change, visual examination of Figures 1 and 3 suggests an alternative estimation approach: perhaps, it is possible to estimate ATT “in reverse”, simply by re-ordering time periods. Intuitively, if one designates the reference period  $\tilde{t}$  to be some period after the policy change, then pre-policy change coefficients  $\hat{\theta}_\ell$  may estimate the ATT of the policy change. In this section we discuss the validity of this approach.

Consider a setting where groups start from different treatment levels and then converge to the same treatment status, i.e. suppose  $d_{\text{pre}} \neq d_0$  and  $d = d_0$  as in Section 1.3. Let the baseline period  $\tilde{t}$  be some post-policy change period, i.e. let  $\tilde{t} \geq t^*$ . The following proposition is equivalent to Proposition 1 applied to convergent treatment status cases.

**Proposition 2 (Reverse DiD)** *Let  $\{\beta_\ell\}_{\{\ell \geq t^*\}}$  and  $\{\theta_\ell\}_{\{\ell \neq t^*, \ell < t^*\}}$  be the linear projection coefficients from Equation (1). When  $d_{\text{pre}} \neq d_0$  and  $d_{\text{post}} = d_0$ ,  $\tilde{t} \geq t^*$ , for  $\ell \geq t^*$ ,  $\ell \neq \tilde{t}$ ,*

$$\beta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0, d_0) - Y_{i\ell}(d_0, d_0) | S_i = 0] - \mathbb{E}[Y_{i\tilde{t}}(d_0, d_{\text{pre}}) - Y_{i\ell}(d_0, d_{\text{pre}}) | S_i = 1]$$

and for  $\ell < t^*$ ,

$$\begin{aligned} \theta_\ell &= \mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\ell}(d_0) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\tilde{t}}(d_0, d_0) - Y_{i\ell}(d_0) | S_i = 0] - \mathbb{E}[Y_{i\tilde{t}}(d_0, d_{\text{pre}}) - Y_{i\ell}(d_0) | S_i = 1]. \end{aligned}$$

If, in addition, Assumption 2 holds, for  $\ell \geq t^*$ ,  $\ell \neq \tilde{t}$ ,

$$\beta_\ell = \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\ell}(d_0) | S_i = 0] - \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\ell}(d_0) | S_i = 1]$$

and for  $\ell < t^*$ ,

$$\begin{aligned} \theta_\ell &\rightarrow_{\mathbb{P}} \mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\ell}(d_0) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\ell}(d_0) | S_i = 0] - \mathbb{E}[Y_{i\tilde{t}}(d_0) - Y_{i\ell}(d_0) | S_i = 1]. \end{aligned}$$

Proposition 2 demonstrates that the pre-policy change coefficients  $\theta_\ell$  equal the ATT

of the policy change,  $\mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\ell}(d_0) | S_i = 1]$ , if the parallel trends Assumption 3 holds and if we are willing to assume that potential outcomes are static (i.e. Assumption 2 holds). Therefore, for settings demonstrated in Panel A of Figures 1 and 3, one can indeed recover the ATT by choosing an alternative reference point.

However, the result does not hold when potential outcomes are dynamic. If switching treatment status leads to different outcome trends compared to the case in which treatment status remains constant throughout the period, then  $Y_{i\tilde{t}}(d_0, d_{\text{pre}}) \neq Y_{i\tilde{t}}(d_0, d_0)$ , and therefore  $\hat{\theta}_\ell$  will not recover the ATT.

The assumption of static potential outcomes is very strong, as it requires, for instance, that average potential outcomes immediately return to their pre-treatment trend once the treatment is removed. Therefore, the reverse estimation strategy is hard to justify, unless researchers have prior knowledge on the nature of outcome dynamics.

### 3.3 Linear Trend Adjustment

An appealing and frequently used solution to account for differential trends in the baseline period is to include a linear trend term in the usual event study specification. For example, Jakobsen et al. (2020) account for differential trends in Figures 2(b) and 4 by including “a linear differential pretrend identified based on  $n$  prereform years (i.e. the omitted years in the first term on the right-hand side).” Intuitively, the goal of the linear trends adjustment is to relax a parallel trends assumption such as Assumption 5 from requiring no differences in trends to allowing for differences that are constant over time, that is:

**Assumption 6 (Constant Differences in Trends)** *For any  $\ell \geq t^*$  and  $\tilde{t} < t^*$ , and for some constant  $\kappa$  that does not depend on  $\ell$  or  $\tilde{t}$ ,*

$$\mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\tilde{t}}(d_{\text{pre}}) | S_i = 1] - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\tilde{t}}(d_0) | S_i = 0] = \kappa.$$

Note that the parallel trends Assumption 5 amounts to setting  $\kappa = 0$ .

Let  $\mathcal{T}_{\text{LA}} \subseteq \{1, \dots, t^* - 1\}$  be the subset of periods used to estimate the linear trend adjustment and let  $T_{\text{LA}} \geq 2$  be the number of elements in  $\mathcal{T}_{\text{LA}}$ . We assume that  $\mathcal{T}_{\text{LA}}$  consists of consecutive periods,  $\mathcal{T}_{\text{LA}} = \{t_m, t_m + 1, \dots, t_M - 1, t_M\}$ , where  $1 \leq t_m < t_M \leq t^* - 1$ . Define the average and the variance of time periods in set  $\mathcal{T}_{\text{LA}}$  as  $\bar{t}_{\text{LA}} = \sum_{t \in \mathcal{T}_{\text{LA}}} t / T_{\text{LA}}$ ,  $V_{\text{LA}} = \sum_{t \in \mathcal{T}_{\text{LA}}} (t - \bar{t}_{\text{LA}})^2 / T_{\text{LA}}$ , and let  $\Delta Y_{it} = Y_{it} - Y_{it-1}$  denote one-period differences. We now consider the following extension of the basic



event-study specification from Equation (1):

$$Y_{it} = \alpha_i + \delta_t + \gamma S_{it} + \sum_{\ell < t^*, \ell \notin \mathcal{T}_{\text{LA}}} \theta_\ell^{\text{LA}} S_i \mathbb{1}(t = \ell) + \sum_{\ell \geq t^*} \beta_\ell^{\text{LA}} S_i \mathbb{1}(t = \ell) + u_{it} \quad (2)$$

where we refer to  $\gamma$  as the linear adjustment coefficient. The following result characterizes the OLS estimators from this equation and is similar to Lemma 1.

**Lemma 2** *Let  $\gamma$ ,  $\{\beta_\ell^{\text{LA}}\}_\ell$  and  $\{\theta_\ell^{\text{LA}}\}_\ell$  be the linear projection coefficients from Equation (2) and  $\hat{\gamma}$ ,  $\{\hat{\beta}_\ell^{\text{LA}}\}_\ell$  and  $\{\hat{\theta}_\ell^{\text{LA}}\}_\ell$  be the corresponding OLS estimators. Then,*

$$\gamma = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma (\mathbb{E}[\Delta Y_{it} | S_i = 1] - \mathbb{E}[\Delta Y_{it} | S_i = 0]),$$

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

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

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

$$\theta_\ell^{\text{LA}} = \mathbb{E} \left[ Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M} \Delta Y_{it} \omega_t^\ell \middle| S_i = 1 \right] - \mathbb{E} \left[ Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M} \Delta Y_{it} \omega_t^\ell \middle| S_i = 0 \right]$$

where

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

Under Assumption 1, as  $n \rightarrow \infty$  with  $T$  fixed,

$$\begin{aligned} \hat{\gamma} &= \sum_{t=t_m+1}^{t_M} \omega_t^\gamma \left( \frac{\sum_{i=1}^n \Delta Y_{it} S_i}{\sum_{i=1}^n S_i} - \frac{\sum_{i=1}^n \Delta Y_{it} (1 - S_i)}{\sum_{i=1}^n (1 - S_i)} \right) \rightarrow_{\mathbb{P}} \gamma, \\ \hat{\beta}_\ell^{\text{LA}} &= \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1 - \bar{S})} \left( Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M} \Delta Y_{it} \omega_t^\ell \right) \rightarrow_{\mathbb{P}} \beta_\ell^{\text{LA}}, \\ \hat{\theta}_\ell^{\text{LA}} &= \sum_i \frac{(S_i - \bar{S})}{n\bar{S}(1 - \bar{S})} \left( Y_{i\ell} - Y_{it_m} - \sum_{t=t_m+1}^{t_M} \Delta Y_{it} \omega_t^\ell \right) \rightarrow_{\mathbb{P}} \theta_\ell^{\text{LA}}. \end{aligned}$$

The proof in the appendix shows that  $\omega_t^\ell \geq 0$  and  $\sum_{t=t_m+1}^{t_M} \omega_t^\ell = \ell - t_m$ , while  $\omega_t^\gamma \geq 0$

and  $\sum_{t=t_m+1}^{t_M} \omega_t^\gamma = 1$ . Also notice that the weights are non-random and observable, since they are functions of the time periods only.

Next, we link the coefficients from Equation (2) to potential outcomes, similarly to Proposition 1.

**Proposition 3 (Linear trend adjustment)** *Let  $\gamma$ ,  $\{\beta_\ell^{\text{LA}}\}_\ell$  and  $\{\theta_\ell^{\text{LA}}\}_\ell$  be the linear projection coefficients from Equation (2). Then,*

$$\gamma = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma (\mathbb{E}[\Delta Y_{it}(d_{\text{pre}})|S_i = 1] - \mathbb{E}[\Delta Y_{it}(d_0)|S_i = 0]),$$

for  $\ell \geq t^*$ ,

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

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

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

with weights  $\omega_t^\gamma$  and  $\omega_t^\ell$  defined in Lemma 2. If, in addition, Assumption 6 holds, then for  $\ell \geq t^*$ ,

$$\beta_\ell^{\text{LA}} = \mathbb{E}[Y_{i\ell}(d_{\text{post}}, d_{\text{pre}}) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}})|S_i = 1].$$

Proposition 3 shows that the linear trend coefficient  $\gamma$  equals a weighted average of pre-policy differences in trends between switchers and non-switchers. These weights  $\omega_t^\gamma$  are non-negative, sum up to one, and are quadratic in nature, with the largest weight given to the middle of the estimation period  $\mathcal{T}_{\text{LA}}$  (specifically, period  $\bar{t} + 0.5$ ), and are decreasing towards early and late periods  $t_m$  and  $t_M$ , respectively.

The coefficient  $\gamma$  is equal to zero if the pre-trends are parallel, that is,  $\mathbb{E}[\Delta Y_{it}(d_0)|S_i = 1] = \mathbb{E}[\Delta Y_{it}(d_{\text{pre}})|S_i = 0]$ , in the chosen estimation period  $\mathcal{T}_{\text{LA}}$ . Therefore, the inclusion

of a linear trend helps evaluate the validity of Assumption 5: a non-zero and statistically significant estimate of  $\gamma$  would imply that this assumption is unlikely to hold. As discussed previously, this assumption imposes restrictions on both the evolution of outcomes under the baseline treatment status and on the evolution of treatment effects.

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 weights  $\omega_t^\gamma$  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.

Proposition 3 further shows that the event study coefficients  $\beta_\ell^{\text{LA}}$  and  $\theta_\ell^{\text{LA}}$  equal the same estimand as the event-study coefficients  $\beta_\ell$  and  $\theta_\ell$  from Proposition 1 with  $\tilde{t} = t_m$ , minus an adjustment term consisting of a linear combination of differences in pre-policy trends between switchers and non-switchers.

Most importantly,  $\beta_\ell^{\text{LA}}$  recovers the ATT if the parallel trends assumption (Assumption 5) holds in all periods or if the differences in trends are constant over time, that is, Assumption 6 holds. Practically this means that including a linear adjustment factor will not bias estimates in settings where the parallel trend assumption actually holds (although doing so may reduce the precision of the estimates). The upside of including a linear trend term is that it will eliminate the asymptotic bias in settings with constant differences in trends, e.g. in settings with unequal baseline status and linear treatment effects (illustrated in Figure 1(b) and 3(b)).

However, inclusion of a linear trend adjustment is no panacea: if the differences in trends between the switchers and non-switchers are not constant over time, then both estimates  $\hat{\beta}_\ell$  and  $\hat{\theta}_\ell$  from Equation (1) and estimates  $\hat{\beta}_\ell^{\text{LA}}$  and  $\hat{\theta}_\ell^{\text{LA}}$  from Equation (2) are inconsistent, and it is not possible in general to determine which asymptotic bias is larger. Appendix A illustrates this fact by considering a simple three-period setting. Under one assumption simple event study produces estimates with smaller bias while under a different assumption the bias is smaller for a specification with a linear term trend.

To conclude, inclusion of a linear trend in event study specification will result in consistent estimates in settings where the parallel trend assumption holds or where the differences in trends are constant over time. However, if the differences in trends are non-constant, using specification (2) instead of (1) may increase the bias. Therefore,

the choice of specification should be guided by economic theory or prior empirical evidence, or direct observation of pre-reform trends, if the data series are sufficiently long.

How many periods should one use to estimate the linear adjustment term, in other words, what periods should be included in the set  $\mathcal{T}_{LA}$ ? Unfortunately, it is not possible to provide a precise recommendation without making further assumptions on the evolution of potential outcomes and treatment effects. It is worth emphasizing, however, that the larger the number of periods used to estimate  $\hat{\gamma}$ , the more the strategy relies on the validity of the linearity assumption. On the other hand, when only two periods are used to estimate the difference in trends, this approach can be thought of as a “non-parametric” adjustment that subtracts an additional difference term from the DiD estimand (see Appendix A for details).

### 3.4 Sensitivity Analysis

In the context of canonical DiDs, recent studies by [Manski and Pepper \(2018\)](#) and [Rambachan and Roth \(2023\)](#) have proposed that researchers evaluate the sensitivity of treatment effect estimates to deviations from traditional 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 over time. We refer the reader to the aforementioned papers for further details on this type of sensitivity analysis.

## 4 Conclusion

In this paper we consider DiD settings in which treatment and control group experience unequal treatment statuses in the baseline period. Our analysis demonstrates that the usual event study estimation approach will not recover the ATT unless the treatment effect is constant over time. Furthermore, we show that the traditional parallel trend test is inappropriate and cannot provide adequate support for or against the parallel trends assumption.

We evaluate two intuitive solutions. First, we show that in settings with convergent treatment status, estimation in reverse will recover the ATT, but only when the

outcomes are static. Second, correcting for divergent trends by including a linear term trend works but only in settings where differences in trends are constant over time. Otherwise, event studies with a linear trend may lead to more biased estimates than event studies without a linear trend.

## References

- 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.
- Barreca, Alan I., Matthew Neidell, and Nicholas J. Sanders**, “Long-run Pollution Exposure and Mortality: Evidence from the Acid Rain Program,” *Journal of Public Economics*, 2021, *200*, 104440.
- 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,” 2021.
- 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,” 2021.
- 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.
- 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, C. and X. D’Haultfœuille**, “Fuzzy Differences-in-Differences,” *The Review of Economic Studies*, April 2018, *85* (2), 999–1028.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110*, 2964–2996.
- and **Xavier D’Haultfœuille**, “Two-way fixed effects and Differences-in-Differences with Heterogeneous Treatment Effects: a Survey,” *The Econometrics Journal*, 06 2022.
- , **Xavier D’Haultfœuille, Félix Pasquier, and Gonzalo Vazquez-Bare**, “Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period,” SSRN Scholarly Paper 4011782, Rochester, NY January 2022.
- 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.
- 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.
- 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.

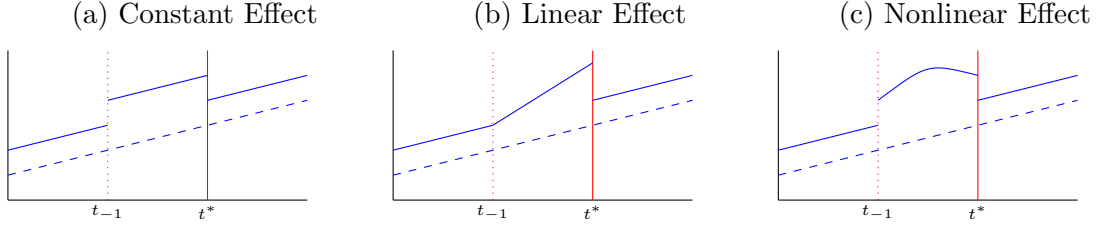
- Lucas, Adrienne M.**, “Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka,” *American Economic Journal: Applied Economics*, April 2010, 2 (2), 46–71.
- Maffioli, Alessandro, Diego Ubfal, Gonzalo Vazquez-Bare, and Pedro Cerdán-Infantes**, “Extension Services, Product Quality and Yields: the Case of Grapes in Argentina,” *Agricultural Economics*, 2011, 42 (6), 727–734.
- 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.
- Moser, Petra and Alessandra Voena**, “Compulsory Licensing: Evidence from the Trading with the Enemy Act,” *American Economic Review*, February 2012, 102 (1), 396–427.
- Namen, Olga**, “The Impact of Encouraging Social Promotion in Schools,” SSRN Scholarly Paper 340219 October 2021.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *The Review of Economic Studies*, 02 2023.
- 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**, “Efficient Estimation for Staggered Rollout Designs,” Mimeo 2021.
- **and —**, “When is Parallel Trends Sensitive to Functional Form?,” *Econometrica*, 2023, 91, 737–747.
- , — , **Alyssa Bilinski, and John Poe**, “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature,” Mimeo 2022.



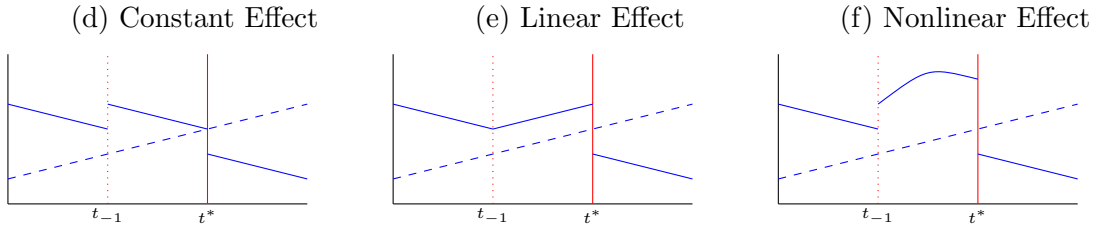
- Sant’Anna, Pedro H.C. and Jun Zhao**, “Doubly Robust Difference-in-Differences Estimators,” *Journal of Econometrics*, 2020, *219* (1), 101–122.
- Sawada, Yasuyuki, Takeshi Aida, Andrew S. Griffen, Eiji Kozuka, Haruko Noguchi, and Yasuyuki Todo**, “Democratic Institutions and Social Capital: Experimental 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: Policy Reversal: Examples of Possible Outcome Paths

**Panel A: Comparable Groups, Immediate Treatment Effect**



**Panel B: Not Comparable Groups, Immediate Treatment Effect**



*Notes:* This figure illustrates the possible evolution of outcomes for treatment (solid line) and control (dashed line) groups under different scenarios, in a setting where a policy is adopted for some individuals in period  $t_{-1}$  and then is cancelled in period  $t^*$ . Panel A assumes that groups are truly comparable and therefore satisfy the “parallel trends” assumption in the pre-baseline period under Assumption 3, while Panel B assumes the opposite. All figures assume that potential outcomes are static, i.e. Assumption 2 holds. Three general types of treatment effects are considered: an immediate constant effect in Figures (a) and (d), a persistent linear effect in Figures (b) and (e), and a nonlinear effect in Figures (c) and (f). For simplicity, all figures assume that treatment leads to an increase in outcome, and that treatment effects are homogenous across groups.

Figure 2: Empirical Illustration – Convergent Treatment Status: [Jakobsen et al. \(2020\)](#)

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

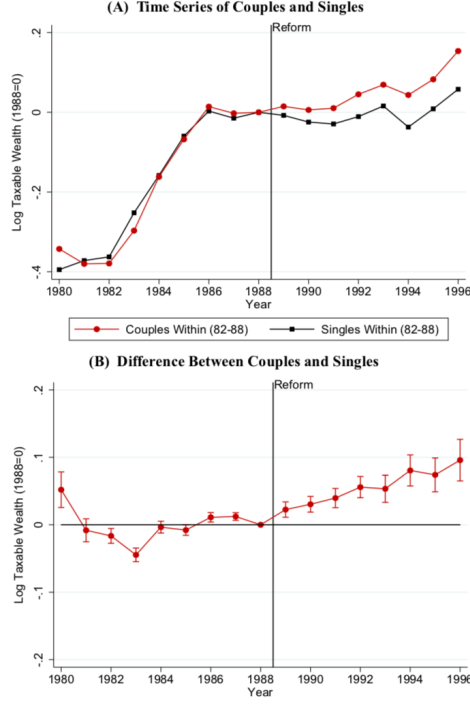
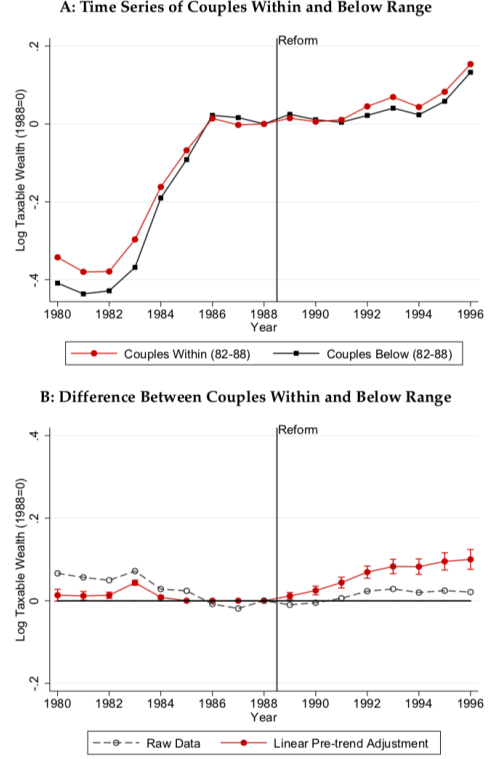


FIGURE IV  
Difference-in-Differences Comparing Couples and Singles within Exempted Range

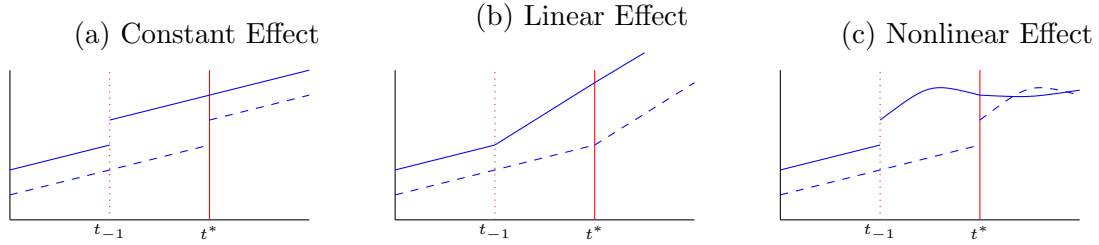
(b) Unequal baseline status:  
treated are taxed but control are not



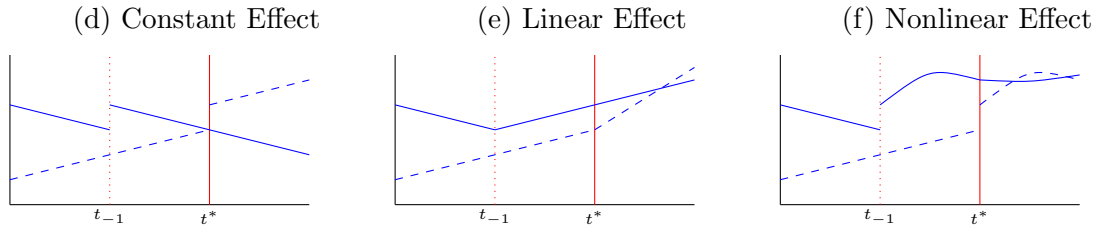
*Notes:* This figure reproduces (a) Figure IV and (b) Appendix Figure A.VII from [Jakobsen et al. \(2020\)](#). These figures show evolution of taxable wealth and difference between treatment and control groups before and after 1989 reform that increased the exemption threshold for couples but not for single individuals. Loosely speaking, in both figures, 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 – “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 in black squares.

Figure 3: Universal Adoption: Examples of Possible Outcome Paths

**Panel A: Comparable Groups, Immediate Treatment Effect**



**Panel B: Not Comparable Groups, Immediate Treatment Effect**



*Notes:* This figure illustrates the possible evolution of outcomes for treatment (solid line) and control (dashed line) groups under different scenarios, in a setting where a policy is adopted for some individuals in period  $t_{-1}$  and then is adopted for all individuals in period  $t^*$ . Panel A assumes that groups are truly comparable and therefore satisfy the “parallel trends” under Assumption 4, while Panel B assumes the opposite. All figures assume that potential outcomes are static, i.e. Assumption 2 holds. Three general types of treatment effects are considered: an immediate constant effect in Figures (a) and (d), a persistent linear effect in Figures (b) and (e), and a nonlinear effect in Figures (c) and (f). For simplicity, all figures assume that treatment leads to an increase in outcome, and that treatment effects are homogenous across groups.

Figure 4: Empirical Illustration – Non-Convergent Treatment Status: [Jakobsen et al. \(2020\)](#)

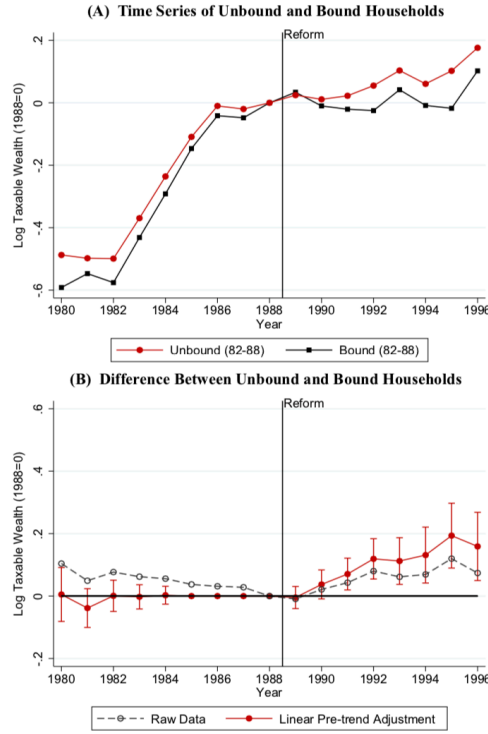


FIGURE VI  
Difference-in-Differences Comparing Households Unbound and Bound  
by Tax Ceiling

*Notes:* This figure reproduces Figure VI from [Jakobsen et al. \(2020\)](#). This figures show evolution of taxable wealth and difference between treatment and control groups before and after 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 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 in black squares.

# APPENDIX

## A Event Studies With or Without Linear Trend Assumption – Illustration

If there are non-linear trends between switchers and non-switchers, then both the event study estimator and the linear adjustment estimator are inconsistent. Importantly, it is not possible to determine which asymptotic bias is larger in general.

To see this, consider the following simple example. Suppose there are two pre-change periods, and one post-change period, i.e.  $t = 1, 2, 3$  and  $t^* = T = 3$ . Consider a policy reversal setting where the switchers are treated in periods 2 and 3 and then the treatment is removed in period 3 while the stayers remain untreated, i.e.  $d = d_0$ . Also suppose that the parallel trends Assumption 3 holds.

We will use the first two periods to estimate the linear trend, so that  $\mathcal{T}_{\text{LA}} = \{1, 2\}$ . The effect of the policy is then estimated in period  $\ell = 3$ . Applying Proposition 3, we find that the event-study and linear-adjusted estimators converge to:

$$\begin{aligned} \hat{\beta}_\ell \rightarrow_{\mathbb{P}} \beta_\ell^{\text{ES}} &= \mathbb{E}[Y_{i3}(d_0, d_{\text{pre}}) - Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E}[Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1], \end{aligned}$$

and

$$\begin{aligned} \hat{\beta}_\ell^{\text{LA}} \rightarrow_{\mathbb{P}} \beta_\ell^{\text{LA}} &= \mathbb{E}[Y_{i3}(d_0, d_{\text{pre}}) - Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E}[Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] \\ &\quad - (\mathbb{E}[Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] - \mathbb{E}[Y_{i1}(d_{\text{pre}}) - Y_{i1}(d_0) | S_i = 1]). \end{aligned}$$

Then the bias of event-study estimator equals to the change in the ATT between periods 2 and 3, whereas the bias of the linear-adjusted estimator is the difference in changes of the ATT between periods 2 and 3 and periods 1 and 2.

We now consider the following four cases, each resulting in an asymptotic bias summarized in Table 1.

**Case 1: constant TE.** Suppose that

$$0 = \mathbb{E}[Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] - \mathbb{E}[Y_{i1}(d_{\text{pre}}) - Y_{i1}(d_0) | S_i = 1]$$

$$0 = \mathbb{E} [Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1].$$

so that the biases are zero for both estimators.

**Case 2: linear TE.** Suppose that for some constant  $\kappa$

$$\begin{aligned} \kappa &= \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] - \mathbb{E} [Y_{i1}(d_{\text{pre}}) - Y_{i1}(d_0) | S_i = 1] \\ \kappa &= \mathbb{E} [Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1]. \end{aligned}$$

**Case 3: nonlinear decreasing TE.** Suppose that for some constant  $\kappa$

$$\begin{aligned} 3\kappa &= \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] - \mathbb{E} [Y_{i1}(d_{\text{pre}}) - Y_{i1}(d_0) | S_i = 1] \\ \kappa &= \mathbb{E} [Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1]. \end{aligned}$$

**Case 4: nonlinear increasing TE.** Suppose that for some constant  $\kappa$

$$\begin{aligned} \kappa &= \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1] - \mathbb{E} [Y_{i1}(d_{\text{pre}}) - Y_{i1}(d_0) | S_i = 1] \\ 3\kappa &= \mathbb{E} [Y_{i3}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i3}(d_0, d_0) | S_i = 1] - \mathbb{E} [Y_{i2}(d_{\text{pre}}) - Y_{i2}(d_0) | S_i = 1]. \end{aligned}$$

	ES estimator	LA estimator
Case 1: constant TE	0	0
Case 2: linear TE	$\kappa$	0
Case 3: nonlinear decreasing TE	$\kappa$	$-2\kappa$
Case 4: nonlinear increasing TE	$3\kappa$	$2\kappa$

Table 1: Asymptotic bias - ES vs LA estimators

## B Proofs

**Setup.** For any variable  $A_{it}$  and any subset  $\mathcal{T} \subseteq \{1, 2, \dots, T\}$  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  $\check{A}_{it}(\mathcal{T}) = Y_{it} - \bar{A}_i(\mathcal{T}) - \mathbb{E}[A_{it}] + \mathbb{E}[\bar{A}_i(\mathcal{T})]$ . Also let  $X_{it}^\ell = \mathbb{1}(t = \ell)S_i$  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:

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

where

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

and the event-study regression with a linear adjustment

$$Y_{it} = \alpha_i + \delta_t + \sum_{\ell \notin \mathcal{T}_{\text{LA}}} \beta_\ell^{\text{LA}} X_{it} + \gamma Z_{it} + u_{it}$$

where

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

and where  $\mathcal{T}_{\text{LA}}$  is the subset of periods used to estimate the linear trend and  $T_{\text{LA}} = |\mathcal{T}_{\text{LA}}|$  (see Section 3.3 for details).

**Proof of Lemma 1.** The population coefficients  $\{\beta_\ell\}_\ell$  are characterized by the system of equations:

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

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

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

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

$$\hat{\beta}_\ell = \frac{\sum_i \ddot{Y}_{i\ell} S_i}{n\bar{S}(1-\bar{S})} + \frac{1}{T} \sum_{\ell \neq \bar{t}} \hat{\beta}_\ell. \quad (3)$$

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

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

and thus

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

using that  $\sum_{\ell \neq \bar{t}} \ddot{Y}_{i\ell} = \sum_\ell \ddot{Y}_{i\ell} - \ddot{Y}_{i\bar{t}} = -\ddot{Y}_{i\bar{t}}$ . Plugging back into (3),

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

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

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

and by the law of large numbers, as  $n \rightarrow \infty$ ,  $\hat{\beta}_\ell \rightarrow_{\mathbb{P}} \mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 1] - \mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 0]$  which gives the desired result.  $\square$

**Proof of Proposition 1.** If  $\ell \geq t^*$ ,

$$\begin{aligned} \mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 1] - \mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 0] &= \mathbb{E}[Y_{i\ell}(d_{\text{post}}, d_{\text{pre}}) - Y_{i\bar{t}}(d_{\text{pre}}) | S_i = 1] \\ &\quad - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\bar{t}}(d_0) | S_i = 0] \\ &= \mathbb{E}[Y_{i\ell}(d_{\text{post}}, d_{\text{pre}}) - Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) | S_i = 1] \\ &\quad + \mathbb{E}[Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) - Y_{i\bar{t}}(d_{\text{pre}}) | S_i = 1] \\ &\quad - \mathbb{E}[Y_{i\ell}(d_0, d_0) - Y_{i\bar{t}}(d_0) | S_i = 0]. \end{aligned}$$

On the other hand, when  $\ell < t^*$ ,

$$\mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 1] - \mathbb{E}[Y_{i\ell} - Y_{i\bar{t}} | S_i = 0] = \mathbb{E}[Y_{i\ell}(d_{\text{pre}}) - Y_{i\bar{t}}(d_{\text{pre}}) | S_i = 1]$$

$$- \mathbb{E}[Y_{i\ell}(d_0) - Y_{i\bar{\ell}}(d_0)|S_i = 0].$$

Finally, when Assumption 2 holds,  $Y_{i\ell}(d_{\text{post}}, d_{\text{pre}}) = Y_{i\ell}(d_{\text{post}})$  and  $Y_{i\ell}(d_{\text{pre}}, d_{\text{pre}}) = Y_{i\ell}(d_{\text{pre}})$  which completes the proof.  $\square$

**Proof of Corollary 1.** Follows from Proposition 1 under Assumption 3.  $\square$

**Proof of Corollary 2.** Follows from Proposition 1 under Assumption 3.  $\square$

**Proof of Corollary 3.** Follows from Proposition 1 under Assumption 4.  $\square$

**Proof of Proposition 2.** The proof is analogous to that of Proposition 1 using Lemma 1.  $\square$

**Proof of Lemma 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  $\{\hat{\beta}_\ell^{\text{LA}}\}_{\ell \notin \mathcal{T}_{\text{LA}}}$  and  $\hat{\gamma}$  are characterized by the system of equations:

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

From the first equation,

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

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

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

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

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

so that

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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  $1 \leq u \leq T_{\text{LA}} - 1$  where  $T_{\text{LA}} = t_M - t_m + 1$ . Next, noting that  $u = t - t_m$  implies  $t_M + 1 - t = T_{\text{LA}} - u$ ,

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

By a similar argument,

$$\begin{aligned}
\sum_{t \in \mathcal{T}_{\text{LA}}} (t - \bar{t}_{\text{LA}})^2 &= \sum_{t \in \mathcal{T}} t(t - \bar{t}) = \sum_{t=t_m}^{t_M} (t - t_m)(t - t_m - (\bar{t}_{\text{LA}} - t_m)) \\
&= \sum_{u=0}^{T_{\text{LA}}-1} u \left( u - \frac{T_{\text{LA}} - 1}{2} \right) \\
&= \sum_{u=1}^{T_{\text{LA}}-1} u \left( u - \frac{T_{\text{LA}} - 1}{2} \right) \\
&= \frac{(T_{\text{LA}} - 1)T_{\text{LA}}(T_{\text{LA}} + 1)}{6} - \frac{T_{\text{LA}}(T_{\text{LA}} - 1)^2}{4} \\
&= \frac{(T_{\text{LA}} - 1)T_{\text{LA}}(T_{\text{LA}} + 1)}{12} \\
&= \frac{1}{2} \binom{T_{\text{LA}} + 1}{3}
\end{aligned}$$

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

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

but

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

so

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

where

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

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

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

and a standard application of the law of large numbers completes the proof.  $\square$

**Proof of Proposition 3.** Plugging in the potential outcomes,

$$\gamma = \sum_{t=t_m+1}^{t_M} \omega_t^\gamma (\mathbb{E}[\Delta Y_{it}(d_{\text{pre}})|S_i = 1] - \mathbb{E}[\Delta Y_{it}(d_0)|S_i = 0])$$

where the equality uses that  $t_M \leq t^*$ . Similarly,

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

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

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

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

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

Finally note that when the difference in trends is constant,

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

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