A Guide to Dynamic Difference-in-Differences Regressions for Political Scientists

Zikai Li*, Anton Strezhnev†

June 25, 2024

Abstract

Difference-in-differences (DiD) designs for estimating causal effects have grown in popularity throughout political science. It is common for DiD studies report their main results using a "dynamic" or "event study" two-way fixed effects (TWFE) regression. This regression combines estimates of average treatment effects for multiple post-treatment time periods alongside placebo tests of the main identifying assumption: parallel trends. Despite their ubiquity, there is little clear and consistent guidance in the discipline for how researchers should estimate dynamic treatment effects. This paper develops a novel decomposition of the dynamic TWFE regression coefficients in terms of their component 2×2 difference-in-differences comparisons in the style of Goodman-Bacon (2021). We use this decomposition to illustrate how bias can result from the incorrect specification of baseline time periods, the inclusion of units and time periods where all observations are treated, and heterogeneity in the dynamic treatment effects across different treatment timing groups. Our results provide additional intuition for the source of bias due to effect heterogeneity—what Sun and Abraham (2021) term "contamination bias"—by directly characterizing the contaminated 2×2 comparisons. We then provide a common framework for connecting the many proposed "heterogeneity-robust" estimators in the literature, noting that they vary primarily in which 2×2 comparisons they choose to include. Through a replication of three studies published in prominent political science journals, we conclude by showing how attentiveness to baseline selection and specification can alter findings.

*Author contact information: PhD Student, University of Chicago, Department of Political Science. Email: zkl@uchicago.edu

†Author contact information: Assistant Professor, University of Chicago, Department of Political Science. Email: astrezhnev@uchicago.edu

‡We thank Molly Offer-Westort, Andy Eggers, Bobby Gulotty, Adam Glynn, Nahomi Ichino, Kosuke Imai, Melody Huang, Matthew Blackwell, Luke Miratrix, Andy Halterman and Amanda Weiss as well as participants at the 2023 APSA Annual Meeting, the Harvard IQSS Applied Statistics workshop, and the Methods and IR seminar for helpful comments and conversations. We are very grateful to Andy Hall, Jesse Yoder, Jake Grumbach, Charlotte Hill and Agustina Paglayan for making available their replication archives and for answering our questions in working through the data and code. This draft supersedes a previous 2018 manuscript: "Semiparametric weighting estimators for multi-period difference-in-differences designs."

1 Introduction

Among social scientists, differences-in-differences (DiD) has become a canonical research design for inferring causal effects from non-experimental settings. Within the field of political science, the popularity of the DiD design has grown rapidly. As of August 2023, a Google Scholar search of the three major general interest political science journals—the American Political Science Review, the American Journal of Political Science, and the Journal of Politics—shows that of the approximately 405 articles that mention some variation of "difference-in-differences," 236, or more than half, were published in the last four years. These designs are popular in part due to their ability to address certain forms of unobserved confounding by leveraging repeated observations of units over time.

The conventional difference-in-differences design consists of observations of treated and control units across two time periods. Some units are treated in the second period while all units are untreated in the first. In the absence of randomized treatment assignment, the mean difference between treated and control units in the second period is biased for the average treatment effect on the treated (ATT). By incorporating an additional difference—the mean difference between treated and control in the baseline period one—the differences-in-differences estimator identifies the ATT under the assumption that the selection-into-treatment bias is constant across periods one and two. This identifying assumption, often referred to as "parallel trends," allows researchers to identify an average treatment effect even in the presence of confounding that is not observed directly.

Most modern difference-in-differences applications go beyond this 2×2 setting and incorporate observations over multiple pre- and post-treatment periods. This has two major benefits. First, it allows researchers to characterize the trajectory of a treatment's impact over time by estimating a separate treatment effect parameter for each post-treatment period. Second, it permits a series of falsification tests commonly referred to as "pre-trends" tests to assess the plausibility of the main identifying assumption. By applying the two-period difference-in-differences estimator to observations prior to the initiation of treatment, where the treatment effect is zero by construction, this test evaluates whether treated observations exhibit pre-treatment trends that differ from those of the control units. Placebo DiD estimates that diverge significantly from zero suggest that the identifying assumption of parallel trends may be invalid.

Regression estimators incorporating unit-specific and time-specific fixed effects parameters

(two-way fixed effects or TWFE) have been a popular method for estimating ATTs under parallel trends due to the numerical equivalence between the TWFE regression and the non-parametric 2×2 difference-in-differences estimator in the two-period setting. There are two general approaches for specifying the treatment effect parameter of interest in these regressions (Sun and Abraham 2021): a "static" specification that estimates a single coefficient on a single indicator regressor for treatment or a "dynamic" specification that includes multiple indicators for each *relative* time period since treatment adoption for each unit.[1] For the dynamic specification, the estimated parameters are often presented in the form of a coefficient plot—referred to as an "event study plot"—where the coefficients on pre-treatment periods are interpreted as placebo tests while those corresponding to post-treatment periods are interpreted as estimates of the treatment effect some number of time periods after adoption.

Because the dynamic TWFE regression provides researchers with both effect estimates and falsification tests in a single ordinary least squares regression, it has become a standard tool in the applied differences-in-differences toolkit. However, its connection to the non-parametric 2×2 differences-in-differences estimator is not as straightforward as one might hope. As a consequence, researchers often lack a proper understanding of the underlying comparisons that are used to construct the estimates and may make specification errors that result in invalid placebo tests and biased dynamic treatment effect estimates.

This paper develops a novel decomposition of the dynamic TWFE estimator into its component 2×2 differences-in-differences comparisons in order to better understand the sources of variation that this estimator leverages to identify each treatment effect. It builds on a recent wave of papers in the econometrics literature that highlight the sensitivity of two-way fixed effects regression estimators to treatment effect heterogeneity when treatment adoption is also *staggered* across units. When units can initiate treatment at different time periods, the equivalence between the TWFE estimator and the non-parametric 2×2 difference-in-differences estimator breaks down. Specifically, we expand on the results of Goodman-Bacon (2021), which decomposes the static two-way fixed effects estimator into an average over 2×2 differences-in-differences terms. In the presence of staggering, only some of these terms will identify a single ATT parameter for a partic-

1. This is also sometimes called a "leads-and-lags" specification as the relative treatment time indicators can be conceptualized as "leads" (for pre-treatment periods) or "lags" (for post-treatment periods) of the treatment indicator variable.

ular treatment timing group[2] and time period. Other terms—what Borusyak, Jaravel, and Spiess (2021) call "forbidden comparisons"—identify the sum of one ATT and the *difference* between two other ATTs for different treatment groups and time periods. The presence of this difference in ATTs results in what is sometimes termed the "negative weights" problem (De Chaisemartin and d'Haultfoeuille 2020) for the static TWFE regression. In the absence of further constraints on effect heterogeneity such that these differences in ATTs are always equal to zero, not all of the component differences-in-differences in the static TWFE regression will themselves identify an effect.

Applying this style of direct decomposition of the OLS estimator to the dynamic TWFE regression, we obtain an expression for the coefficient on the indicator for a particular relative treatment time in terms of six component 2×2 difference-in-differences terms. All but one of these terms is "contaminated" by the presence of one or more treatment effects for other, irrelevant treatment timing groups. We highlight three notable features of the dynamic TWFE regression, particularly in comparison with the static specification. First, we clarify the differences in the various effect homogeneity assumptions that researchers might wish to impose and how they facilitate the identification of treatment effect averages in both the static and dynamic TWFE regressions. We show that the effect homogeneity assumption required for the dynamic specification is *not* simply a weaker version of the homogeneity assumption required for the static specification. Namely, if treatment effects are homogeneous over time within each treated group but *vary* across treatment timing group (e.g. if late adopters have larger effects than early adopters), the static TWFE regression will identify a convex average of ATTs but the dynamic TWFE coefficients will not.

Second, we show that, even under constant effects, the dynamic TWFE coefficient does not generally have a simple interpretation as an average over differences-in-differences estimators. This is in contrast to the static TWFE regression, where an appropriate effect homogeneity assumption yields a static TWFE coefficient in which each component 2×2 can be interpreted as identifying a single ATT parameter, justifying the common interpretation of the TWFE coefficient as a convex average over "differences-in-differences." In the dynamic TWFE, while some component 2×2s are "clean" differences-in-differences that identify a single ATT, most component terms identify a mixture of relevant and irrelevant effects and some identify *only* irrelevant ATTs for other relative

2. That is, a group of units that initiate treatment at the same time.

time periods. Weights on each 2×2 comparison can be negative – in contrast to the static case. As such, the dynamic TWFE is better understood as a kind of "sequential" differences-in-differences estimator akin to triple-differences.

Third, we highlight the importance of the omitted *baseline* time period in the construction of the dynamic TWFE estimator and in the interpretation of the resulting estimates. We caution against selecting omitted relative time periods that do not appear for some treatment timing groups as this raises the sensitivity of the dynamic TWFE estimator to violations of the constant effects assumption. Intuitively, the estimator contains *no* clean difference-in-differences terms that will identify effects for these timing groups.

We apply the intuitions gained from this decomposition to further guide researchers in understanding the variety of "heterogeneity-robust" estimators that have been proposed in the recent difference-in-differences literature. These estimators can be understood in much the same way as the static and dynamic TWFE regressions—as choices over which 2×2 comparisons are valid for identifying a particular treatment effect or average of treatment effects. Unlike the conventional TWFE regressions, these choices are made *explicitly* with the aim of eliminating comparisons that are only valid under additional homogeneity assumptions. We emphasize that variation across these proposed estimators is driven primarily by differences in which units are considered controls and which time periods are defined as the baseline. We discuss the various approaches to constructing pre-trend placebo tests using these new methods, again noting the importance of clarity in defining the baseline comparison period(s) for each placebo estimate.

The paper proceeds as follows. The next subsection summarizes the recent literature on differences-in-differences under staggered adoption and effect heterogeneity and discusses where this paper is situated within the broader literature. Section 2 develops our primary theoretical contribution. It begins by reviewing the relevant causal estimands and core identification assumptions for the differences-in-differences design under staggered adoption. It then continues to a discussion of the static TWFE estimator and develops a decomposition of this estimator into three categories of component 2×2 terms and characterizes the effect homogeneity assumptions required for identification. It concludes by applying this decomposition approach to the dynamic TWFE regression, obtaining an expression in terms of six categories of 2×2 terms. In contrast to the static TWFE, this dynamic TWFE cannot be expressed as a convex average over 2×2 difference-in-differences

if treatment adoption is staggered. Section 3 reviews the proposed solutions to this problem, illustrating how they avoid the problem by deliberately selecting the 2×2 comparisons that will identify the treatment effect parameter of interest. Section 4 goes on to replicate three recent publications from prominent political science journals that present event study plots as part of their core analysis. It illustrates and corrects some common errors in the specification of the dynamic TWFE, such as failure to include sufficient lead and lag terms, the inclusion of "always-treated" units as controls, and the use of too few "never-treated" units. Our replication results show that conclusions from the dynamic TWFE regression can change substantially even under ostensibly minor specification changes. It then goes further to relax the assumption of homogeneous effect trajectories by estimating cohort-specific treatment effects using the approaches of Callaway and Sant'Anna (2021) and Sun and Abraham (2021), averaging over the distribution of cohorts to obtain valid relative-treatment time effect estimates and placebo tests. In all three replications, we find notably different interpretations of the results, even when the core finding of the paper appears to hold. The conclusion, Section 5, provides recommendations to researchers and readers for how to best use and evaluate event study plots in difference-in-differences designs.

1.1 Related literature

This paper builds on the extensive recent literature on estimating treatment effects in difference-in-differences designs when treatment adoption times vary across units and the failure of the popular two-way fixed effects estimator to recover unbiased treatment effect estimates. It is most closely related to Goodman-Bacon (2021) and Sun and Abraham (2021), which analyze and characterize the source of the bias in TWFE in the static and dynamic settings, respectively. Goodman-Bacon (2021) provides a decomposition of the OLS TWFE regression estimator for the coefficient on the static treatment indicator in terms of its component 2×2 difference-in-differences comparisons. It shows that this estimator identifies an average of treatment effects when effects vary across units but are constant across time, but this average imposes a non-uniform weighting on the ATTs. Effects for units that are treated "early" or "late" receive less weight due to the variance-weighting induced by the OLS estimator.[3] We develop a variation of the static Goodman-Bacon

3. See Aronow and Samii (2016) for an extended discussion of this feature of regression adjustment with a binary treatment.

(2021) decomposition that retains the 2×2 component comparisons but articulates the estimand in terms of averages over "group-time ATT" (Callaway and Sant'Anna 2021) effects.[4] This allows us to clarify the particular effect homogeneity assumptions required for the static TWFE to identify a (weighted) average of ATTs. We find that, in *addition* to the within-unit homogeneity assumption mentioned in Goodman-Bacon (2021), the TWFE estimator will also identify a (variance-weighted) average of group-time ATTs when effects are allowed to vary with *calendar time* but not *relative treatment time*.

We then apply this same decomposition approach to the dynamic TWFE regression analyzed in Sun and Abraham (2021). Our proof differs from that of Sun and Abraham (2021) as we work directly with the OLS estimator rather than decompose the population regression coefficient. This allows us to obtain an expression for the OLS estimator of a particular relative treatment time coefficient in terms of its component 2×2 comparisons and characterize the contribution of each comparison to the estimator. We recover the original contamination bias result and show—as in Sun and Abraham (2021)—that the regression identifies the relative treatment time effect only under the assumption that the source of effect heterogeneity is exclusively in relative time. Applying the same decomposition strategy to the static and dynamic regressions generates some novel insights about how these two estimators are related to the non-parametric 2×2 difference-in-differences. We show that the dynamic TWFE is closer to a kind of "triple" differences estimator rather than a strict average over DiDs. Each component 2×2 term does not necessarily identify a single group-time ATT, even when one imposes an effect homogeneity assumption—in stark contrast to the static TWFE. Rather, the regression weights are constructed such that treatment effect contamination in one set of 2×2 comparisons is eliminated by subtracting off one or more additional 2×2 comparisons involving other treated units or time periods.

We then review of a number of alternative estimators that have been proposed in this literature and provide a common framework for characterizing the differences between them. We focus on two main strands: estimators that directly select the relevant "clean control" 2×2 comparisons (Sun and Abraham 2021; Callaway and Sant'Anna 2021) and estimators that do so implicitly by imputing counterfactuals from a TWFE model fit to control observations (Borusyak, Jaravel, and Spiess 2021; Liu, Wang, and Xu 2022; Gardner 2022). We primarily focus on the staggered

4. Sun and Abraham (2021) refer to these as "cohort" ATTs.

adoption setting and do not review estimators that generalize to settings with treatment reversal, such as methods that define sets of treatment "switchers" versus units with "stable" treatment for each time period of interest (De Chaisemartin and d'Haultfoeuille 2020; Imai and Kim 2021). This includes extensions involving incorporating covariates by matching switchers to similar non-switchers (Imai, Kim, and Wang 2021), long versus short differences to capture different treatment effect dynamics, and continuous treatments (De Chaisemartin and d'Haultfoeuille 2024; Dube et al. 2023; Callaway, Goodman-Bacon, and Sant'Anna 2024). In each of these general cases, the proposed estimator under binary treatment or staggered adoption is equivalent to a variation of the Callaway and Sant'Anna (2021) estimator.

Our paper is also closely related to the "second wave" of papers that summarize and synthesize these contributions for applied researchers in order to develop guidelines on their use, such as De Chaisemartin and d'Haultfoeuille (2023), Roth et al. (2023), Baker, Larcker, and Wang (2022), Chiu et al. (2023), Hassell and Holbein (2023). We diverge from this work primarily through our new approach to decomposing and understanding the dynamic event study regression. Our aim is to help researchers build a better intuition for understanding how this operates by illustrating the implicit comparisons that it leverages for identification. This paper does not attempt to provide a comprehensive overview of difference-in-differences as practiced in the existing literature, nor does it provide a review of methods for relaxing the conventional difference-in-difference identification assumptions. Rather, our goal is to ensure that researchers have a clearer comprehension of the relationship between the dynamic TWFE regression and difference-in-differences design as well as a common framework for understanding how recently proposed new estimators relate to the regression. While we agree with Chiu et al. (2023) that failures of the parallel trends are arguably the most pervasive problem in the existing difference-in-differences literature, we also note that the task of constructing an appropriate *test* for the failure of parallel trends needs to be attentive to issues surrounding effect heterogeneity. We are concerned not only about the prevalence of parallel trends violations, but also that many published *evaluations* of the parallel trends assumption can be misleading. As we show in our replications, improper specifications of the pre-trends placebo test can result in estimates that are close to zero even when substantial violations are present.

2 Understanding the Dynamic TWFE Regression

In this section, we review the staggered adoption difference-in-differences design and the corresponding two-way fixed effects regression estimators. We first define the causal estimand(s) of interest and the main identification assumptions and connect these to the non-parametric 2×2 difference-in-differences-in-means estimator. We then turn to the "static" two-way fixed effects regression, providing a decomposition in the style of Goodman-Bacon (2021) to characterize the regression coefficient in terms of its component 2×2 comparisons. We define two potential effect homogeneity assumptions that allow the static TWFE to identify a (weighted) average of treatment effects across time and treatment group. We then move to the "dynamic" or "event study" TWFE specification to develop a novel proof of the "contamination bias" problem identified in the literature (Sun and Abraham 2021). Following a similar strategy to the static TWFE decomposition, we characterize the regression coefficient on any relative treatment time indicator as a sum over a series of 2×2 differences-in-differences terms. We show how the contamination bias is eliminated under the absence of staggering, in which case each coefficient in the event study specification corresponds to the non-parametric 2×2 difference-in-differences estimator relative to the baseline omitted period.

However, if treatment initiation is staggered, the average over 2×2 terms is not convex. Estimates of each dynamic treatment effect rely on matched comparisons with both never-treated units and units that have already received treatment. These latter terms will identify *differences* between different treatment effect terms, giving rise to the *negative weights* problem of De Chaisemartin and d'Haultfoeuille (2020). Crucially, identification requires that the relative-time effects to be homogeneous between units that initiate treatment at varying times. Under this assumption, the dynamic TWFE regression can be understood in terms of a series of differences *across* differences-in-differences comparisons. The weights on these successive differences are constructed in a manner that eliminates the "contamination" in each DiD that is induced by "irrelevant" treatment effects. Notably, we show that the assumption required for the dynamic TWFE is not weaker than the effect homogeneity assumptions required for identification under the static TWFE estimator. For the dynamic TWFE estimator, while the assumption required allows treatment effects to vary in relative time, it *does not permit* any other form of cross-unit effect heterogeneity.

2.1 The differences-in-differences design

We begin by defining the conventional differences-in-differences set-up, drawing notation primarily from the exposition in Roth et al. (2023). Consider a standard balanced panel design with a sample of N units observed over T time periods. Let $\mathcal{Y}$ and $\mathcal{D}$ denote the $N \times T$ matrices of observed outcomes and observed treatments respectively. Each element Y_{it} of $\mathcal{Y}$ denotes the outcome observed for unit i in time period t. Likewise, $D_{it} \in \{0, 1\}$ is an indicator for whether unit i is under treatment at time t. We make a number of initial restrictions on the structure of the treatment. First, we assume that units that initiate treatment remain under treatment for all time periods after initiation—there are *no treatment reversals*. This allows us to summarize the vector of treatment indicators for unit i with a single scalar. Next, we assume that all units are untreated at time period 1—that there are no "always-treated" units—and that at least some units never receive treatment during the period under observation—that there are no "always-treated" time periods. In a setting where all units eventually receive treatment, the sample can be trimmed in time such that the last units to initiate treatment become the "never-treated" group.[5]

Let $\mathcal{G} = \{2, 3, 4, 5, \ldots, T\}$ be the set of possible treatment initiation times. G_i denotes the time period in which unit i initiates treatment. Units with the same G_i belong to a common "treatment timing group" (Callaway and Sant'Anna 2021) or cohort (Sun and Abraham 2021) and share the same treatment history. We use the notation $G_i = \infty$ to represent units that are "never-treated" and remain under control for all time periods under analysis. We denote the number of units in a given treatment timing group g as N_g with $\sum_{g=1}^{T} N_g + N_\infty = N$.

Figure 1 provides a simple example of a setting with three time periods and two distinct treatment timing groups, those that adopt early ($G_i = 2$) and those that adopt in the later period ($G_i = 3$). An important feature of the staggered adoption setting is that the *calendar time* indexed by $t = 1, 2, 3, \ldots, T$ is not aligned with the *relative treatment time* for each unit that receives treatment. At any given time period t, the number of time periods that a treated unit i is away from treatment is equal to $t - G_i$.

Treatment effects are conventionally defined in terms of contrasts in potential outcomes (Rubin 1974). Define $Y_{it}(g)$ as the potential outcome we would observe at time t for unit i if, possibly con-

5. Notably, this implies that treatment effects cannot be estimated for time periods where all units are under treatment or for units that are never under control without imposing assumptions beyond those conventional to the difference-in-differences design.

	Time		
G_i	1	2	3
2	0	1	1
3	0	0	1
∞	0	0	0

Treatment status

	Time		
G_i	1	2	3
2	-1	0	1
3	-2	-1	0
∞	$-\infty$	$-\infty$	$-\infty$

Relative treatment time

Figure 1: Treatment timing groups and relative treatment times for $T = 3$

trary to fact, unit i initiated treatment at time g. A version of the standard consistency/SUTVA assumption (Rubin 1980; VanderWeele 2009) connects the observed outcomes to the potential outcomes.

Assumption 1 *Stable Unit Treatment Value*

$$Y_{it}(g) = Y_{it} \text{ if } G_i = g$$

In other words, the observed outcome for unit i at time t is equivalent to the potential outcome $Y_{it}(g)$ if unit i is observed to have initiated treatment at time g.

One natural causal quantity of interest is the effect of starting treatment at a *particular* period g on the outcome at a *particular* time t. We formalize this as the difference, among units observed initiating treatment at time g, between the average potential outcome at time t of starting treatment at time g and the average potential outcome at time t of never initiating treatment. This quantity is referred to as the group-time average treatment effect on the treated (ATT) in Callaway and Sant'Anna (2021) or the cohort ATT/CATT in Sun and Abraham (2021).

Definition 1 *Group-time Average Treatment Effect on the Treated*

$$ATT_g(t) = E[Y_{it}(g)|G_i = g] - E[Y_{it}(\infty)|G_i = g]$$

Note that the definition of the group-time ATT conditions on the observed treatment allocation $\mathcal{D}$ or the *design* – to use the language of De Chaisemartin and d'Haultfoeuille (2024). We

adopt this approach throughout the paper and treat all causal estimands as conditional on the observed distribution of treatment timing groups. This is in contrast to "random treatment timing" approaches to the panel setting, which explicitly characterize the treatment assignment process (Athey and Imbens 2022; Roth and Sant'Anna 2021). We consider the conditional estimand approach to be the more common justification for the DiD identifying assumptions in the existing literature. We motivate uncertainty as being driven by i.i.d. sampling from a target population, as in Callaway and Sant'Anna (2021) and Sun and Abraham (2021).[6]

The group-time ATTs form the building blocks of other causal quantities of interest. For example, one could consider aggregating all $ATT_g(t)$ into a single summary ATT using a set of weights on each of the group-time (g, t) combinations that sum to 1. One aggregate quantity of particular interest for the "dynamic" difference-in-differences specification is the "relative-time" ATT, which we define in Section 2.3, as a weighted average over all group-time ATTs, where t is a particular number of periods away from g.[7] We'll first consider identification of a *particular* group-time ATT.

While $E[Y_{it}(g)|G_i = g]$ can be identified directly under Assumption 1, identifying the counterfactual $E[Y_{it}(\infty)|G_i = g]$ from the observed data requires additional assumptions. The differences-in-differences design accomplishes this by imposing a *parallel trends* assumption and a *no anticipation* assumption. Parallel trends states that in the absence of treatment, the *trends* in the average potential outcomes across different treatment timing groups would be equivalent.

Assumption 2 *Parallel trends*

For all $t \neq t'$ and $g \neq g'$

$$E[Y_{it}(\infty) - Y_{it'}(\infty)|G_i = g] = E[Y_{it}(\infty) - Y_{it'}(\infty)|G_i = g']$$

We note that this is the strongest possible form of this assumption, where parallel trends is assumed with respect to all time periods and all treatment timing groups.[8] Strictly speaking,

6. An alternative motivation for uncertainty assumes a model for the potential outcomes under control that contains a stochastic error component (Borusyak, Jaravel, and Spiess 2021; Liu, Wang, and Xu 2022). For our purposes, the choice between these two frameworks is not consequential. See Appendix A2 of Borusyak, Jaravel, and Spiess (2021) which connects these two approaches by showing that i.i.d. sampling and the identifying assumptions below imply that particular model for the potential outcomes in a complete panel.

7. See Callaway and Sant'Anna (2021) for more extensive discussion of possible aggregation choices.

8. See Callaway and Sant'Anna (2021) for a more extended discussion of weaker versions of parallel trends that

identification of the ATTs requires only that parallel trends holds with respect to one pre-treatment time period and the *post-treatment* time periods, thus allowing for "pre-treatment" violations of parallel trends. However, as noted in Borusyak, Jaravel, and Spiess (2021), it is difficult to justify such an assumption ex-ante. Moreover, this stricter version of parallel trends directly justifies the task of testing for pre-trends since it restricts all pre-treatment group-time ATTs to be zero when combined with the no anticipation assumption.

Just as the group-time ATT is the building block of any causal estimand in the staggered adoption setting, the non-parametric 2×2 difference-in-differences comparison is the building block of *any estimator* that is motivated by a parallel trends assumption. Denote the sample average of the observed outcome at time t among timing group g as $\bar{Y}_{g,t} = \frac{1}{N_g} \sum_{i:G_i=g} Y_{it}$. We define the nonparametric 2×2 estimator, $\hat{\tau}_{gt}(g', t')$ as the difference in the difference in outcome means among groups g and g' at two distinct time periods: t and t'.

Definition 2 *2×2 Difference-in-Differences Estimator*

$$\begin{aligned}\hat{\tau}_{gt}(g', t') &= \left(\frac{1}{N_g} \sum_{i:G_i=g} Y_{it} - \frac{1}{N'_g} \sum_{i:G_i=g'} Y_{it} \right) - \left(\frac{1}{N_g} \sum_{i:G_i=g} Y_{it'} - \frac{1}{N'_g} \sum_{i:G_i=g'} Y_{it'} \right) \\ &= \bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}\end{aligned}$$

Under SUTVA and parallel trends, the 2×2 differences-in-differences estimator identifies a combination of four group-time ATTs (Proposition 1).

Proposition 1 *Under Assumptions 1 and 2*

$$E[\hat{\tau}_{gt}(g', t')|\mathcal{D}] = ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t')$$

Although parallel trends allows us to assign a *causal* interpretation to the difference-in-differences estimator, it does not suffice to identify one causal effect by itself. We can eliminate some of the treatment effects by considering differences-in-differences with respect to the "never-treated" groups $g' = \infty$ since $ATT_\infty(t) = ATT_\infty(t') = 0$ by definition. However, even in this case, the difference-in-differences estimator reduces to a difference between two treatment effects: $ATT_g(t)$ and $ATT_g(t')$

still permit identification of the group-time ATTs.

To actually identify a causal parameter, we need to impose an additional assumption: "no anticipation." This assumption states that units' pre-treatment potential outcomes are equivalent to their potential outcomes had they never initiated treatment. Future treatments cannot affect past outcomes and units cannot "anticipate" receipt of treatment and alter outcomes in pre-adoption periods.

Assumption 3 *No anticipation*

$$Y_{it}(g) = Y_{it}(\infty) \quad \textit{if } t < g$$

No anticipation guarantees that $ATT_g(t) = 0$ for any $t < g$. Therefore, we can construct a "clean" 2×2 difference-in-differences estimator that identifies a single treatment effect, $ATT_g(t)$, simply by selecting $g' = \infty$ and any $t' < g$. In the absence of staggering, $g' = \infty$ is the only possible $g' \neq g$ for any treatment effect of interest. However, under staggering, choosing *any* treatment timing group with $g' > t$ will yield a "clean" 2×2 comparison. Multiple clean 2×2 comparisons can be averaged to yield an estimator for a single $ATT_g(t)$ or an average of $ATT_g(t)$. Indeed, the careful choice of these comparisons underlies the "heterogeneity-robust" estimators which we discuss in Section 3.

2.2 Static TWFE

We now turn to estimation of treatment effects using fixed effects regression models and characterize the connection between these estimators and the non-parametric 2×2. The classic approach to effect estimation under the difference-in-differences assumptions relies on an ordinary least squares regression that includes "two-way fixed effects"—a set of unit-specific and a set of period-specific intercepts. There are two general approaches to parameterizing the treatment effect within the TWFE regression—a "static" specification where an indicator for whether a unit is under treatment at a given time period, D_{it} is included as a regressor and a single treatment effect parameter is estimated and a "dynamic" specification which includes multiple indicators for the relative time before/after treatment initiation. In this section, we first review and decompose the static 2×2 regression and express the estimand in terms of its component 2×2 differences-in-differences.

The static TWFE regression specification takes the form:

$$Y_{it} = \tau D_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

where α_i denotes the unit fixed effect parameter for unit i, γ_t denotes the time fixed effect parameter for time t, and ε_{it} is a mean-zero error term.

When does $\hat{\tau}$ identify a (weighted) average of group-time ATTs? To answer this, we can re-write $\hat{\tau}$ in terms of an average over 2×2 difference-in-differences comparisons. Goodman-Bacon (2021) provides one such decomposition in terms of differences across units in pre- and post-treatment averages. Our decomposition takes on a similar form, but defines the component 2×2 comparisons between two particular time periods as in Definition 2 rather than averages over entire pre- and post-periods for each unit. This allows us to better connect the decomposition to the group-time ATT estimand from Callaway and Sant'Anna (2021) and reveals an alternative constant effects assumption under which the static TWFE is valid for a convex average of ATTs.

We start our proof by applying the Frisch-Waugh-Lovell (FWL) theorem to obtain an expression for the OLS estimator $\hat{\tau}$. Define the one-way and two-way means of the treatment indicator D_{it} as:

$$\bar{D}_t = \frac{1}{N}\sum_{i=1}^{N} D_{it} \qquad \bar{D}_i = \frac{1}{T}\sum_{t=1}^{T} D_{it} \qquad \bar{\bar{D}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}$$

We use FWL to re-write the two-way fixed effects regression as a bivariate regression of the outcome on the "double-demeaned" treatment, leveraging the connection between the two-way fixed effects estimator and the two-way Mundlak (1978) estimator (Wooldridge 2021). Re-arranging the components of the sum in the numerator of this bivariate regression coefficient yields Lemma 1.1, which states that the OLS coefficient can be expressed as a uniform average over all difference-in-differences comparisons matching a treated unit i at time t to another unit i' and another time period t'.

Lemma 1.1 *The static TWFE OLS estimator $\hat{\tau}$ can be written as:*

$$\hat{\tau} = \frac{\sum_{it}\sum_{i't'} D_{it}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big)}{NT\sum_{it}\left(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}}\right)^2}$$

Simplifying this expression and re-writing further in terms of sums over treatment timing groups g and g' reveals that there are three categories of 2×2 difference-in-differences comparisons included it in this average as many component difference-in-differences terms cancel with one another. Proposition 2 states that $\hat{\tau}$ can be written as a convex weighted average over three sets of 2×2 non-parametric differences-in-differences.

Proposition 2 *The static TWFE OLS estimator $\hat{\tau}$ can be written as:*

$$\hat{\tau} = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\sum_{t'=1}^{g-1} N_g N_\infty \overbrace{\left(\bar{Y}_{g,t} - \bar{Y}_{\infty,t} - \bar{Y}_{g,t'} + \bar{Y}_{\infty,t'}\right)}^{(1)} + \sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=1}^{g-1} N_g N_{g'} \overbrace{\left(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}\right)}^{(2)}}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=g'}^{T} N_g N_{g'} \overbrace{\left(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}\right)}^{(3)}}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

The first (1) set of 2×2s involves comparisons between one treatment timing group and the never-treated units across each combination of a post-treatment time period t matched to a pre-treatment time period t'. The second (2) set involves comparisons between one timing group and another timing group that initiates treatment at a later period where each time period t after the earlier timing group's start time but before the later timing group's start time is matched to each t'. The third (3) set has the same pair of early/late treatment timing groups and the same set of time periods t but these are instead matched to t' *after* the later timing group's start time (where both pairs of units are under treatment). Figure 2 provides an example of each of these types of 2×2s in the setting with $T = 3$.

In the Appendix, we show that the weights on these comparisons sum to 1. It is straightforward

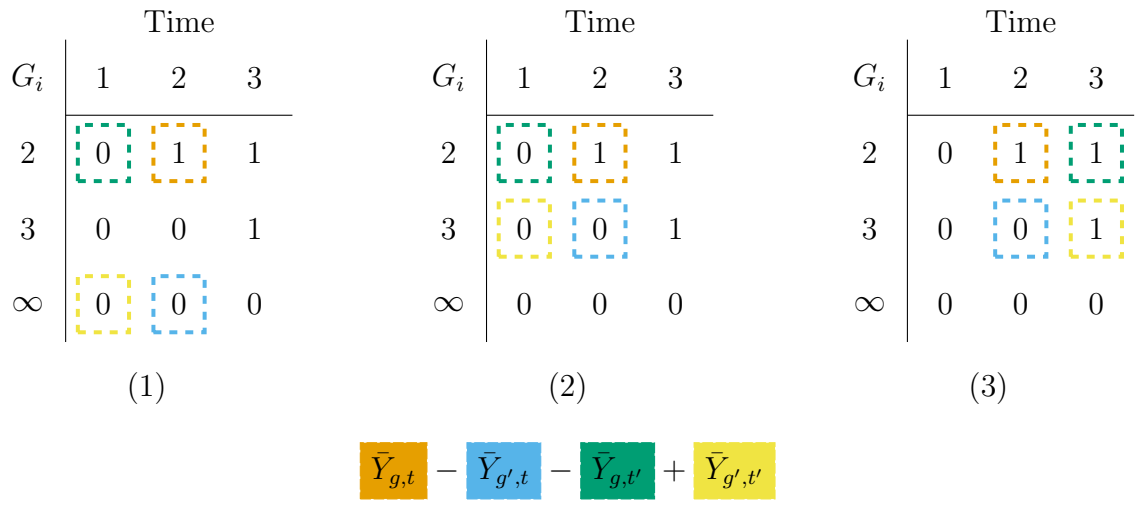

Figure 2: Types of 2×2 comparisons included in the static TWFE regression

to see that when treatment is not staggered, only the first category of 2×2 difference-in-differences will appear in the static TWFE expression. Since t' in each of these 2×2s is prior to treatment initiation, each DiD identifies a single group-time ATT under parallel trends and no anticipation. The static TWFE regression coefficient can be interpreted as a uniform average over the 2×2 difference-in-differences for each time period and, when there are only two time periods, the decomposition reduces to a single DiD term.

While the equivalence between the static TWFE regression and the non-parametric DiD is well-known in the non-staggered setting, this equivalence breaks down when we introduce staggered adoption. The terms in group (1) and group (2) will identify a single group-time average treatment effect on the treated, $ATT_g(t)$, under the difference-in-differences assumptions alone, but the terms in group (3) identify a combination of three treatment effects: $ATT_g(t) - ATT_g(t') + ATT_{g'}(t')$. Despite the fact that the static TWFE can still be written as a convex average over 2×2 difference-in-differences comparisons, this does not guarantee that the quantity that it identifies is a convex average over treatment effects. The "forbidden comparisons" (to use the terminology of Borusyak, Jaravel, and Spiess (2021)) lead to the "negative weights" problem as described in De Chaisemartin and d'Haultfoeuille (2020).

These comparisons fail to identify just one group-time ATT because the time period t' is *post-treatment* to both timing groups. Therefore, "no anticipation" does not restrict these treatment effects to be zero as it does for *pre-treatment* time periods. In order for these terms to identify

a single ATT parameter—and for the static TWFE to identify a convex average of group-time ATTs—we need to make an additional effect homogeneity assumption. $ATT_g(t')$ must equal either $ATT_g(t)$ or $ATT_{g'}(t')$ such that two of the effects in these terms cancel out.

Assumption 4 *Homogeneous within-unit effects*

$$ATT_g(t) = ATT_g(t')$$

for all $t \geq g$, $t' \geq g$

The first effect homogeneity assumption (Assumption 4), which was considered in Goodman-Bacon (2021), states that effects do not differ over time for units in the same treatment timing group. Under this assumption, $ATT_g(t')$ cancels with $ATT_g(t)$ and each "forbidden comparison" identifies $ATT_{g'}(t')$ or the effect of adopting treatment "later" in the later time period.

However, this is not the only effect homogeneity assumption that permits identification.

Assumption 5 *Homogeneous calendar time effects*

$$ATT_g(t) = ATT_{g'}(t)$$

for all $t \geq g$, $t \geq g'$

An alternative assumption (Assumption 5) that effects are homogeneous in calendar time has the "forbidden comparisons" instead identify the effect of adopting treatment "earlier" in the earlier time period t. Under this assumption, $ATT_g(t')$ would instead cancel with $ATT_{g'}(t')$. Intuitively, this assumption permits effects to vary over time but not in the *relative* time since treatment. In the presence of time-varying effect heterogeneity, Assumption 4 may not be justifiable, but Assumption 5 may still hold if the effect heterogeneity is driven by factors unrelated to the roll-out of treatment.

Proposition 3 characterizes the estimand identified by the static TWFE under one of these two constant effects assumptions. In both cases, the static TWFE recovers a variance-weighted average of the group-time ATTs. However, because of differences in *which* group-time ATT is identified by the "forbidden comparisons," the weights assigned to each effect differ slightly.

Proposition 3 *Under Assumptions 1, 2, 3, and 4, the static TWFE identifies:*

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\left(N_g(g-1)\left[N_\infty + \sum_{g'>g} N_{g'}\mathbf{1}(t<g')\right] + N_g\left[\sum_{g'<g}(g-g')N_{g'}\right]\right) \times \left(ATT_g(t)\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

Under Assumption 5 instead of Assumption 4

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\left(N_g N_\infty (g-1) + \sum_{g'>g} N_g N_{g'}(T-g'+g)\mathbf{1}(t<g')\right) \times \left(ATT_g(t)\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

Again, these weights are convex and sum to 1. Our decomposition of the static TWFE captures many of the same phenomena identified by Goodman-Bacon (2021)—notably the implicit "variance-weighting" of treatment effects and the importance of effect homogeneity assumptions. What our version clarifies is precisely which effect homogeneity assumptions are necessary. We show that there is a form of effect heterogeneity in time under which the static TWFE will still identify a convex average of ATTs, though not the *same* convex average under a within-unit effect homogeneity assumption.

Lastly, our decomposition of the static TWFE provides researchers with an additional means of quantifying the potential sensitivity of their static TWFE regression to violations of effect homogeneity. This comes from recognizing that the denominator of Proposition 2 can be interpreted as the total number of 2×2 comparisons that enter into the estimator. The share of "forbidden comparisons", π, can be written as:

$$\pi = \frac{\sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T+1-g')}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T+1-g')}$$

Computing π for a given distribution of treatment assignment $\mathcal{D}$ allows researchers to gauge the magnitude of "staggering" in their particular staggered adoption design.

2.3 Dynamic TWFE

In practice, both of the two effect homogeneity assumptions that would justify the static TWFE are often implausible. Some interventions manifest their full effect over many time periods. Other interventions that have an instantaneous impact nevertheless diminish over time as the treated and control outcomes eventually converge. Both of these patterns of treatment effects would be ruled out by Assumption 4 or Assumption 5.

To address this problem, researchers often turn to the "dynamic" two-way fixed effects specification to estimate separate effects for each time period following treatment initiation. This specification also has the added benefit of directly incorporating placebo tests for the parallel trends assumption, making it standard practice for many applied researchers using difference-in-differences methods. Instead of a single treatment indicator, this regression includes a set of relative treatment time indicators for each unit that initiates treatment at some point in time.[9] Denote the relative treatment time indicators for relative time q by $D_{it}^{(q)}$, which takes on a value of 1 if $t - G_i = q$ and 0 otherwise. In this formulation, $q = 0$ denotes the first time period that a unit is under treatment and $q = -1$ the last time period prior to treatment initiation. Let $\mathcal{Q} = \{-(T+1), \ldots, -1, 0, 1, \ldots, T-1\}$ denote the set of all possible relative treatment times.[10]

In the dynamic specification, researchers are typically not interested in an average over all ATTs, but rather an average over a *subset* of group-time ATTs that share a common relative-time. We define the relative-time ATT (RTT) as a weighted average over all $ATT_g(t)$ where t is q periods from the initiation of treatment g.

Definition 3 *Relative-time Average Treatment Effect on the Treated*

$$RTT(q) = \sum_{t=1}^{T} \sum_{\substack{g \in \mathcal{G} \\ g: g = t - q}} w_{gt} ATT_g(t)$$

where w_{gt} denotes a set of weights over group g and time t that sum to 1.

While the choice of which weights to place on each group-time ATT is ultimately up to the researcher, this choice is only meaningful when considering "heterogeneity-robust" estimators that

9. These are sometimes referred to as "leads" and "lags" of the treatment.

10. Note that under the restriction that no unit has $G_i = 1$, there are a total of $2T - 1$ possible relative treatment time periods that could be observed.

directly estimate each component $ATT_g(t)$. As we will show, the dynamic TWFE regression only identifies the $RTT(q)$ under the assumption that all of the component $ATT_g(t)$ are equivalent, making the weights irrelevant.

The conventional dynamic two-way fixed effects regression specification retains the two-way fixed effect structure of the static regression but incorporates a separate coefficient for each of the relative treatment time indicators.

$$Y_{it} = \sum_{q \neq -1} \tau^{(q)} D_{it}^{(q)} + \alpha_i + \gamma_t + \varepsilon_{it}$$

where α_i denotes the unit fixed effect, γ_t denotes the time fixed effect, and ε_{it} is a mean-zero error term.

Note that this specification is what is termed "fully dynamic" by Sun and Abraham (2021) as it includes the complete set of relative treatment time indicators, omitting a single baseline period, -1.[11] Omitting at least one period is required to avoid multi-collinearity. Often event study specifications will omit additional periods or bin together certain treatment indicators. As discussed in Sun and Abraham (2021), this can induce additional biases if omitted periods are *post-treatment* or effects are not homogeneous in the binned periods. Under no anticipation and parallel trends, $RTT(q) = 0$ if $q < 0$, motivating the use of estimates of these quantities as "placebo" or "falsification" tests.

Can the OLS estimator for some period of interest, denoted q', also be written as an average over 2×2 differences-in-differences terms? We show that this estimator *can* be decomposed into six categories of 2×2 comparisons. Only one of these categories identifies a single "uncontaminated" group-time ATT while the remainder identify some combination of multiple treatment effects, including those that are associated with *irrelevant* relative time periods. Additionally, unlike the static TWFE, the estimator is *not* a convex average over these 2×2 terms. Even when researchers make the necessary effect homogeneity assumptions, not all of the component 2×2s will identify a single group-time ATT.

Our results provide new intuition for the "contamination bias" result of Sun and Abraham (2021), which shows through a decomposition of the population regression coefficient that each

11. Typical practice is to omit the period immediately prior to the onset of treatment. Here, this is period -1, but other expositions refer to this as period 0 as in Roth et al. (2023).

coefficient $\hat{\tau}^{(q')}$ is "contaminated" by the presence of treatment effects for other relative time periods $q \neq q'$. If relative-time treatment effects are not homogeneous across timing groups, this will result in biased estimates of the relative-time treatment effects, making both post-treatment effect estimates and pre-trends tests potentially invalid. The dynamic TWFE regression may, therefore, fail to detect a pre-trend when one exists or improperly attribute a pre-trend to effect heterogeneity.

We recover this result through a different proof strategy, decomposing the *sample* dynamic TWFE coefficient into an average over the component 2×2 differences-in-differences terms. This expression allows us to illustrate how contamination bias is a consequence of the dynamic regression identifying the treatment effect of the relative time indicator $D_{it}^{(q)}$ not only through differences-in-differences comparisons with the "never treated" units but also through comparisons with units that initiate treatment at a different time period. Because the relative time (q) and the "calendar time" (t) do not perfectly align, this latter set of 2×2 terms capture the differences between two different sets of group-time ATTs and by extension, different sets of relative treatment time effects. Even under an appropriate effect homogeneity assumption, the dynamic TWFE cannot be interpreted as a convex average over non-parametric differences-in-differences terms. Rather, the intuition behind how the dynamic TWFE uses its component 2×2 comparisons is very similar to the "triple differences" estimator (Olden and Møen 2022). For each difference-in-differences term that identifies an ATT of interest that is also "contaminated" by irrelevant ATTs, the dynamic TWFE in effect subtracts off *additional* 2×2 terms that exclusively identify some part of the contamination. Just as some treated units act as "controls" in the "forbidden comparisons" of the static TWFE under staggered adoption, some DiD terms in the dynamic TWFE receive *negative* weights because the treatment effect of interest is identified only by (multiple) differences in differences-in-differences.

We start our proof by again applying the Frisch-Waugh-Lovell (FWL) theorem to obtain an expression for the OLS estimator of the relative time effect for a particular period of interest q' in the form of a bivariate regression coefficient. First, define some preliminaries. Let $\mathcal{Q}_g$ denote the set of relative time periods associated with treatment timing group g. For example, if there are $T = 4$ time periods, then for the group starting treatment in the third time period, $G_i = 3$ and $\mathcal{Q}_3 = \{-2, -1, 0, 1\}$. We use the notation $\mathcal{Q}_{G_i}$ to denote the set of relative treatment periods

associated with observation i. Note that in the absence of staggered adoption, $\mathcal{Q}_g = \mathcal{Q}, \forall g \neq \infty$. Conversely, let $\mathcal{G}_q$ denote the set of timing groups for which the relative time indicator q exists. For example, for $T = 4$, $\mathcal{G}_{-2} = \{3, 4\}$.

Next, as with D_{it}, we define the one- and two-way averages of the relative treatment time indicator $D_{it}^{(q)}$:

$$\bar{D}_t^{(q)} = \frac{1}{N}\sum_{i=1}^{N} D_{it}^{(q)} \qquad \bar{D}_i^{(q)} = \frac{1}{T}\sum_{t=1}^{T} D_{it}^{(q)} \qquad \bar{\bar{D}}^{(q)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}^{(q)}$$

Let q' be the relative treatment time of interest and $\hat{\tau}^{(q')}$ the OLS estimator of the coefficient on the relative treatment time indicator $D_{it}^{(q')}$. Denote sums over all units i and time periods t by $\sum_{it}$ and sums over all relative time periods q by $\sum_q$. Our first result mirrors Lemma 1.1 but for the dynamic TWFE case

Lemma 3.1 *The dynamic TWFE OLS estimator $\hat{\tau}^{(q')}$ can be written as:*

$$\hat{\tau}^{(q')} = \frac{\sum_{it}\sum_{i't'} D_{it}^{(q')}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right) - \sum_{q \notin \{q',-1\}} \omega_q \sum_{it}\sum_{i't'} D_{it}^{(q)}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)}{NT\sum_{it}\left((D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')}) - \sum_{q \notin \{q',-1\}} \omega_q\left(D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)}\right)\right)^2}$$

where ω_q is the coefficient on $D_{it}^{(q)}$ from a two-way fixed effects regression of $D_{it}^{(q')}$ on all other indicators $D_{it}^{(q)}$, $q \neq q', q \neq -1$

See Appendix B.5 for the proof. Lemma 3.1 provides an initial intuition for the identifying variation used by the dynamic regression estimator. As in the static specification, we can write the regression estimator in terms of averages over 2×2 comparisons between different units i and i' across different time periods t and t'. These 2×2 difference-in-differences terms: $Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}$ identify some combination of treatment effects under the parallel trends assumption as the bias in the cross-sectional comparisons is assumed to be constant across time. However, unlike the conventional static specification, the numerator consists of *differences* between the 2×2 DiD terms associated with the indicator $D_{it}^{q'}$ and terms associated with all other included indicators $D_{it}^{(q)}$. As in Sun and Abraham (2021), the weights ω_q from the auxiliary regression of $D_{it}^{q'}$ on the other indicators enter into the decomposition.

We prove a few additional lemmas on these regression coefficients ω_q for $q \notin \{q', -1\}$ that will be useful later on. We will define $\omega_{q'} = -1$ and $\omega_{-1} = 0$ for notational simplicity (as these are the implied definitions that arise from our decomposition). Applying FWL again, Lemma 3.2 shows that we can write any of the auxiliary coefficients ω_q as a sum over every other $\omega_{q^*} \neq q$

Lemma 3.2 *For any $q \notin \{q', -1\}$, defining $\omega_{q'} = -1$, $\omega_{-1} = 0$*

$$\omega_q = -\sum_{q^* \neq q} \omega_{q^*} \frac{\sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q)}}{\sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)}}$$

Equivalently

$$\omega_q = -\sum_{q^* \neq q} \omega_{q^*} \frac{\left(\sum_{g \in \mathcal{G}_q} N_g\right)\left(\sum_{g \in \mathcal{G}_{q^*}} N_g\right) - T\sum_{g \in \mathcal{G}_q} N_g N_{g+q-q^*} - N\sum_{g \in G_q \cap G_{q^*}} N_g}{(T-1)\left[\sum_{g \in \mathcal{G}_q} N_g(N - N_g)\right] + \left(\sum_{g \in \mathcal{G}_q} N_g\right)^2 - \sum_{g \in \mathcal{G}_q} N_g^2}$$

See Appendix B.6 for the full proof.

We also obtain an expression for the denominator in terms of the number of time periods, the relative sizes of each treatment timing group, and the auxiliary regression weights.

Lemma 3.3 *The denominator of $\hat{\tau}^{(q)}$ can be expressed as:*

$$\begin{aligned} NT\sum_{it}\left(\tilde{D}_{it}^{(q')} - \sum_{q \notin \{q', -1\}} \omega_q \tilde{D}_{it}^{(q)}\right)^2 &= NT\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q \notin \{q', -1\}} \omega_q \sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q')}\right] \\ &= \left\{(T-1)\left[\sum_{g \in \mathcal{G}_{q'}} N_g(N - N_g)\right] + \left(\sum_{g \in \mathcal{G}_{q'}} N_g\right)^2 - \sum_{g \in \mathcal{G}_{q'}} N_g^2\right\} - \\ &\quad \sum_{q \notin \{q', -1\}} \omega_q \left\{\left(\sum_{g \in \mathcal{G}_q} N_g\right)\left(\sum_{g \in \mathcal{G}_{q'}} N_g\right) - T\sum_{g \in \mathcal{G}_q} N_g N_{g+q-q'} - N\sum_{g \in \mathcal{G}_q, \mathcal{G}_{q'}} N_g\right\} \end{aligned}$$

See Appendix B.7 for the full proof.

We then proceed to re-arrange the numerator into a weighted sum over the component 2×2 differences-in-differences terms, rewriting the sum over i as a sum over treatment timing groups g. We leverage the fact that $D_{it}^{(q)}$ only takes on a value of 1 for a single time period for a given

unit i. Define $\Omega_g = \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \omega_q$ as the sum of the auxiliary regression weights for all relative time periods q that pertain to timing group g. Let $\tilde{D}_{it}^{(q)} = D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)}$ denote the double de-meaned relative treatment time indicator.

Proposition 4 *The dynamic TWFE OLS estimator $\hat{\tau}^{(q')}$ can be written as a sum over :*

$$
\begin{aligned}
\hat{\tau}^{(q')} = \Bigg[& \sum_{g \in \mathcal{G}_{q'}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g \Big) \overbrace{\Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big)}^{(1)} \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g \Big) \overbrace{\Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big)}^{(2)} \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g + T\omega_q \Big) \overbrace{\Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{\infty,g+q} \Big)}^{(3)} \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g + T\omega_q \Big) \overbrace{\Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{g',g+q} \Big)}^{(4)} \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(N_g N_\infty \Big) \Big(\Omega_g - T\omega_q \Big) \overbrace{\Big(\bar{Y}_{g,g+q} - \bar{Y}_{\infty,g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big)}^{(5)} \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(\Omega_g - T\omega_q \Big) \overbrace{\Big(\bar{Y}_{g,g+q} - \bar{Y}_{g',g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big)}^{(6)} \Bigg] \times \\
& \Bigg[(T-1) \Bigg[\sum_{g \in \mathcal{G}_{q'}} N_g (N - N_g) \Bigg] + \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g \Bigg)^2 - \sum_{g \in \mathcal{G}_{q'}} N_g^2 \\
& - \sum_{q \notin \{q', -1\}} \omega_q \Bigg\{ \Bigg(\sum_{g \in \mathcal{G}_q} N_g \Bigg) \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g \Bigg) - T \sum_{g \in \mathcal{G}_q} N_g N_{g+q-q'} - N \sum_{g \in \mathcal{G}_q \cap \mathcal{G}_{q'}} N_g \Bigg\} \Bigg]^{-1}
\end{aligned}
$$

See Appendix B.8 for the full proof. The proposition above represents the regression estimator for relative period q' as a sum over 2×2 non-parametric difference-in-differences terms. It is similar in form to Proposition 1 in Sun and Abraham (2021) with the primary difference being the explicit

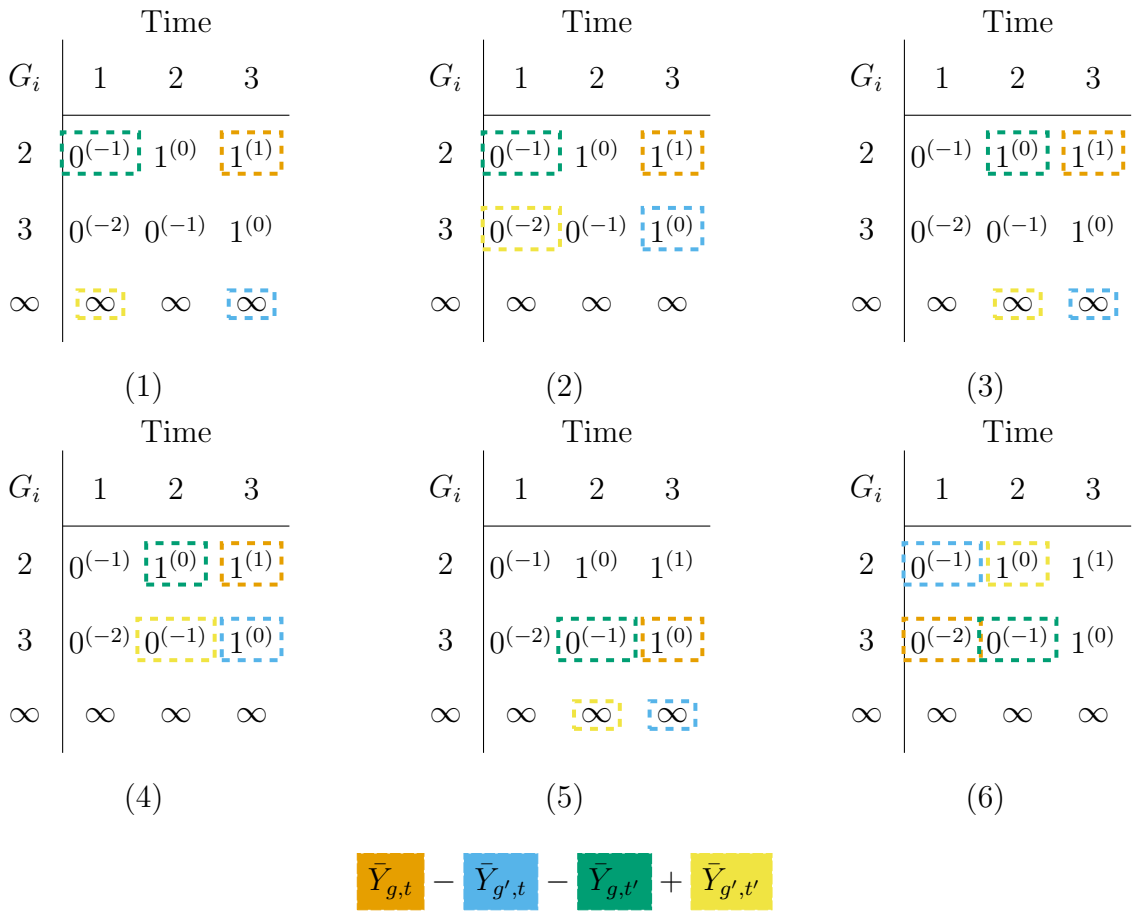

Figure 3: 2×2 comparisons included in the dynamic TWFE regression for $q' = 1$

characterization of the matched "baseline" period and unit in each 2×2. We illustrate the six types of 2×2 comparisons in Figure 3 in our simple $T = 3$ setting for relative-treatment-time period 1. Note that despite the absence of any observations with that relative-treatment-time in treatment timing group $G_i = 3$, that timing group still enters into the dynamic TWFE coefficient through multiple DiD terms.

The first (1) set of 2×2s are the only set of comparisons that always identify a single group-time ATT that pertains to the relative treatment time of interest q' since they involve matching each timing group g that contains q' to the never-treated units and taking the difference-in-differences between time period $t = g + q'$ – the relative-time period of interest – with time period $t' = g - 1$ – the held-out baseline. All other comparisons are potentially contaminated by other non-zero treatment effects. Set two (2) also involves a comparison across the time period of interest $t = g+q'$ and the baseline $t' = g - 1$, but timing group g is matched not to the never-treateds, but rather

another timing group $g' \neq g$. Even under no-anticipation with respect to the baseline -1, these terms will identify the combination of three effects: $ATT_g(g+q') - ATT_{g'}(g+q') + ATT_{g'}(g-1)$. Contamination for terms in groups three (3) and four (4) arises further from matching the period of interest $t = g+q'$ to a non-held-out but irrelevant baseline period q. Lastly, the terms in groups five (5) and six (6) include all differences-in-differences with entirely irrelevant timing groups—those that are not in the set of timing groups that contain relative treatment time q'. While terms in groups 1-4 at least identify *one* group-time ATT relevant to $RTT(q')$, these comparisons do not identify a single relevant ATT!

Given this representation of the estimator, we can show that the OLS estimator of $\tau^{(q)}$ identifies the relative treatment time effect $RTT(q')$ only under two conditions. The first is when there is no staggered adoption and there exists only one timing group that receives treatment at g^*. In this case, there is a single group-time ATT that corresponds to $RTT(q') = ATT_{g^*}(g^*+q')$. We can show that the regression coefficient is equivalent to a simple, non-parametric 2×2 differences-in-differences estimator between time period g^*+q' and the baseline period g^*-1.

While this result is well-known, we show that we can obtain it directly from Proposition 4 using our results on the auxiliary regression coefficients. In particular, we show that in the absence of staggering $\omega_q = -\frac{1}{2}$ for all of the irrelevant relative time periods.

Lemma 4.1 *Under a common treatment initiation time g^*, $G_i \in \{g^*, \infty\} \forall i$, $\omega_q = -\frac{1}{2} \forall q \in \mathcal{Q}, q \notin \{q', -1\}$*

See Appendix B.9 for the proof. When treatment is not staggered, 4 also only contains DiD terms in category (1) and category (3) since treatment timing group g^* can only be matched to the never-treated units (eliminating groups (2) and (4)) and there are no timing groups that do not contain the relative treatment time of interest (eliminating (5) and (6)). When $\omega_q = -\frac{1}{2}$, the weights on each of the comparisons in (3) are zero since $1 - \Omega_g = -T\omega_q = \frac{T}{2}$ leaving only the comparisons in (1), which identify a single group-time ATT without any additional constant effects assumptions. Therefore, the coefficient on $D_{it}^{(q')}$ reduces to a single non-parametric difference-in-difference between the relative period q' and the baseline period -1. Under our standard difference-in-difference assumptions, this non-parametrically identifies the group-time $ATT_{g^*}(g^*+q')$ (Proposition 1), which, in the absence of any other treatment timing groups, is $RTT(q')$, the relative treatment time effect of interest.

Proposition 5 *Under a common treatment initiation time g^*, $G_i \in \{g^*, \infty\} \forall i$, the OLS estimator $\hat{\tau}^{(q')}$ is equivalent to:*

$$\hat{\tau}^{(q')} = \left[\frac{1}{N_{g^*}} \sum_{i:G_i=g^*} Y_{i,g^*+q'} - \frac{1}{N_\infty} \sum_{i:G_i=\infty} Y_{i,g^*+q'}\right] - \left[\frac{1}{N_{g^*}} \sum_{i:G_i=g^*} Y_{i,g^*-1} - \frac{1}{N_\infty} \sum_{i:G_i=\infty} Y_{i,g^*-1}\right]$$

See Appendix B.10 for the proof.

Under staggered adoption, however, the dynamic TWFE only identifies a relative treatment time effect under the effect homogeneity assumption in Sun and Abraham (2021): that the group-time ATTs are equivalent if they correspond to the same relative time ATT.

Assumption 6 *Homogeneous relative-time ATT*

For all $g \neq g^*$, $t \neq t^*$, such that $q = t - g = t^* - g^*$,

$$ATT_g(t) = ATT_{g^*}(t^*) = RTT(q)$$

While this assumption allows effects to vary over time, it imposes some additional restrictions on treatment effects that are not implied by either Assumptions 4 or 5. Therefore, while it may be more plausible in some settings, it is not a *weaker* assumption compared to those required for the static TWFE to identify a convex average of ATTs. Relative-time ATT homogeneity, for example, rules out the possibility that calendar time might also modify the treatment effects. This may be implausible in many political science settings. For example, when looking at election turnout, the impact of a treatment on turnout may depend on whether the first election after the intervention is held in a mid-term year or a presidential year.

With this assumption, and taking the expectation of our expression for $\hat{\tau}^{(q')}$ in Proposition 4 conditional on the design $\mathcal{D}$, we recover the main result from Sun and Abraham (2021).

Proposition 6 *Under homogeneous relative-time ATTs, no anticipation, and parallel trends with respect to relative time period -1*

$$E[\hat{\tau}^{(q')}|\mathcal{D}] = \sum_{\substack{q \in \mathcal{Q} \\ q \neq -1}} RTT(q) \frac{\left[-\omega_q \sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)} - \sum_{q^* \neq q} \omega_q \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q)}\right]}{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \notin \{q', -1\}} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q^*)}\right]}$$

The weights on each $RTT(q), q \neq q'$ are equal to zero, and the weight on $RTT(q')$ is equal to one:

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] &= RTT(q') \frac{\left[-\omega_{q'} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q'} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q')} \right]}{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q',-1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q^*)} \right]} \\
&= RTT(q') \frac{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q',-1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q')} \right]}{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q',-1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q^*)} \right]} \\
&= RTT(q')
\end{aligned}
$$

The full proof is in Appendix B.11. Crucially, the final steps for determining the weights on each relative-treatment time effect rely on our result in Lemma 3.2.

The decomposition highlights a few central features of the dynamic TWFE regression as compared to the static TWFE. The first is the importance of the choice of the omitted relative time period. Note that the only 2×2 comparisons that are *not* contaminated by other relative treatment time effects are those involving the omitted baseline period and the never-treated units. Omitting a baseline period that does not appear for all units in the sample means that the treatment effects for those units where the omitted period is absent are only identified indirectly through the contaminated ATTs. This places a much greater weight on the effect homogeneity assumption since there are no "clean" difference-in-difference comparisons that contribute to the point estimate. Second, it highlights that the mechanics behind the dynamic TWFE and the connection between it and the non-parametric difference-in-differences estimator are markedly different than the static TWFE. While the static TWFE can still be interpreted as a convex weighted average over 2×2 differences-in-differences terms, this is not true of the dynamic TWFE. Even when the effect homogeneity assumption holds (Assumption 6), the component 2×2 terms still identify combinations of treatment effects. Some 2×2 terms include *only* irrelevant effects. The dynamic TWFE regression identifies the relative-treatment time effects of interest by appropriately *weighting* these 2×2 comparisons such that the overall weight placed on irrelevant treatment effects is zero. Effectively, the dynamic TWFE regression functions as a kind of triple (or more) differences estimator under staggered adoption.

3 Heterogeneity-robust estimators

Because the conventional event study regression is biased under unrestricted heterogeneity in the group-time ATTs, a variety of alternative estimators have been proposed in the literature. These include the "interaction-weighted" estimator of Sun and Abraham (2021), the "doubly-robust" estimator of Callaway and Sant'Anna (2021), the two-way fixed effects counterfactual imputation estimator of Borusyak, Jaravel, and Spiess (2021), the "leavers" and "joiners" estimator of De Chaisemartin and d'Haultfoeuille (2020), and the local projections estimator of Dube et al. (2023) among others. Analogues of many of these estimators have been proposed in the political science literature as well. Notably, a version of the "leavers" and "joiners" estimator was developed in Imai and Kim (2021) with an additional pre-processing matching step in Imai, Kim, and Wang (2021), and a counterfactual imputation estimator using a two-way fixed effects model also appears in Liu, Wang, and Xu (2022). It is easy for applied researchers to become overwhelmed by all of the options. However, all of these methods aim at the same common goal: to construct an estimator that incorporates only "clean" 2×2 comparisons such that identification relies only on the conventional DiD assumptions and nothing else. In fact, some of these estimators nest others as special cases. Generally speaking, differences between these estimators primarily concern: 1) which cross-sectional units to include in the "clean controls", 2) which pre-treatment periods to use as the baseline comparison periods, and 3) how to incorporate covariates.

Rather than estimating the relative treatment time effect of interest in a single regression,[12] the heterogeneity-robust estimators separate this task into two steps. First, they construct an estimator for each group-time ATT associated with that particular relative-time period using only valid 2×2 difference-in-differences comparisons. Second, they aggregate the group-time ATTs according to some pre-defined weighting function to obtain a target parameter such as the relative-time treatment effect. The choice of weights to assign to each treatment effect is in some sense arbitrary, but a typical choice for the RTT is the proportion of units in the associated timing group. Options for aggregation weights are discussed in greater detail in Callaway and Sant'Anna (2021) and Sun and Abraham (2021) and we leave the question of how best to choose among different weighting approaches outside the scope of this paper.

12. Although, see Wooldridge (2021) for a discussion of how one can obtain each of these estimators through properly specified fixed effects regressions with appropriately interacted and de-meaned regressors.

Below, we briefly summarize the two categories of heterogeneity-robust estimators for the staggered adoption setting, describe how each constructs its component 2×2 difference-in-differences comparisons, and how the chosen control observations and baseline comparison periods differ across method.

3.1 Direct specification of the 2×2 DiDs

The methods described by Sun and Abraham (2021) and Callaway and Sant'Anna (2021) are built around estimating each effect of interest via a single, valid 2×2 difference-in-differences comparison. Sun and Abraham (2021) approach this by fixing the event study regression to eliminate the staggering in treatment adoption, recovering the relationship between the TWFE regression and the non-parametric difference-in-differences. By including distinct indicator terms for each treatment timing group's relative treatment times – or equivalently, interacting the relative treatment time indicators with indicators for treatment timing group – Sun and Abraham (2021) obtain separate estimates for each timing group's set of $ATT_g(t)$s. Intuitively, the fully-interacted regression is is equivalent to running separate event study regressions for all possible timing groups g that include observations from only that timing group $G_i = g$ and the never-treated units $G_i = \infty$. And as we show in Proposition 5, the dynamic TWFE regression with only a single treatment timing group and the never-treated *is* equivalent to the non-parametric difference-in-differences estimator. Estimates of the relative-time effects $RTT(q)$ can be obtained by taking a weighted average over the estimates of the relevant $ATT_g(t)$. Essentially, this approach uses only the comparisons falling under group (1) in the static TWFE decomposition (Proposition 2).

One downside to the interaction-weighted estimator is that it only leverages observations from the "never-treated" units to construct the differences-in-differences control group. This may result in very imprecise estimates of the individual group-time treatment effects if most units in the sample initiate treatment and the relative number of never-treated observations is very small. Additionally, if researchers believe parallel trends hold across all treatment-timing groups, there may be efficiency gains to using additional observations that are eventually treated but "not yet" treated in a given time period. In other words, we may want to include comparisons of type (2) from the static TWFE decomposition in addition to those in type (1) while still avoiding the "contaminated" terms in category (3).

The estimator proposed by Callaway and Sant'Anna (2021) does precisely this. Like the estimator in Sun and Abraham (2021), it is also constructed by identifying the valid control groups and comparison times associated with each $ATT_g(t)$ of interest. However, in contrast to the interaction-weighted estimator in Sun and Abraham (2021), this set of clean controls can also include those units that are not yet treated at both time t and the baseline period but are eventually treated at some time $g > t$. As in the Sun and Abraham (2021) approach, relative-time effects are obtained by aggregating the relevant $ATT_g(t)$ parameters.

One advantage of Callaway and Sant'Anna's (2021) approach is that, by reducing the task of estimating treatment effects under staggered adoption to a series of 2×2 differences-in-differences, conventional covariate-adjustment techniques developed for the two-period setting can be used to improve precision or allow for identification under conditional versions of parallel trends. The "double-robustness" of the Callaway and Sant'Anna (2021) estimator comes from the incorporation of two models: the inverse-propensity of treatment weighting adjustment of Abadie (2005) and an outcome model. If either the model for treatment adoption or the model for the outcome is correct, the DR estimator identifies the group-time ATT of interest In the absence of covariates, however, the Callaway and Sant'Anna (2021) estimator for each group-time ATT reduces to a simple 2×2 difference-in-differences with a carefully constructed control groups. If researchers restrict the control group to include only never-treated units, it is exactly equivalent to Sun and Abraham (2021).

3.2 Implicit specification through regression imputation

While the above estimators attempt to address the problems with two-way fixed effects by explicitly selecting the set of acceptable 2×2 comparisons, an alternative approach instead starts by directly modeling the potential outcomes. The "imputation" method discussed in Borusyak, Jaravel, and Spiess (2021), Liu, Wang, and Xu (2022), and Gardner (2022) begins by assuming a particular model for the potential outcomes under control: $Y_i(\infty)$. One such model is the conventional two-way fixed effects specification: $Y_{it}(\infty) = \alpha_i + \delta_t + \varepsilon_{it}$. However, more flexible forms involving covariates or latent factors can be seen as extensions of this fundamental approach. In general, the outcome model is characterized by what Liu, Wang, and Xu (2022) call a "strict exogeneity" assumption in which the probability of treatment assignment in a given time period

does not depend on the lagged outcomes. In the case of the two-way fixed effects structure, these assumptions imply the general form of the parallel trends assumption.

The imputation approach then fits the selected model to a subset of the data—units and time periods that are under control—and generates imputed counterfactuals for each treated unit-time in the sample. Estimates of the relevant group-time average treatment effects $ATT_g(t)$ and their aggregates are then constructed by averaging over differences between observed outcomes for treated units and the imputed counterfactual outcome from the model.

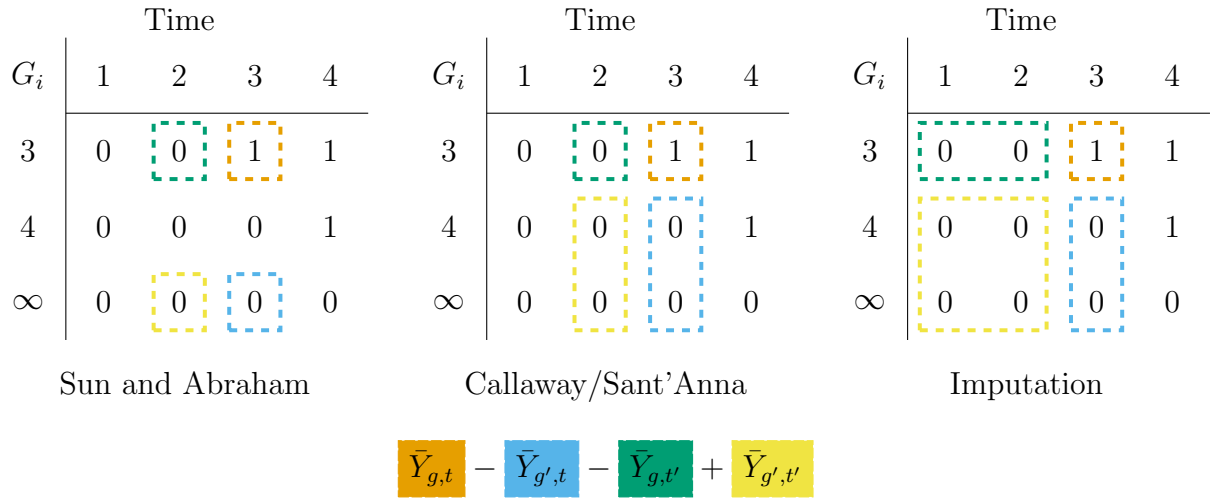

Figure 4: Visualizing the implied 2×2 comparisons in "heterogeneity-robust" DiD methods – $ATT_3(3)$

Imputation differs from the two previous methods in that parallel trends are a *consequence* of the assumptions made about the outcome model rather than an assumption made outright. As such, Chiu et al. (2023) characterize these methods as distinct from the 2×2 difference-in-differences approaches discussed above. However, they can still be described practically in terms of the implied 2×2 difference-in-differences comparisons that comprise each imputation in the case where the control model only includes unit and time fixed effects and no additional covariates. In the case of non-staggered treatment, the imputation estimator is an average over the 2×2 differences-in-differences between the period of interest and every single pre-treatment period.[13] Extension to staggered adoption is complicated by the non-uniform weights placed on each component 2×2 DiD due to regression variance weighting. However, the underlying principle remains

13. See Appendix A for a proof of this result. Also see the discussion in Roth et al. (2023).

the same – the imputation estimator is also selecting the "clean" 2×2 differences-in-differences terms albeit implicitly. Figure 4 summarizes the implied 2×2 comparisons for each estimator in a hypothetical setting with $T = 4$.

In contrast to the dynamic TWFE regression in which the pre-treatment placebo estimates are constructed in the exact same way as the post-treatment effect estimates, most of the new heterogeneity-robust methods allow researchers greater flexibility in how they estimate the pre-trends coefficients. As Roth (2024) notes, a potential downside to this flexibility is that researchers need to be more attentive to how researchers present and interpret the classic "event study plot" using these new approaches. Because the pre-trends tests do not necessarily have the same common baseline as the effect estimates, the signs and magnitudes for the placebo coefficients for each period may not have the same interpretation as they would in the dynamic TWFE case.

The crucial issue highlighted in Roth (2024) is that researchers need to be attentive to *baseline choice* when constructing placebos. For example, common implementations of the Callaway and Sant'Anna (2021) estimator allow for placebo tests to be constructed with a *varying* baseline period – in contrast to the effect estimates which are always relative to a *universal* pre-treatment period. With this approach, the pre-treatment placebos are estimated across adjacent time periods ("short diferences") while the gap in time between the baseline and period of interest for each effect estimate can be as large as the number of post-treatment periods ("long differences"). As a consequence, the magnitudes of the placebo estimates cannot be directly compared with the effect sizes as researchers would a conventional event study plot. "Short differences" may understate the magnitude of the pre-trends violations, especially if violations build slowly over the pre-treatment period.

Imputation approaches introduce an additional complication by having *multiple* baseline periods for each pre-treatment placebo. Liu, Wang, and Xu (2022) describe two approaches to constructing placebo tests for imputation. The first is to leave out observations with the relative treatment time for which the placebo is being constructed, re-estimate the two-way fixed effects model on the remaining control periods and impute the counterfactual for the "held-out" observations as though they were treated. This "leave-one-out" approach averages over both the short and long placebo differences-in-differences in the pre-treatment period. The other approach – described as a "placebo test" – involves holding out some number of sequential pre-treatment periods

and using the remainder as the control observations on which the imputation model is fit. While this limits the number of additional regressions that need to be run, it requires users to explicitly partition the pre-treatment period into "placebo" and "baseline" periods. Borusyak, Jaravel, and Spiess (2021) adopt this approach with the additional innovation of fitting the dynamic TWFE model on the control observations as opposed to using a held-out imputation approach. This takes advantage of the efficiency gains from the "indirect" comparisons of the dynamic TWFE and because the parallel trends assumption implies that *all* pre-treatment effects are zero, the effect homogeneity assumption for this subsample is guaranteed by the standard DiD identifying assumptions. However, because the baseline periods for these placebo tests are *prior* to the periods for which the placebo effects are being estimated – in contrast to the "universal baseline" approach in which the typical baseline of -1 is *after* the placebo periods – the signs of the estimates are opposite to what one would obtain from a standard dynamic TWFE event study plot (Roth 2024).

All of the above approaches provide valid placebo effect estimates, so researchers need only to be cautious in understanding what underlying 2×2 comparisons contribute to each. However, there is one implementation that is currently in use that does *not* provide unbiased placebo estimates and should not be used: "in-sample imputation". This approach is implemented as the default in the `fect` package as of version `1.0.0` with the other two placebo tests included as options to be enabled by the researcher. In this approach, rather than leaving out the pre-treatment periods for which imputation is carried out (as in the leave-one out approach), these periods are included when fitting the two-way fixed effects regression. Although this saves computational time by requiring only a single regression, it does not yield unbiased estimates of the pre-trends. We show in Appendix A that "in-sample imputation" as opposed to "leave-one out" imputation generates placebo estimates that are artificially attenuated towards zero. This is because some of the implied 2×2 comparisons incorporated in the imputation are zero by construction since they use either the same unit or time period twice. The magnitude of this bias is larger when there are fewer pre-treatment periods and fewer "never-treated" units. Simulation evidence shows considerable under-rejection of the null of no effect compared to the leave-one out approach meaning that researchers may fail to detect substantial pre-trends violations as a consequence of this attenuation bias. When using one of the new imputation approaches, researchers should ensure that any periods for which placebo

estimates are obtained are *held out* from the model fit to generate the imputed counterfactual. In-sample imputation risks over-confidence in the parallel trends and no anticipation assumptions.

To retain maximum comparability with the dynamic TWFE specification when illustrating the sensitivity of estimates to constant effects violations and to ensure that we can detect longer-running parallel trends violations, all of our replications in the subsequent section use a common, universal baseline period of -1 for all pre-treatment placebos and post-treatment effects when implementing the Callaway and Sant'Anna (2021) estimator. For our replications of Paglayan (2019) and Grumbach and Hill (2022), we construct control groups using both never-treated units and units that have not yet been treated but eventually receive treatment in order to improve efficiency. Because the data in Hall and Yoder (2022) contain a significantly large number of observations, we are able to estimate each timing group's treatment effect trajectory separately using only the never-treated units as controls.

4 Replications

In the sections above, we have highlighted a number of potential pitfalls of event study regressions that can result in invalid and biased estimates of dynamic treatment effects and pre-trends. Some of these problems exist irrespective of treatment effect homogeneity. For example, in many applications that we have surveyed, researchers inappropriately omit multiple relative time indicators from their event study regressions, including indicators associated with post-treatment relative time periods. Omitting indicators for post-treatment periods risks biasing all of the event study estimates due to contamination by treatment effects from these baselines. This issue exists even when there is no staggering as the omitted indicators in the regression define the baseline comparison periods for the implied 2×2 difference-in-differences. Additionally, omitting pre-treatment relative time indicators that are too far from the start of treatment risks placing greater weight on the effect homogeneity assumption, particularly when there are observations for which no relative time indicators are omitted. For these treatment timing groups, effects are identified indirectly, *exclusively* off of the differences in the component differences-in-differences. We recommend that, if using a dynamic DiD estimator with only a single baseline comparison time, researchers select a period that is common to all observations. One advantage of the standard convention of using

relative period $q = -1$—the last period under control—is that it will always satisfy this criterion as long as there are no always-treated units.

We also advise that researchers exercise caution in the selection of their control group. In two of our applications below, the never-treated units were discarded due to concerns about parallel trends. While in some settings, it may be the case that parallel trends only hold with respect to the units that receive treatment at some point, we show in our replications of these studies that observed pre-trends violations are actually more acute when only using sometimes-treated units as controls than when using both sometimes-treated and never-treated units. As noted in Baker, Larcker, and Wang (2022) and shown in our decomposition, settings with fewer never-treated units are also potentially more susceptible to problems arising from treatment effect heterogeneity, since comparisons across the different timing groups drive a larger share of the comparisons that enter into the regression estimator.

Lastly, we show that treatment effect heterogeneity has significant consequences for estimated pre-trends and dynamic treatment effects. While we replicate fewer papers compared to the analysis of static two-way fixed effects estimates in Chiu et al. (2023) due to the relative lack of consistent published event study plots in the political science difference-in-differences literature, our conclusions are slightly less sanguine regarding the threat of treatment effect heterogeneity for inference. While Chiu et al. (2023) tend to find that static TWFE estimates match the heterogeneity-robust "static" estimates, we think researchers should be cautious in assuming that this extends to the dynamic TWFE estimates. We reach different substantive interpretations in most of our replications—conventional event study plots and heterogeneity-robust event study plots provide contradictory conclusions about both the presence of pre-trends and the trajectory of the dynamic treatment effects. For example, Chiu et al. (2023) highlight Hall and Yoder (2022) as an example of a study where results do not change much as a consequence of the new methods. While this does appear to be the case for the static TWFE, it does not extend to the dynamic TWFE. Our replication, by contrast, shows that the published event study plots mask a significant pre-trends violation driven by one particular treatment timing group.

We speculate that part of the explanation for why static and dynamic regression estimates could vary in their sensitivity to treatment effect heterogeneity lies in the different effect homogeneity assumptions needed for each setting. While the static regression identifies (some) weighted average

of ATTs under the assumption that group-time ATTs are homogeneous in time even if these effects vary by group, the event study regression requires an assumption that group-time ATTs are homogeneous across initiation groups even though they are permitted to vary over time. If the impact of treatment is unchanged after initiation, but the magnitude of this initial shock varies by early versus late adopters, then static two-way fixed effects regressions may remain unchanged while event study regression estimates change substantially.

4.1 Paglayan (2019)

Our first replication is of Paglayan (2019), which examines the impact of collective bargaining rights for teacher unions on the size of state government in the United States. The paper concludes that, contrary to expectation, mandatory collective bargaining had no discernible effect on state investment in education, potentially due to the incorporation of anti- as well as pro-union provisions in these laws. The primary differences-in-differences design involves analyzing a longitudinal dataset that includes all states before and after they granted collective bargaining rights to teachers. Figure 5 plots the distribution of treatment and control units using the `panelview` R package (Mou, Liu, and Xu 2023). The earliest "treated" states adopted mandatory collective bargaining laws in 1965, while the last, Nebraska, adopted the policy in 1987. Notably, although the primary analysis excludes the never-treated units, the data collected extends beyond 1987 to 1997 and no state in the sample implements a collective bargaining law between 1987 and 1997.

The event study specification reported in the original paper took the form of the following dynamic event-study specification, fit on observations from 1959 to 1997 among states that implemented a collective bargaining law at some point:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{q=-6}^{10} \tau^{(q)} D_{it}^{(q)} + \varepsilon_{it} \tag{1}$$

Figure 6 replicates the exact estimates as reported in the original paper on the four measures of educational investment: student-teacher ratios, log of real teacher salaries, log of real per-pupil current expenditures, and log of real per-pupil nonwage current expenditures. In all four cases, the results suggest no significant evidence for pre-trends in the six periods prior to treatment initiation as well as no discernible short or long-term effect of collective bargaining in the 10

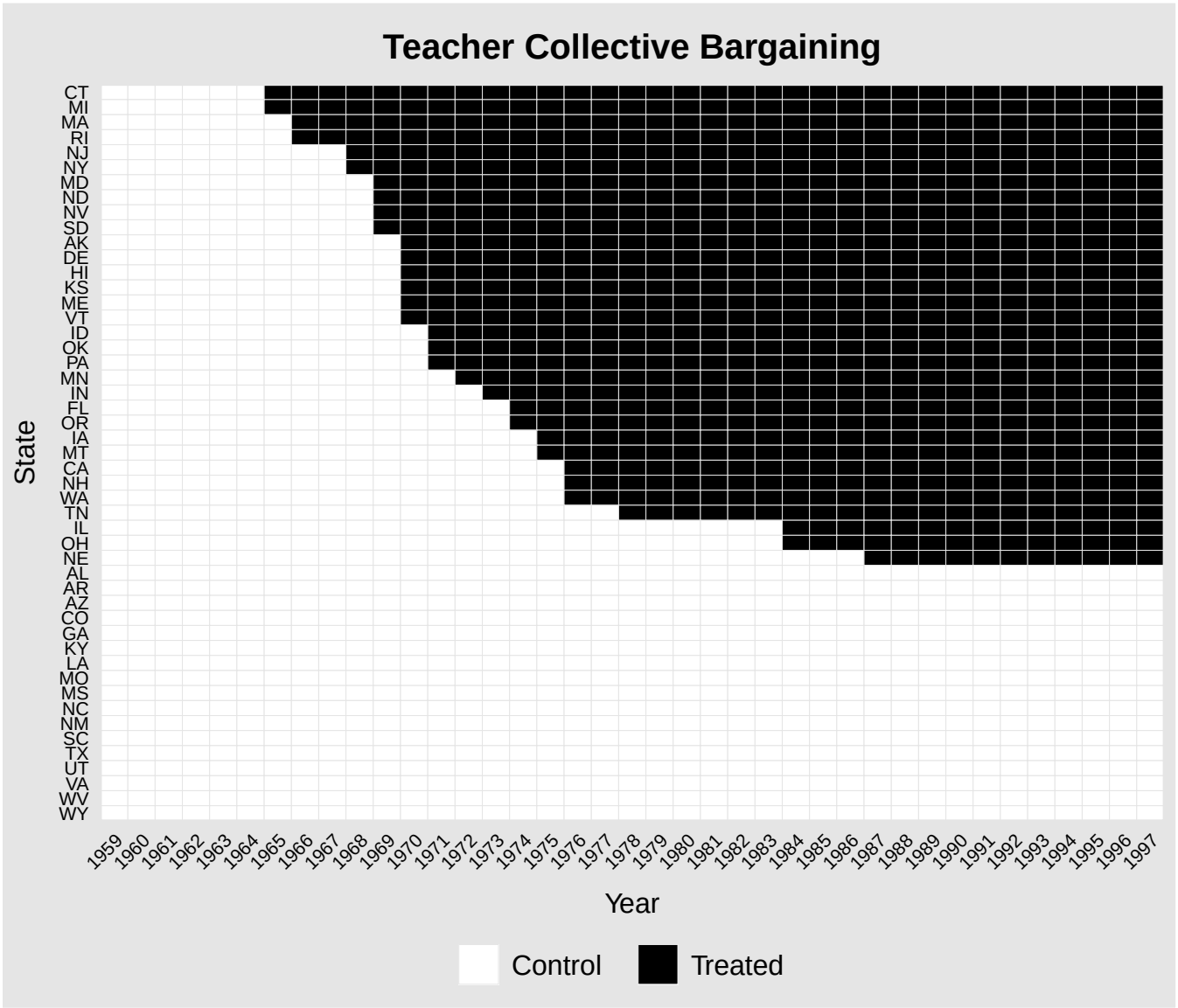

Figure 5: Distribution of treatment and control units in Paglayan (2019)

periods following implementation.

Before turning to the heterogeneity-robust estimates, we examine two problems with this specification. The first is which dummy indicators are included and excluded. Rather than fully saturate the regression with all but one relative-time indicator, the regression omits all relative time indicators prior to -6 (while still including the indicator for the period immediately prior to treatment) and omits all relative time indicators after 10 (rather than binning the post-treatment indicators as in Hall and Yoder (2022)). As shown in Figure 5, the full dataset contains units with as many as 28 potential pre-treatment "lead" periods (for Nebraska) and 32 post-treatment "lags"

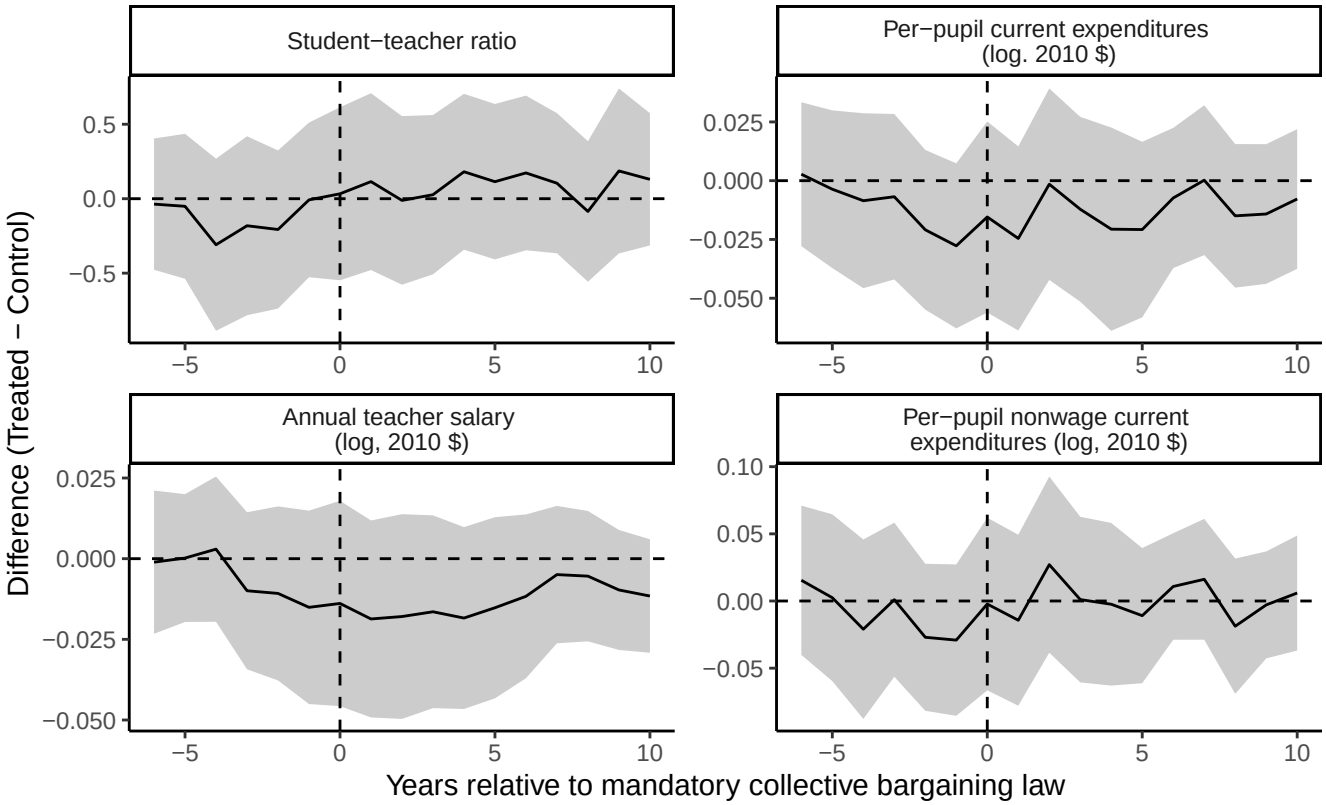

Figure 6: Exact replication of the event-study plot in Paglayan (2019).

(for Connecticut and Michigan). As illustrated in Sun and Abraham (2021) and our decomposition above, the omitted indicators define the baseline comparison times for the underlying difference-in-differences. Identification imposes the assumption that the relative time effects for these periods are *zero*. Under no anticipation and parallel trends, these periods must therefore be *pre-treatment*. Therefore, the omission of indicators for post-treatment periods could potentially bias the pre-trends and dynamic treatment effect estimates even before we consider problems arising from effect heterogeneity.

Additionally, the regression includes many time periods in which all units are treated since the never-treated units are removed but the sample is not truncated. This results in some treated observations contributing to the treatment effect estimates only through differences with other units under treatment. Intuitively, no 2×2 difference-in-differences can identify the treatment effects in these periods as no units are under control after 1987. Indeed, as noted by Borusyak, Jaravel, and Spiess (2021) and Sun and Abraham (2021), if we use the conventional fully-dynamic

specification with a single omitted baseline, it will still suffer from multi-collinearity as there are no "never-treated" units. As is increasingly standard practice, we remove time periods after 1986 from the analysis such that Nebraska, the "last treated" unit in the data becomes the "never-treated" unit in the truncated sample.[14]

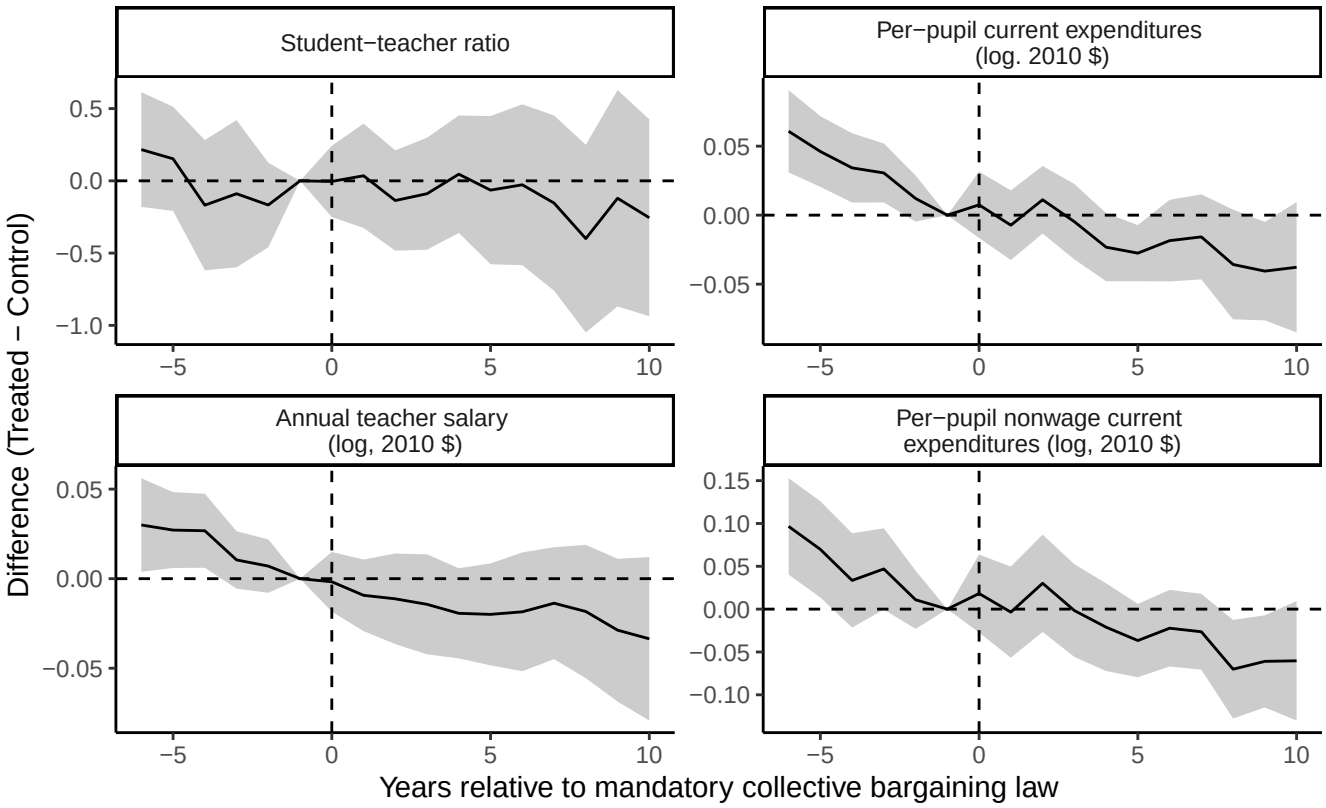

Figure 7: Fully dynamic event-study plot for Paglayan (2019). Includes only states which implemented collective bargaining. 1959-1986

We examine the results from a fully dynamic event study specification fit to the period 1959-1986, still excluding the states that never implemented a collective bargaining law. We omit relative period −1 as the "baseline." Although we estimate effects for relative time indicators beyond 10 periods post-treatment and 6 periods pre-treatment, to retain maximum visual comparability with the original results, we truncate the figures to only include the same periods as in the original paper. Figure 7 shows the results. We find that both the treatment effect and pre-trends estimates are noticeably different than those presented in the original figure. We es-

14. Note that all heterogeneity-robust estimators also drop these "all-treated" time periods from the analysis as there is no estimator for those group-time ATTs that does not impose additional homogeneity assumptions.

timate a general negative trend in the effects of collective bargaining on per-pupil expenditures (both current and non-wage current) as well as on annual teacher salary. For some post-treatment periods, these estimates are statistically significant at $p < .05$. However, the validity of these estimates is immediately called into question by the presence of clear pre-trends, suggesting that the underlying identifying assumptions for differences-in-differences may not hold.

To what extent do these results change if we instead include the never-treated units? While the paper motivates the use of only those units that implemented a collective bargaining law by arguing that conditional on ever passing such a law, treatment uptake times are largely random, this may not be the case from our initial examination of the pre-trends among these units. Moreover, Figure 3 in the original paper suggests that outcome trends in the never-treated states and the states that were at some point treated largely evolve in parallel even though the levels differ substantially.[15] Although the random treatment time argument is often used to motivate a differences-in-differences estimator[16], it is not strictly necessary to justify the parallel trends assumption. Non-random selection into treatment is permitted if we are willing to believe that the selection bias remains constant over time (parallel trends).

Figure 8 plots the estimated dynamic treatment effects for 6 periods prior to treatment initiation and 10 periods after treatment initiation from a fully dynamic specification that includes all of the never-treated states.[17] Including these units permits us to extend the analysis to the original time window of 1959 to 1997. We find that the pre-trends evident in Figure 7 are attenuated slightly, although we still find some evidence of a pre-trends violation for per-pupil current expenditures. We also find some weak evidence for a pre-trend in the student-teacher ratio, although, again, no single pre-trend coefficient within the 6-period pre-treatment window is statistically significant at $p < .05$. The dynamic treatment effect estimates also strongly suggest null effects for all outcomes except for student-teacher ratio where we see an apparent positive and growing treatment effect. Considering the observed pre-trend, this is also likely consistent with a violation of parallel trends rather than any real effect of collective bargaining.

Lastly, do these apparent pre-trends remain even after using a heterogeneity-robust estimator, or are they an artifact of contamination bias? Figure 9 plots the dynamic treatment effects esti-

15. The main exception here seems to be student-teacher ratios.

16. See Athey and Imbens (2022) and Roth and Sant'Anna (2021).

17. Although we estimate all of the dynamic effects, we constrain the plot to allow for maximum comparability with the original event study figures.

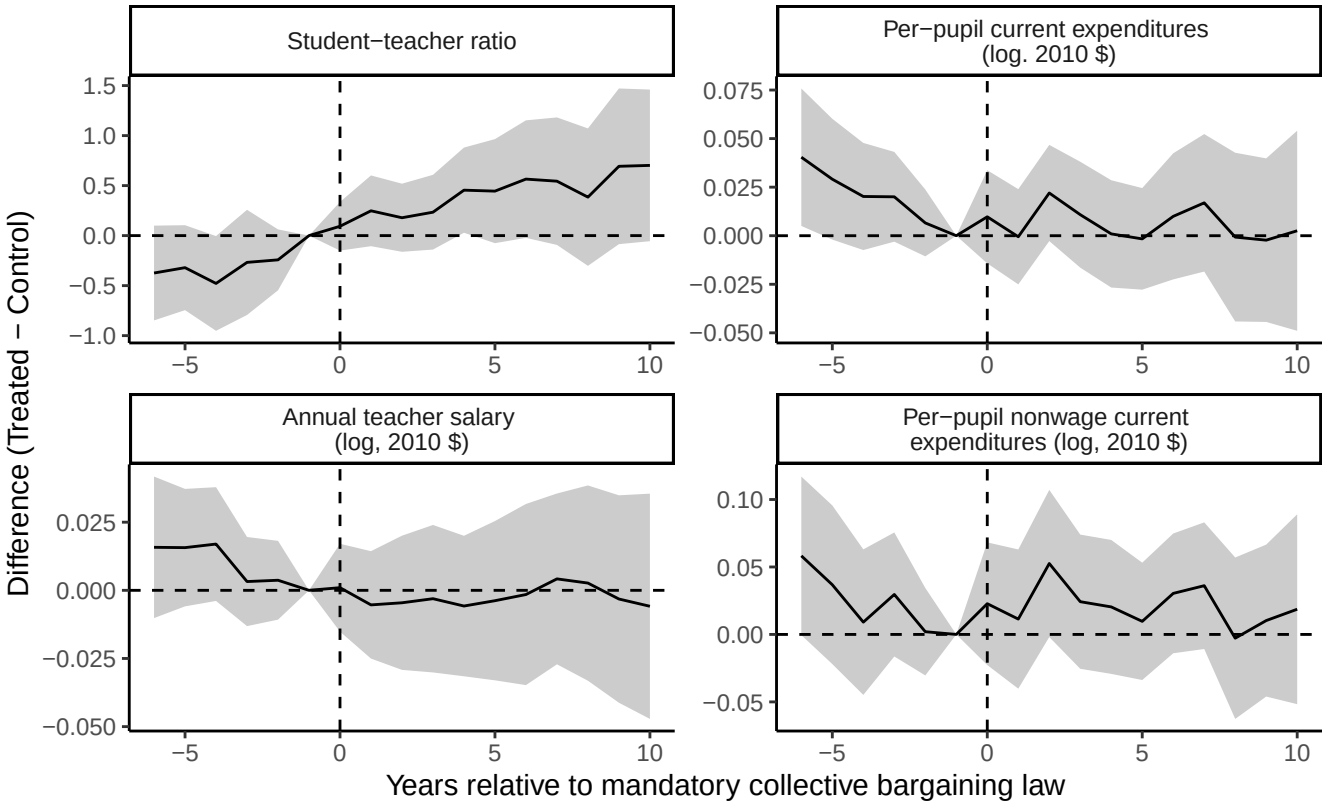

Figure 8: Fully-dynamic event-study plot for Paglayan (2019). Includes never-treated states. 1959-1997

mated using the Callaway and Sant'Anna (2021) estimator. Given the distribution of treatment and the small number of available "clean controls," this estimator has substantially higher variance than the conventional regression event study approach. The low variance of the original reported event study plots was in part an artifact of the restrictions imposed on effect heterogeneity. In general, our heterogeneity-robust estimates fail to find significant pre-trends or treatment effects estimates, although the upward trend for student-teacher ratio is suggestive of a potential parallel trends violation. Overall, these results are largely consistent with the original argument in the paper—there is no clear evidence that the passage of mandatory collective bargaining laws for teachers impacted state spending on education. Although we are slightly more concerned about possible violations of the identifying assumptions, particularly if we are also willing to make the sorts of effect homogeneity assumptions that appear to be necessary for a sufficiently well-powered null, there is little to suggest that the original argument is incorrect. To the extent that some relative-time effects are statistically significant, we conclude this is likely due to violations of the

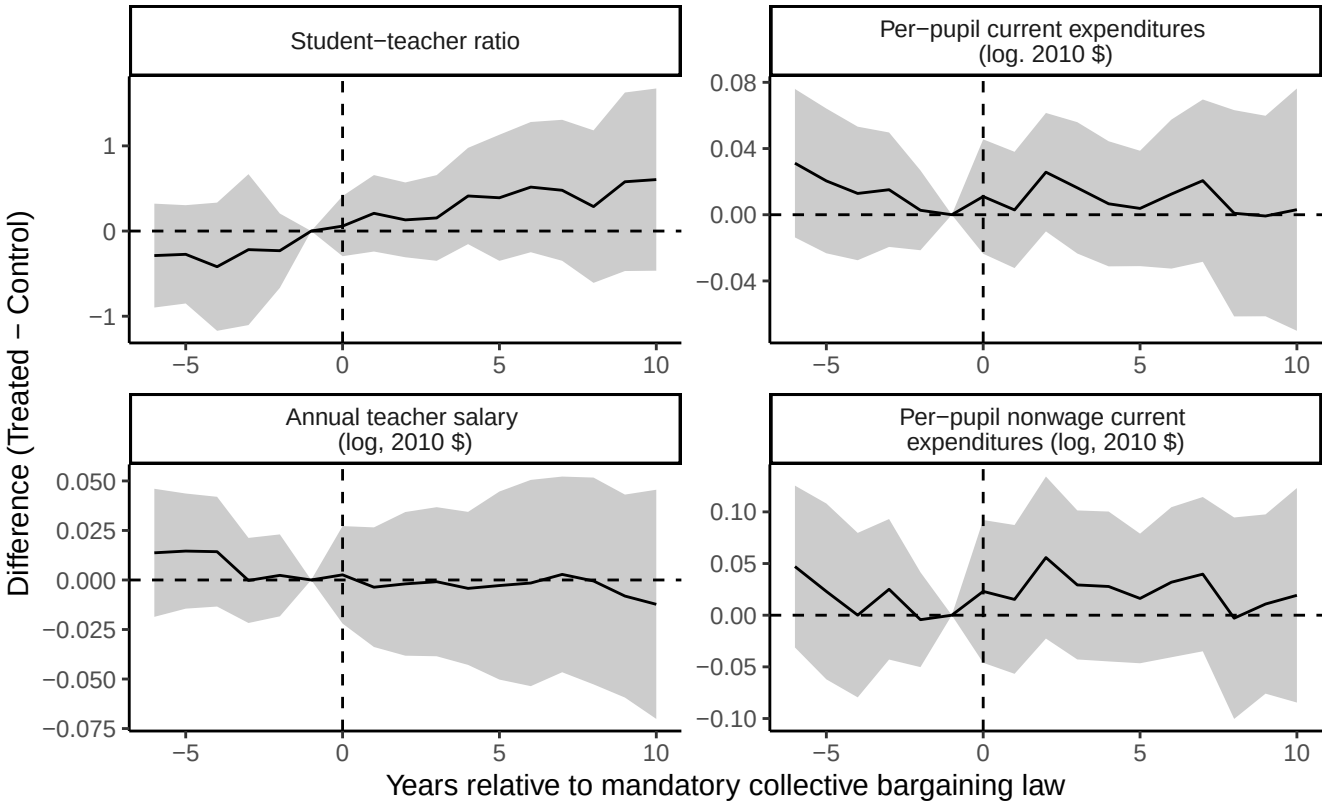

Figure 9: Event-study plot for Paglayan (2019) using the Callaway and Sant'Anna (2021) estimator. Uses not-yet-treated and never-treated states as controls. 1959-1997

identifying assumptions rather than due to the presence of a real effect. Notably, Ben-Michael, Feller, and Rothstein (2022) also replicate this same paper using a different method—the augmented synthetic control estimator. This approach attempts to attain balance between treated and control groups on the pre-treatment outcomes through a weighting approach. Using this method and obtaining highly balanced pre-trends, they likewise find insignificant or weakly negative effects of unionization on spending outcomes, consistent with the paper's original argument.

4.2 Grumbach and Hill (2022)

Grumbach and Hill (2022) examines the effect of state-level same-day registration (SDR) laws on voter turnout. It specifically focuses on effect heterogeneity among voters in different age groups and examines turnout in elections from 1978 to 2018. Figure 10 plots the distribution of treatment and control units and periods. While the majority of states are never-treated, we see

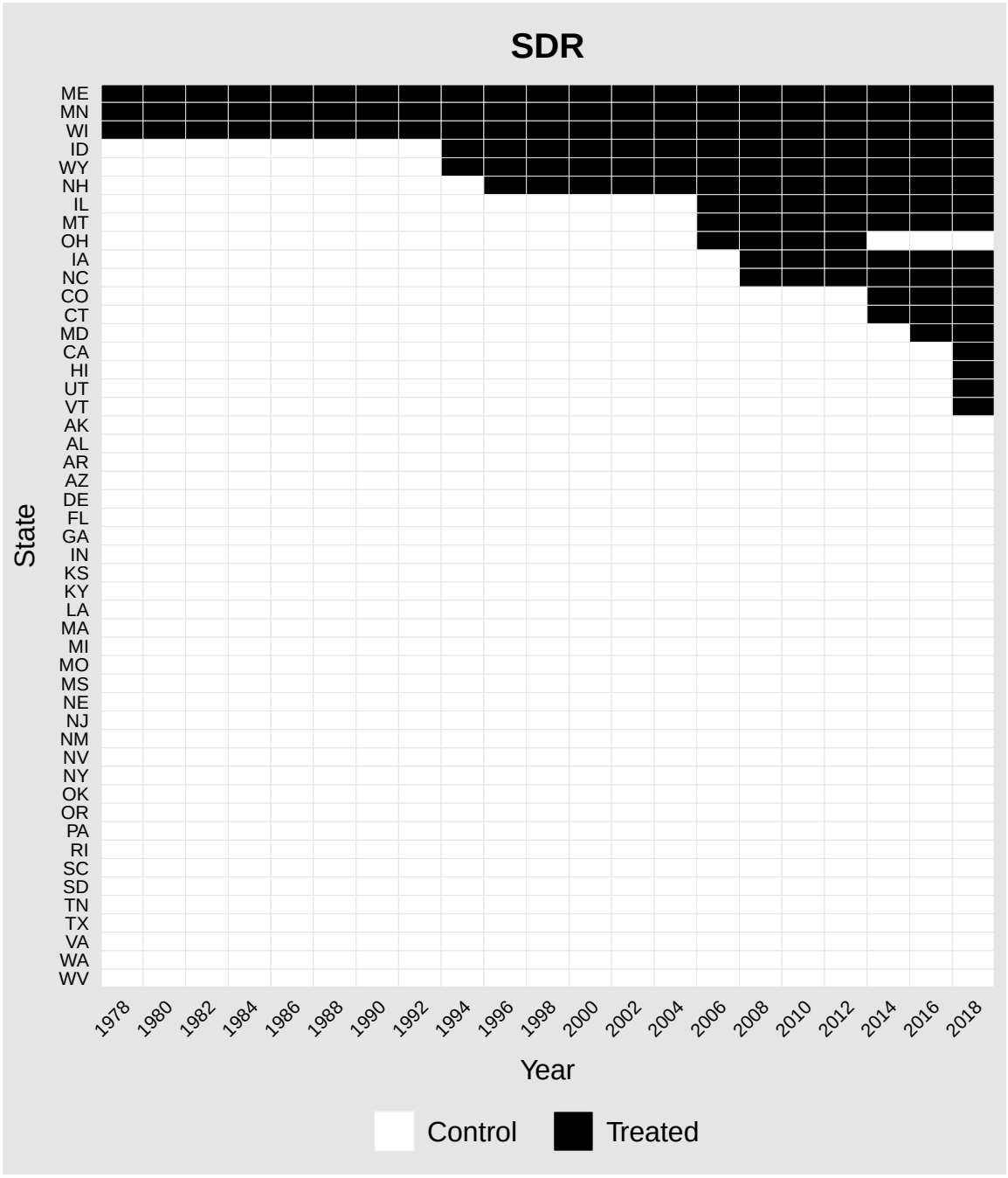

Figure 10: Distribution of treatment and control units in Grumbach and Hill (2022)

staggered adoption of SDR among a handful of states from the 1990s to the 2000s and 2010s. Note that since elections are biannual, we only observe turnout every two years and as a result, the relative time indicators are measured in increments of two in the original paper.

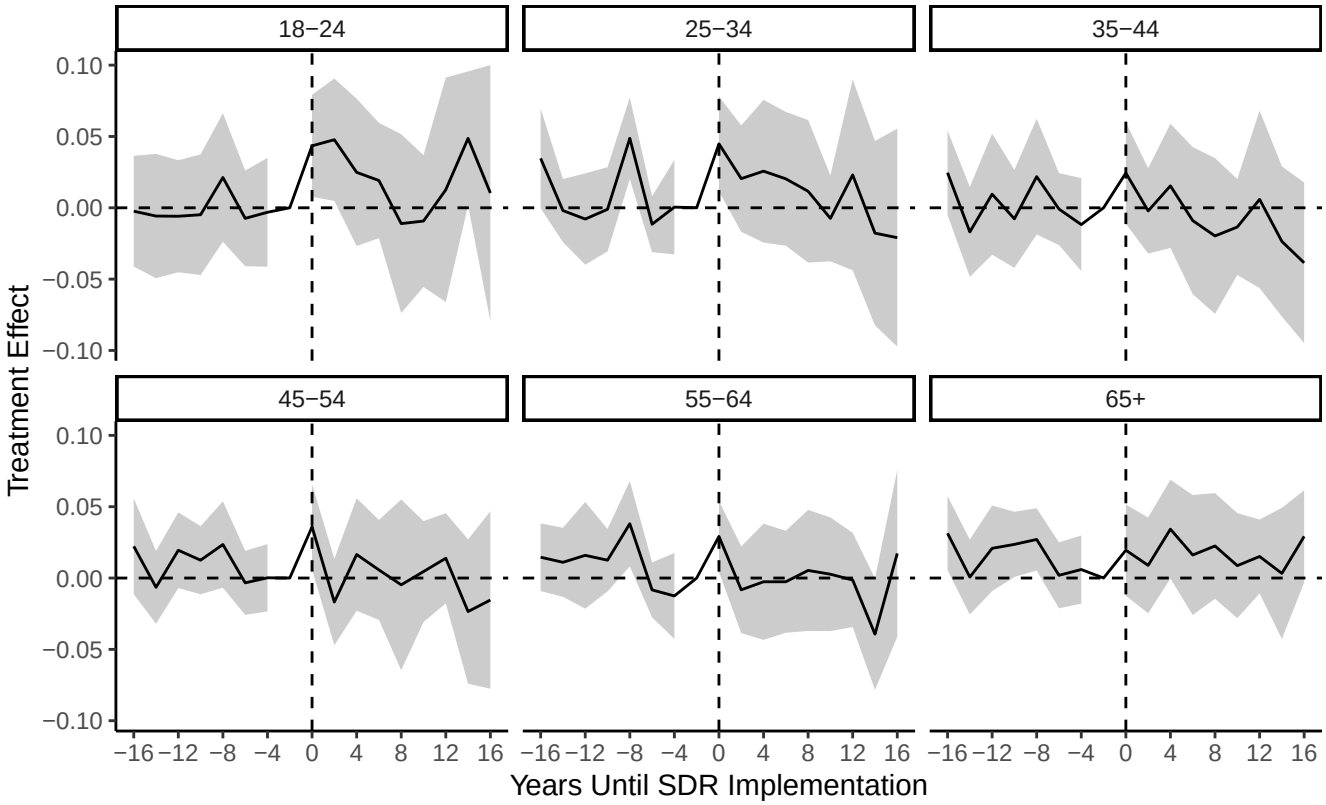

Figure 11: Exact replication of the pre-trend plot in Grumbach and Hill (2022).

We start by replicating the event study plot from the appendix of the original paper.[18] This plot avoids the issues highlighted in Paglayan (2019) by being fully dynamic and by incorporating never-treated states. Although the plots constrain themselves to the 16 years before and after implementation, the particular specification used in the paper is equivalent to the fully saturated regression with a single baseline period (−2 years) omitted. From these estimates, SDR laws appear to increase voter turnout primarily among 18-24-year-olds with largely null results across all other age categories. Pre-trends also appear to be statistically indistinguishable from zero. However, the plot of the effect trajectory suggests that these effects are subject to some decay over time. While the point estimate for the relative time indicator immediately after and 2 years after implementation is statistically distinguishable from 0 at $p < .05$, most of the other effect estimates appear to be statistically insignificant.

18. Note that this figure differs from the published figure due to an error in the appendix replication code which resulted in standard error estimates that were too small. We correct this error in all replications presented in this paper.

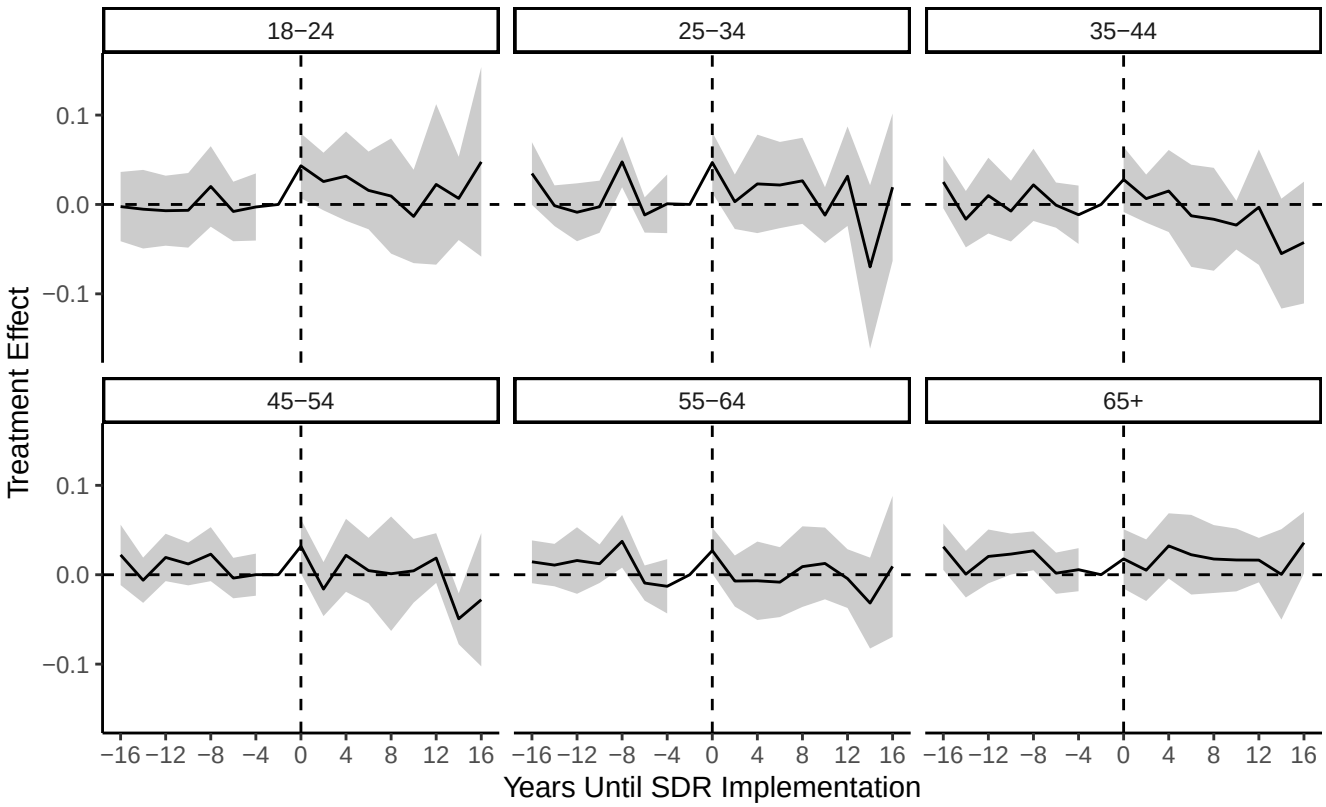

Figure 12: Fully dynamic event study plot Grumbach and Hill (2022). No always-treated units

There are two features of this data structure that raise some issues for the event-study specification. The first is that there appears to be treatment reversal as Ohio implemented SDR in 2006 but eliminated it for the 2014 election and for subsequent election years. The original analysis redefines treatment as a state *ever* having implemented SDR and codes all relative treatment time indicators for Ohio relative to 2006. This approach of converting a setting with some treatment reversal to a conventional staggered adoption structure is generally acceptable and, with minor treatment reversal, should not cause substantial problems for the interpretation of the results. But, for our analysis, it likely results in heterogeneous treatment effect trajectories across different treatment timing groups as one of these timing groups is actually *unexposed* during many of the post-treatment periods. The second feature is the presence of "always-treated" units: Maine, Minnesota and Wisconsin. This is an artifact of the limitations of the data window. Maine implemented its SDR law in 1973 while Minnesota and Wisconsin followed in 1974 and 1975 respectively. As a consequence, there is no pre-treatment period for these units. However, the specification used

in the paper considers these observations treated in 1978 and includes only post-treatment relative time indicators for these observations. Because there is no baseline period, treatment effects for these "always-treated" units are by definition identified indirectly—through differences with other treated units—as there is no way to construct any non-parametric difference-in-differences using only "clean controls" due to the absence of an untreated time period. Prior to implementing the Callaway and Sant'Anna (2021) estimator, we consider an alternative specification of the event study regression that simply removes these three states from the analysis.

We find that omitting these always-treated units has some noticeable consequences for the dynamic treatment effect estimates (Figure 12). While the pre-trends estimates remain unaltered, the effect of SDR on 18-24-year-old turnout is only statistically significant for the period immediately after implementation and decays over time. The confidence intervals also become somewhat wider as we are removing observations, but this does suggest that the peculiar spike in the treatment effect 14-years post-treatment in the original specification may have been attributable to the inclusion of these always-treated units and violations of effect homogeneity.

In fact, the original paper strongly suggests that treatment effect trajectories are likely to be heterogeneous as the effect of SDR on youth turnout is stronger for presidential elections than it is for non-presidential elections. Since election years align with calendar time and not relative time, those states that first implemented SDR in a presidential election should expect to have larger first-period effects than those that implemented SDR first in a non-presidential election. Results from applying the Callaway and Sant'Anna (2021) estimator suggest that the effect homogeneity assumptions underlying the conventional specification are indeed highly consequential. Figure 13 plots the estimated relative-time treatment effect averages. It shows that, rather than trending downward, the treatment effect remains quite stable in the post-treatment window and, in fact, trends upward in the later periods. Again, due to lower precision, we fail to reject the null of no effect for the immediate post-treatment period but do reject the null of no effect for 14 and 16 years post-treatment. Moreover, when aggregating the dynamic Callaway and Sant'Anna (2021) estimates across the entirety of the post-treatment window, we obtain a positive average of the group-time ATTs that is statistically significant at $p < .05$. We do not find a significant effect for any of the other age sub-groups. Overall, while the general evidence for a treatment effect in this sub-group remains, the interpretation of the post-treatment trajectory of estimates changes

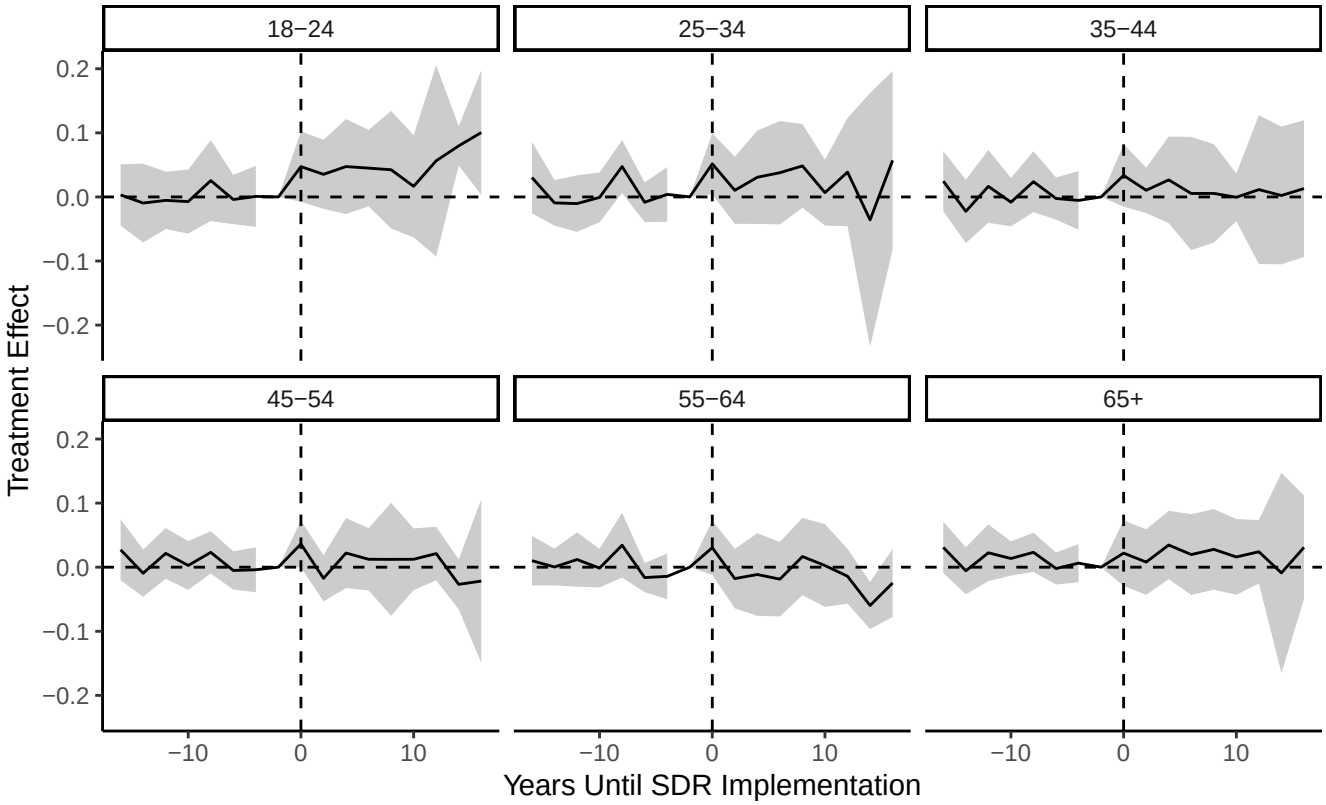

Figure 13: Pre-trend plot for Grumbach and Hill (2022) using the Callaway and Sant'Anna (2021) estimator. Uses not-yet-treated and never-treated states as controls.

substantially. In particular, the sharp decline in the turn-out effect after 4 years that appears in the original plot is likely an artifact of effect heterogeneity.

4.3 Hall and Yoder (2022)

Finally, we re-analyze Hall and Yoder (2022), which looks at the effect of homeownership on turnout. Unlike the other replications, the unit of analysis is the individual. The dataset consists of homeownership data from property ownership records linked to voter file data. Homeownership in time t is defined as owning a home in the locality where one lives during the two-year period after the election at $t-1$ and before the election at time t. We focus on replicating the main analysis, which examines turnout in odd-numbered local elections from 2001–2017, where periods are coded as local elections rather than years (as in Grumbach and Hill (2022)).

Unlike the previous replications, we refrain from generating a plot of the treatment distribution

as the dataset consists of over 7.6 million individual-year observations. However, we note two relevant features. The first is that there appear to be 8 distinct treatment timing groups associated with each odd-numbered election year. Some of these units are coded as homeowners in the entire post-treatment period while others appear to be coded as homeowners only once and revert their treatment. The analysis treats these units in an identical manner and codes the treatment time indicators relative to the first period that an individual appears as a homeowner in the data. The second is that the number of never-treated units is extremely large relative to any of the treatment timing groups and that the group that initiates treatment in 2017 appears to be surprisingly small relative to the other timing groups. It is unclear why this is the case from the description of the data merging process.

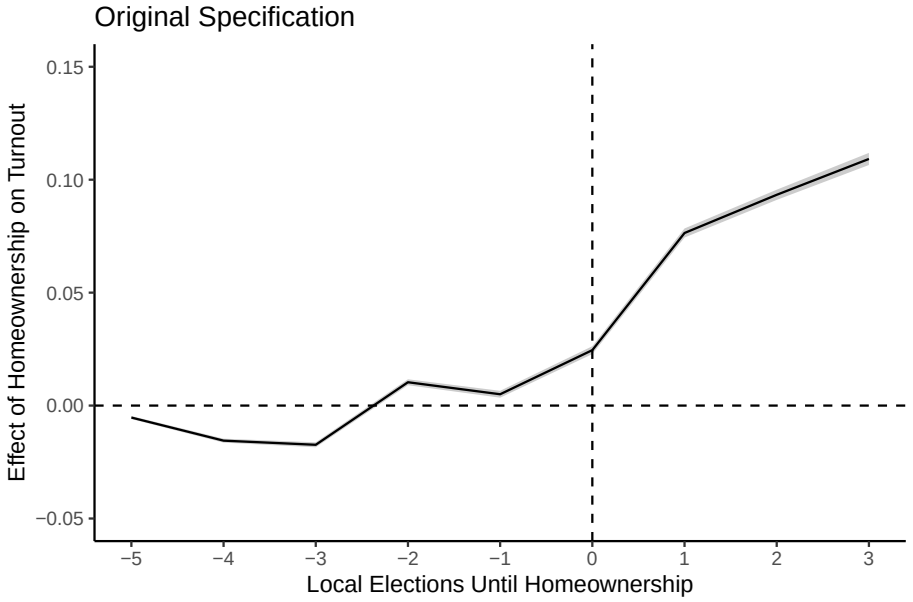

Figure 14: Exact replication of the event-study plot in Hall and Yoder (2022).

As in Paglayan (2019), the original event study specification was not fully dynamic. Although the regression did not omit any post-treatment period as it binned all post-treatment periods after 3, it omitted only the pre-treatment periods prior to -5. The analysis also dropped all units that were either always-treated or never-treated. We replicate this result exactly as presented in the original paper in Figure 14. Although the pre-trends estimates are statistically significant, this is largely due to the extremely large sample size. The paper argues that these pre-trends violations are comparatively small relative to the post-treatment estimates, suggesting that the

parallel trends assumptions is feasible. Additionally, it concludes that the treatment effect builds over time, with subsequent elections post-homeownership.

One problem with omitting only the relative periods prior to -5 is that for over half of the treatment timing groups in the data, no baseline period is omitted. We start by generating an alternative plot using the last period prior to treatment (-1) as the baseline instead. We also avoid binning given the implied effect homogeneity assumptions this imposes. Since there are so few pre- and post-treatment periods, we plot estimates for all of the 8 pre-treatment timing groups and all 7 of the post-treatment timing groups. As the first panel of Figure 15 shows, we draw substantially different conclusions regarding the plausibility of parallel trends compared to the specification presented in the original paper. The pre-trends estimates beyond 5 pre-treatment periods are substantially larger than any of the post-treatment treatment effect estimates. Additionally, the un-binned post-treatment estimates show possible evidence of a decaying effect after 3 years. The treatment effect estimates beyond 3 post-treatment periods are roughly half of what they are in the original, non-fully-dynamic specification (5 percentage points rather than 10 percentage points), suggesting the original estimates exaggerate the long-run impact of homeownership.

Given the substantial parallel trends violations, does the inclusion of non-homeowners make the parallel trends problem worse? As the second panel of Figure 15 illustrates, this does not appear to be the case, similar to what we see when including never-treated units in Paglayan (2019). Most of the pre-trends appear to be very close to zero, apart from relative time periods -7 and -8. In the paper, there appears to be little justification for omitting never-treated periods as individuals who are not homeowners in 2017 could still plausibly become homeowners in 2019 or 2021. The distinction between "never-" and "not-yet"- treated units in this setting is largely an artifact of the observation window rather than a substantive one. Indeed, as we show in Figure 16, although the never-treated units are generally younger than earlier homeowner cohorts, they are not particularly different from the later cohorts in terms of this covariate. We therefore again favor difference-in-differences estimates that incorporate the never-treated units.

The existence of pre-trends for only the extreme time periods (-8 and -7) is puzzling as there are only two treatment timing groups that could possibly contribute to these estimates. We suspect that there may be idiosyncrasies in the dynamic effects across the different timing groups. Given the extremely large number of observations, it does not reduce power significantly to estimate each

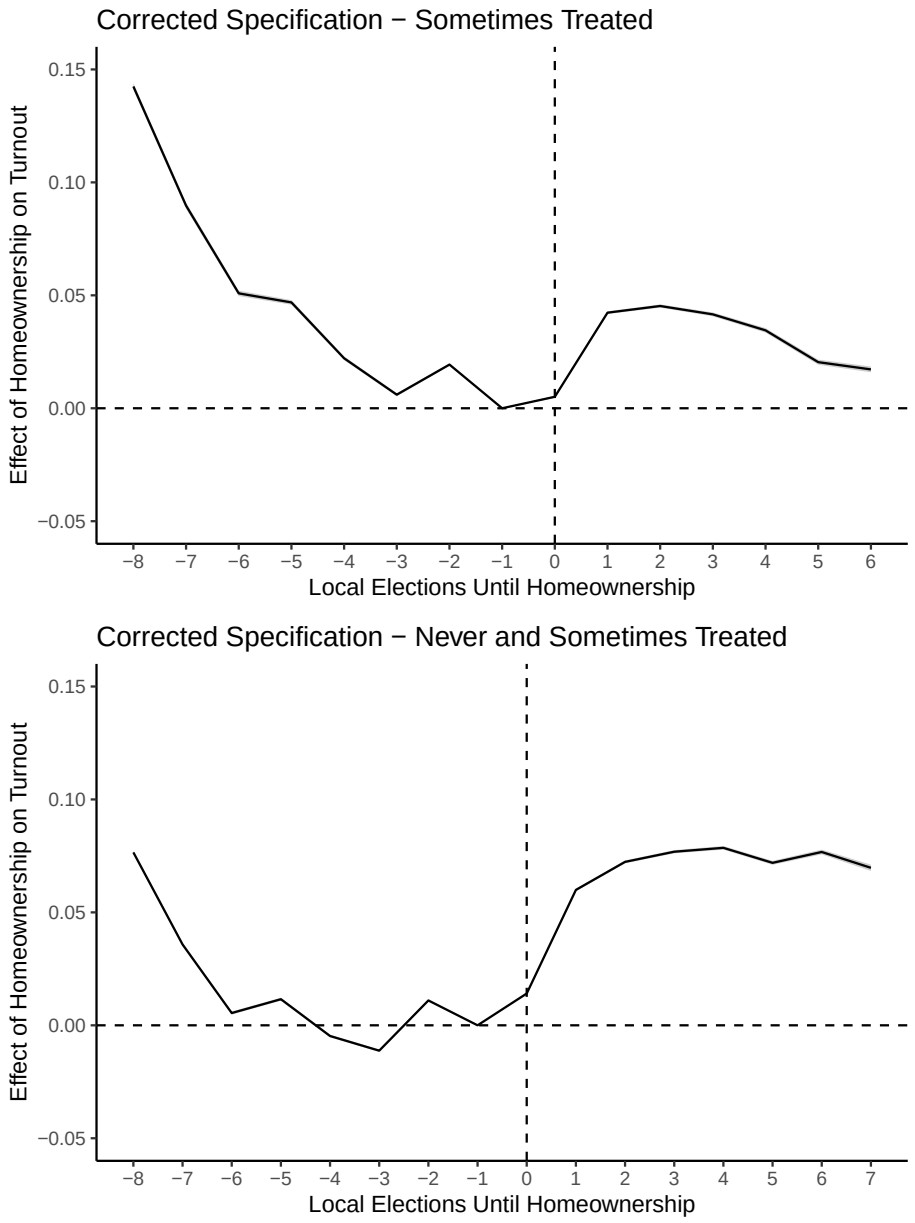

Figure 15: Fully dynamic event-study plot for Hall and Yoder (2022). Sometimes-treated and Sometimes/Never-treated control groups

treatment timing group's ATT trajectory separately. It also allows us to more clearly indicate which cohorts contribute to each relative-time effect and pre-trend. For this sample size, we find it most feasible to use a standard fully saturated two-way fixed effects event study regression applied to each timing group and the never-treateds separately. We do this for each of the 8 treatment timing groups from 2003 to 2017 and present these estimates in Figure 17. Because treatment is staggered, not all pre-trends can be estimated for each timing group. For example, there are no

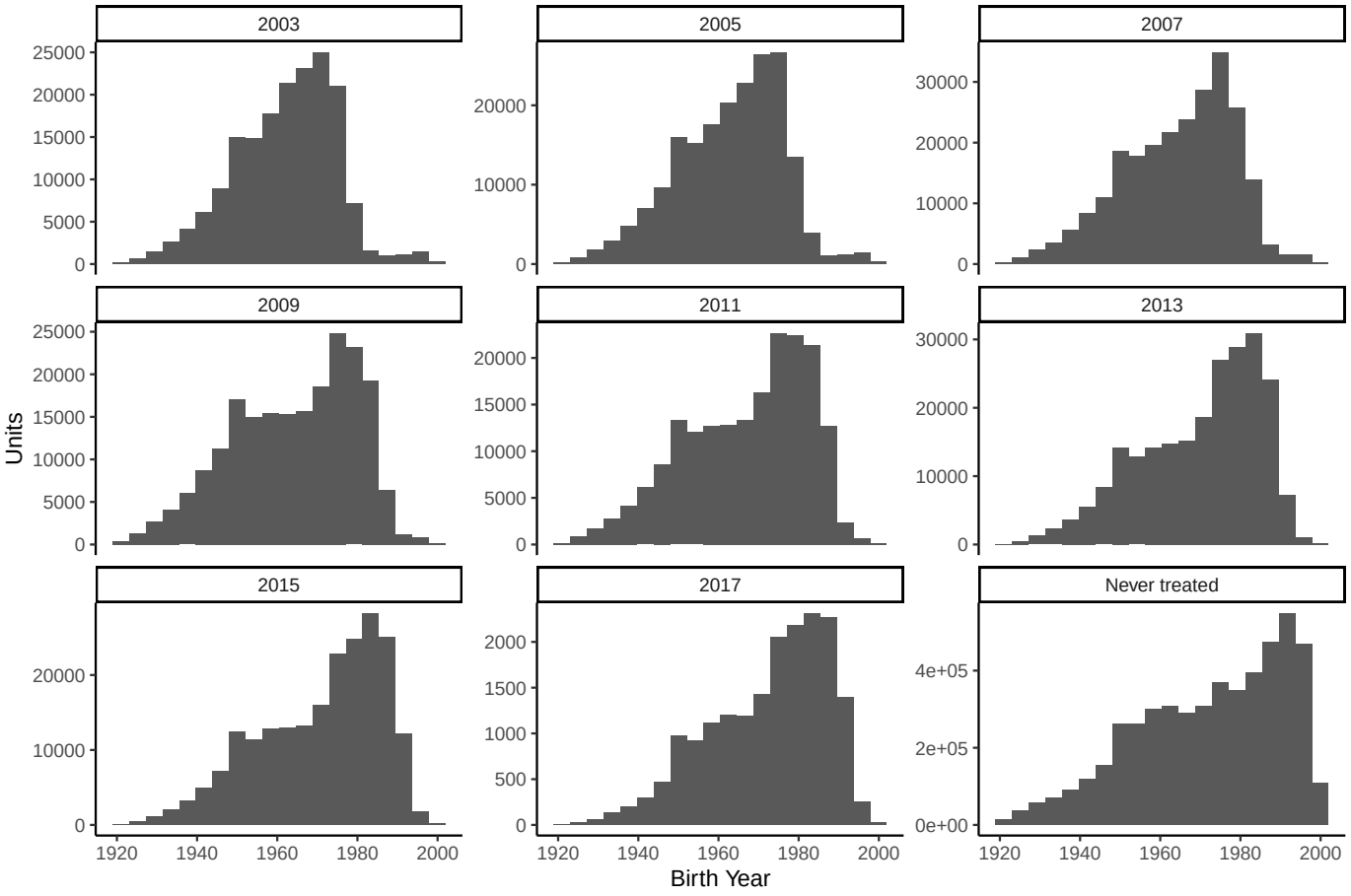

Figure 16: Distribution of birth year by treatment timing group in Hall and Yoder (2022).

pre-treatment placebo tests for the 2003 timing group (as the dataset ends in 2001), while the 2017 timing group contains only a single post-treatment period but 7 pre-treatment placebos.

The results strongly suggest that the pre-trends issues we observe are almost entirely attributable to the 2017 treatment timing cohort, whose pre-treatment placebos are extremely large and unstable. We are unsure of the reasons for this but suspect it may be an issue in the data merging process as this group also has comparatively fewer observations relative to other timing groups. For the individuals who became homeowners in 2007, 2009 and 2011 we see pre-trends very close to zero and treatment effects in the post-treatment period within the 5 percentage point to 10 percentage point range—smaller than the dynamic effect sizes reported in the original paper but not considerably so. Pre-trends for the 2013 and 2015 timing groups appear to be slightly negative and the estimated post-treatment effects are likewise slightly smaller compared to other cohorts. We conclude that, although there are notable pre-trends issues that are obscured by the

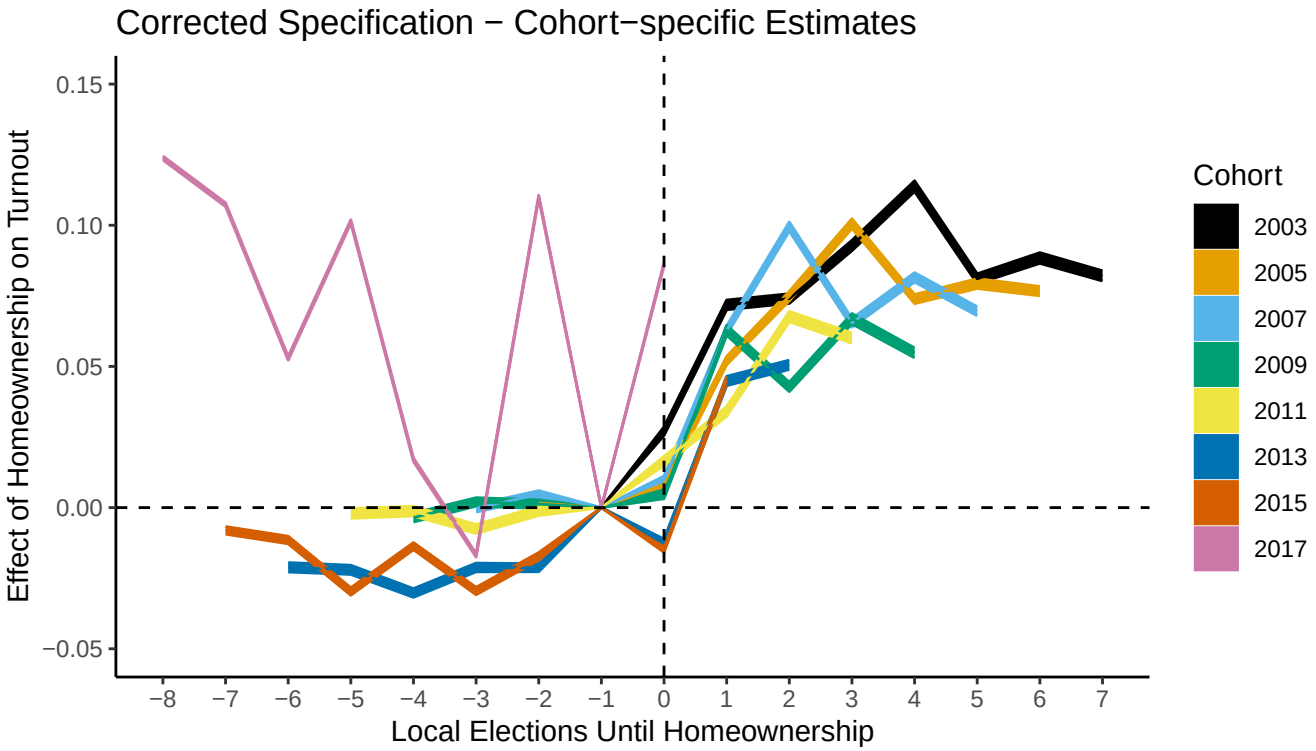

Figure 17: Fully dynamic event-study plot for Hall and Yoder (2022). Group-time ATTs

original specification, these issues are largely attributable the result of a single idiosyncratic timing group in the data and estimates from other timing groups are consistent with the substantive claims made in the original paper. When researchers face settings with such a large sample size and a sufficiently small number of timing groups, they should consider directly estimating each group-time ATT $ATT_g(t)$ separately and plotting the results, if only as a diagnostic measure.

5 Conclusion

As differences-in-differences designs become ubiquitous in applied political science research, researchers using them need to ensure that the underlying identifying assumptions are plausible. Recent work by Chiu et al. (2023) and Hassell and Holbein (2023) suggests that pre-trends violations are extremely common in published political science DiD papers. As a result, political scientists should, at a minimum, be sure to incorporate some placebo diagnostics involving pre-treatment observations when using differences-in-differences designs with multiple time periods.

While practice around the static TWFE regression model has been relatively standardized in the discipline, this is not the case for the dynamic TWFE regression.

In the absence of consistent guidelines for the construction of even study plots, even published studies that include dynamic treatment effect and pre-trends plots may be failing to properly identify evidence of parallel trends violations and mis-characterizing the trajectory of the treatment effects. Some of these problems are readily fixed by re-specifying the dynamic TWFE regression. Improper omission of relative time indicators associated with post-treatment periods can result in significant biases as these omitted periods act as the "baseline" period when constructing the component 2×2 differences-in-differences. Identification requires that there is no treatment effect in these periods, which is not guaranteed if these periods are post-treatment. If researchers wish to estimate the dynamic TWFE, we strongly recommend that they use a fully-dynamic specification with a baseline period that is common to all units even if they do not choose to report all of the relative treatment time coefficients.

However, even this specification may fail to provide unbiased estimates of the treatment effects and pre-trends tests under parallel trends and no anticipation when treatment adoption is staggered. When the path of treatment effects varies across units that adopt treatment early compared to those that adopt treatment later, coefficients on the relative time indicators remain contaminated by the treatment effects from all other time periods included in the regression. We develop a novel decomposition of the sample regression coefficient to help illustrate the intuition behind the contamination bias results in Sun and Abraham (2021). It provides a new interpretation of the dynamic TWFE that contrasts with the static TWFE. While the static TWFE regression still has an interpretation as a convex average over 2×2 differences-in-differences comparisons even under effect heterogeneity, this is not the case for the dynamic TWFE. Rather, the dynamic TWFE should instead be understood as a kind of "sequential" differences-in-differences estimator in which contamination in one 2×2 is eliminated by subtracting off estimates of the contaminating effects from other 2×2 DiD terms. This is accomplished through the implicit weighting on each 2×2 term (Proposition 4) such that all irrelevant effects receive a weight of zero. However, this is only possible if treatment effects are heterogeneous *only* in relative treatment time. In settings where units' response to treatment depends not just on the time-since-treatment but also on other factors such as whether that unit is an early or late-adopter, the dynamic TWFE will fail to identify an

average relative treatment-time effects.

These new findings regarding effect heterogeneity also raise concerns about other common practices. First, there is unclear guidance as to which observations to include as controls in a staggered adoption design. We find that some studies in political science choose to drop never-treated units and assume parallel trends holds only with respect to units that receive treatment at some point in time. While this may be theoretically justified, the absence of a large pool of never-treated units can exacerbate problems due to treatment effect heterogeneity (in addition to reducing power). As we find in our replications, there is no necessary guarantee that pre-trends are necessarily better with respect to these "sometimes-treated" units as opposed to the never-treated units. Second, the failure to remove either time periods where all units are treated or units that are never under control places additional emphasis on effect homogeneity assumptions as does binning observations at the extremes (Borusyak, Jaravel, and Spiess 2021; Schmidheiny and Siegloch 2020). Intuitively, there are *no* non-parametric 2×2 comparisons that can identify treatment effects for the always-treated observations.

Luckily, heterogeneity-robust estimators can help address some of these issues. Although many papers have proposed slightly different methods, all share some common themes. First, they aim to only use valid 2×2 difference-in-differences to identify each group-time ATT. Second, they separate the process of estimating these group-time effects from the process of aggregating them into summary estimands (either full-sample ATTs or the relative-time ATTs we discuss here). In general, the choice of heterogeneity-robust estimator is arguably less consequential than whether one is used at all. However, we emphasize that researchers should be attentive to how their particular choice of estimator selects the component 2×2 comparisons and what the baseline is – particularly for pre-trends estimates.

Although this paper focuses on how to obtain valid dynamic treatment effects and pre-test estimates, it does not discuss some of the other recent related literature on pre-trend testing as a general practice. Notably, underpowered designs can result in pre-trends tests that fail to reject the null even when a substantial trend is present. A number of papers including Bilinski and Hatfield (2018) and Liu, Wang, and Xu (2022) have suggested instead using equivalence-testing approaches as opposed to conventional tests for the null of no pre-trends in order to better characterize the range of possible violations that the design is capable of ruling out. Additionally,

as Roth 2022 notes, the practice of conditioning analyses on a passed pre-test may distort the sizes and confidence intervals for estimated treatment effects. Moreover, we also do not discuss methods that attempt to address parallel trends violations, either by way of covariate adjustment, a latent factor model (Xu 2017; Ben-Michael, Feller, and Rothstein 2022; Athey et al. 2021) or bounds (Rambachan and Roth 2023). We also do not discuss adjustments that incorporate parametric estimates of the trends such as the use of linear and quadratic unit-trends to correct for observed pre-trends as suggested by Hassell and Holbein (2023). However, we urge some caution in viewing polynomial time trends as a universal panacea as they too impose additional assumptions onto the conventional difference-in-differences framework. While in some settings a clear linear trend may be evident in the pre-treatment period, extrapolating this trend still requires researchers to defend their chosen parametric assumption. We also do not discuss the interpretation of many of the new estimators in terms of more flexible fixed effects regression specifications as in Wooldridge (2021).

We hope that this paper aids researchers in better understanding the mechanics underpinning popular two-way fixed effects estimators. In particular, we emphasize that the static TWFE and dynamic TWFE specifications use the data in markedly different ways under staggered adoption. As such, the latter is not simply a weaker or more robust version of the former and they rely on distinct constant effects assumptions for identification when treatment is staggered. We hope also that our review of the "heterogeneity-robust" estimators helps to clarify the specific problems that they address with both of the static and dynamic TWFE specifications. Ultimately, it is essential for researchers estimating effects under differences-in-differences to be able to characterize the underlying variation that contributes to their estimator, whatever method they choose.

References

Abadie, Alberto. 2005. "Semiparametric difference-in-differences estimators." *The review of economic studies* 72 (1): 1–19.

Aronow, Peter M, and Cyrus Samii. 2016. "Does regression produce representative estimates of causal effects?" *American Journal of Political Science* 60 (1): 250–267.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association* 116 (536): 1716–1730.

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

Baker, Andrew C, David F Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics* 144 (2): 370–395.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2022. "Synthetic controls with staggered adoption." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84 (2): 351–381.

Bilinski, Alyssa, and Laura A Hatfield. 2018. "Seeking evidence of absence: Reconsidering tests of model assumptions." *arXiv preprint arXiv:1805.03273.*

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting event study designs: Robust and efficient estimation." *arXiv preprint arXiv:2108.12419.*

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024. *Difference-in-differences with a continuous treatment.* Technical report. National Bureau of Economic Research.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of econometrics* 225 (2): 200–230.

Chiu, Albert, Xingchen Lan, Ziyi Liu, and Yiqing Xu. 2023. "What To Do (and Not to Do) with Causal Panel Analysis under Parallel Trends: Lessons from A Large Reanalysis Study." *Available at SSRN 4490035.*

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2024. "Difference-in-differences estimators of intertemporal treatment effects." *Review of Economics and Statistics,* 1–45.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9): 2964–2996.

———. 2023. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey." *The Econometrics Journal* 26 (3): C1–C30.

Dube, Arindrajit, Daniele Girardi, Oscar Jorda, and Alan M Taylor. 2023. *A local projections approach to difference-in-differences event studies.* Technical report. National Bureau of Economic Research.

Gardner, John. 2022. "Two-stage differences in differences." *arXiv preprint arXiv:2207.05943.*

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

Grumbach, Jacob M, and Charlotte Hill. 2022. "Rock the registration: Same day registration increases turnout of young voters." *The Journal of Politics* 84 (1): 405–417.

Hall, Andrew B, and Jesse Yoder. 2022. "Does homeownership influence political behavior? Evidence from administrative data." *The Journal of Politics* 84 (1): 351–366.

Hassell, Hans, and John B. Holbein. 2023. "Navigating Potential Pitfalls in Difference-in-Differences Designs: Reconciling Conflicting Findings on Mass Shootings' Effect on Electoral Outcomes." *American Political Science Review.*

Imai, Kosuke, and In Song Kim. 2021. "On the use of two-way fixed effects regression models for causal inference with panel data." *Political Analysis* 29 (3): 405–415.

Imai, Kosuke, In Song Kim, and Erik H Wang. 2021. "Matching methods for causal inference with time-series cross-sectional data." *American Journal of Political Science.*

Liu, Licheng, Ziyi Liu, Ye Wang, and Yiqing Xu. 2022. "Package 'fect'."

Liu, Licheng, Ye Wang, and Yiqing Xu. 2022. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data." *American Journal of Political Science.*

Mou, H., L. Liu, and Y. Xu. 2023. "Panel Data Visualization in R (panelView) and Stata (panelview)." *Available at SSRN 4202154.*

Mundlak, Yair. 1978. "On the pooling of time series and cross section data." *Econometrica: journal of the Econometric Society,* 69–85.

Olden, Andreas, and Jarle Møen. 2022. "The triple difference estimator." *The Econometrics Journal* 25 (3): 531–553.

Paglayan, Agustina S. 2019. "Public-sector unions and the size of government." *American Journal of Political Science* 63 (1): 21–36.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A more credible approach to parallel trends." *Review of Economic Studies,* rdad018.

Roth, Jonathan. 2022. "Pretest with caution: Event-study estimates after testing for parallel trends." *American Economic Review: Insights* 4 (3): 305–22.

———. 2024. "Interpreting Event-Studies from Recent Difference-in-Differences Methods." *arXiv preprint arXiv:2401.12309.*

Roth, Jonathan, and Pedro HC Sant'Anna. 2021. "Efficient estimation for staggered rollout designs." *arXiv preprint arXiv:2102.01291.*

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics.*

Rubin, Donald. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66 (5): 688.

———. 1980. "Discussion of" Randomization analysis of experimental data in the Fisher randomization test" by D. Basu." *Journal of the American statistical association* 75:591–593.

Rubin, Donald. 1981. "The bayesian bootstrap." *The annals of statistics,* 130–134.

Schmidheiny, Kurt, and Sebastian Siegloch. 2020. "On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization." *Journal of Applied Econometrics.*

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

VanderWeele, Tyler J. 2009. "Concerning the consistency assumption in causal inference." *Epidemiology* 20 (6): 880–883.

Wooldridge, Jeffrey M. 2021. "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators." *Available at SSRN 3906345.*

Xu, Yiqing. 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis* 25 (1): 57–76.

Appendix

Table of Contents

A Imputation estimators and the problem of "in-sample" placebos

In this appendix, we characterize the regression imputation estimator of Borusyak, Jaravel, and Spiess (2021) and Liu, Wang, and Xu (2022) in terms of the implied 2×2 difference-in-differences. We focus on the non-staggered setting for ease of exposition, but note that the intuition does extend to the staggered setting as well. We decompose two versions of the imputation estimator: one that generates "out-of-sample" imputations for units and periods that are not included in the regression and another that generates "in-sample" imputations on units that are also used to fit the regression model. All the treatment effect estimates involve "out-of-sample" imputations since treated observations are not included in the regression. However, some approaches for generating pre-trends tests—notably the default method for the `fect` R package as of version 1.0.0 (Liu et al. 2022)—use "in-sample" imputations. We caution against this practice and recommend researchers instead use any of the other "out-of-sample" placebo tests (which are also implemented in the `fect` package). We show below that the in-sample two-way fixed effects imputation estimator will tend to under-estimate the magnitude of any pre-trends violation as some of its component 2×2 differences-in-differences are zero by construction.

Assume a design with T periods where treatment is initiated for some units at time period $g^* > 1$. All other units are under control for the entire period. The imputation estimator for the group-time treatment effect for the treated group at time t $ATT_{g^*}(t)$ can be written as:

$$\widehat{ATT}_{g^*}(t) = \frac{1}{N_{g^*}} \sum_{i:G_i=g^*} Y_{it} - \hat{Y}_{it}$$

where $\hat{Y}_{it}$ are the fitted values from the two-way fixed-effects regression fit only to observations with $D_{it} = 0$:

$$\hat{Y}_{it} = \hat{\alpha}_i + \hat{\gamma}_t.$$

The least-squares minimization problem for this regression is:

$$\hat{\alpha}, \hat{\gamma} = \underset{\alpha,\gamma}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{i:D_{it}=0} (Y_{it} - \alpha_i - \gamma_t)^2$$

Solving for $\hat{\alpha}_i$ and $\hat{\gamma}_t$

$$\begin{aligned} 0 &= \sum_{t:D_{it}=0} \Big(\hat{\alpha}_i + \hat{\gamma}_t - Y_{it} \Big) \\ \hat{\alpha}_i[T \times \mathbf{1}(G_i = \infty) + (g^* - 1) \times \mathbf{1}(G_i = g^*)] &= \sum_{t:D_{it}=0} Y_{it} - \sum_{t:D_{it}=0} \hat{\gamma}_t \\ \hat{\alpha}_i &= \frac{\sum_{t:D_{it}=0} Y_{it} - \sum_{t:D_{it}=0} \hat{\gamma}_t}{T \times \mathbf{1}(G_i = \infty) + (g^* - 1) \times \mathbf{1}(G_i = g^*)} \end{aligned}$$

and

$$\begin{aligned} 0 &= \sum_{i:D_{it}=0} \Big(\hat{\gamma}_t + \hat{\alpha}_i - Y_{it} \Big) \\ \hat{\gamma}_t[N \times \mathbf{1}(t < g^*) + N_\infty \mathbf{1}(t > g^*)] &= \sum_{i:D_{it}=0} Y_{it} - \sum_{i:D_{it}=0} \hat{\alpha}_i \\ \hat{\gamma}_t &= \frac{\sum_{i:D_{it}=0} Y_{it} - \sum_{i:D_{it}=0} \hat{\alpha}_i}{N \times \mathbf{1}(t < g^*) + N_\infty \mathbf{1}(t \geq g^*)} \end{aligned}$$

We have four types of fixed effect terms. Let $\bar{i}$ denote a treated unit with $G_{\bar{i}} = g^*$ and $\underline{i}$ denote a control unit with $G_{\underline{i}} = \infty$. Let $\underline{t}$ denote any $t < g^*$ and $\bar{t}$ denote any $t \geq g^*$. Then

$$\hat{\alpha}_{\bar{i}} = \frac{1}{g^* - 1} \sum_{t=1}^{g^*-1} Y_{it} - \frac{1}{g^* - 1} \sum_{t=1}^{g^*-1} \hat{\gamma}_t$$

$$\hat{\alpha}_{\underline{i}} = \frac{1}{T} \sum_{t=1}^{T} Y_{it} - \frac{1}{T} \Bigg[\sum_{t=1}^{g^*-1} \hat{\gamma}_t + \sum_{t=g^*}^{T} \hat{\gamma}_t \Bigg]$$

$$\hat{\gamma}_{\underline{t}} = \frac{1}{N} \Bigg[\sum_{i:G_i=\infty} Y_{it} + \sum_{i:G_i=g^*} Y_{it} \Bigg] - \frac{1}{N} \Bigg[\sum_{i:G_i=\infty} \hat{\alpha}_i + \sum_{i:G_i=g^*} \hat{\alpha}_i \Bigg]$$

$$\hat{\gamma}_{\bar{t}} = \frac{1}{N_\infty}\Big[\sum_{i:G_i=\infty} Y_{it}\Big] - \frac{1}{N_\infty}\Big[\sum_{i:G_i=\infty} \hat{\alpha}_i\Big]$$

Substituting $\hat{\gamma}_{\underline{t}}$ into $\hat{\alpha}_{\bar{i}}$

$$\hat{\alpha}_{\bar{i}} = \frac{1}{g^*-1}\sum_{t'=1}^{g^*-1} Y_{it'} - \frac{1}{N(g^*-1)}\sum_{t'=1}^{g^*-1}\sum_{i':G_{i'}=\infty} Y_{i't'} - \frac{1}{N(g^*-1)}\sum_{t'=1}^{g^*-1}\sum_{i':G_{i'}=g^*} Y_{i't'} + \frac{1}{N}\Big[\sum_{i':G_{i'}=\infty}\hat{\alpha}_{i'} + \sum_{i':G_{i'}=g^*}\hat{\alpha}_{i'}\Big]$$

We consider first the "in-sample" imputation estimator of the ATT for a "pre-treatment" time period $\underline{t}$. As of version 1.0.0 of the `fect` R package, these are the estimates that are reported for pre-treatment periods under the default settings (`loo = FALSE` and `placeboTest = FALSE`) when using the main `fect()` function.

For any "pre-treatment" time period $\underline{t}$, the in-sample imputation estimator can be written as:

$$\widehat{ATT}_{g^*}(\underline{t}) = \frac{1}{N_{g^*}}\sum_{i:G_i=g^*}\Big\{Y_{i\underline{t}} - \frac{1}{(g^*-1)}\sum_{t'=1}^{g^*-1} Y_{it'} - \frac{1}{N}\sum_{i':G_{i'}=\infty} Y_{i'\underline{t}} - \frac{1}{N}\sum_{i':G_{i'}=g^*} Y_{i't} + \frac{1}{N(g^*-1)}\sum_{t'=1}^{g^*-1}\sum_{i':G_{i'}=\infty} Y_{i't'} + \frac{1}{N(g^*-1)}\sum_{t'=1}^{g^*-1}\sum_{i':G_{i'}=g^*} Y_{i't'}\Big\}$$

Rearranging terms to write in terms of individual "difference-in-differences"

$$\widehat{ATT}_{g^*}(\underline{t}) = \frac{1}{N_{g^*}(g^*-1)N}\sum_{t'=1}^{g^*-1}\sum_{i:G_i=g^*}\sum_{i':G_i=\infty}\Big[Y_{i\underline{t}} - Y_{it'} - Y_{i'\underline{t}} + Y_{i't'}\Big] + \frac{1}{N_{g^*}(g^*-1)N}\sum_{t'=1}^{g^*-1}\sum_{i:G_i=g^*}\sum_{i':G_i=g^*}\Big[Y_{i\underline{t}} - Y_{it'} - Y_{i'\underline{t}} + Y_{i't'}\Big]$$

It is straightforward to show that the second triple-sum is equal to zero

$$\widehat{ATT}_{g^*}(\underline{t}) = \frac{1}{N_{g^*}(g^*-1)N}\sum_{t'=1}^{g^*-1}\sum_{i:G_i=g^*}\sum_{i':G_i=\infty}\Big[Y_{i\underline{t}} - Y_{it'} - Y_{i'\underline{t}} + Y_{i't'}\Big] + \frac{1}{N_{g^*}(g^*-1)N}\sum_{t'=1}^{g^*-1}\sum_{i:G_i=g^*}\sum_{i':G_i=g^*}\Big[Y_{i\underline{t}} - Y_{it'}\Big] - \frac{1}{N_{g^*}(g^*-1)N}\sum_{t'=1}^{g^*-1}\sum_{i:G_i=g^*}\sum_{i':G_i=g^*}\Big[Y_{i'\underline{t}} - Y_{i't'}\Big]$$

Swapping indices i and i'

$$\widehat{ATT}_{g^*}(\underline{t}) = \frac{1}{N_{g^*}(g^*-1)N} \sum_{t'=1}^{g^*-1} \sum_{i:G_i=g^*} \sum_{i':G_i=\infty} \left[Y_{i\underline{t}} - Y_{it'} - Y_{i'\underline{t}} + Y_{i't'} \right]$$
$$+ \frac{1}{N_{g^*}(g^*-1)N} \sum_{t'=1}^{g^*-1} \sum_{i:G_i=g^*} \sum_{i':G_i=g^*} \left[Y_{i\underline{t}} - Y_{it'} \right] - \frac{1}{N_{g^*}(g^*-1)N} \sum_{t'=1}^{g^*-1} \sum_{i:G_i=g^*} \sum_{i':G_i=g^*} \left[Y_{i\underline{t}} - Y_{it'} \right]$$

The difference-in-differences term is also equal to zero when $\underline{t} = t'$.

$$\widehat{ATT}_{g^*}(\underline{t}) = \frac{1}{N_{g^*}(g^*-1)N} \sum_{\substack{t'=1 \\ t' \neq \underline{t}}}^{g^*-1} \sum_{i:G_i=g^*} \sum_{i':G_i=\infty} \left[Y_{i\underline{t}} - Y_{it'} - Y_{i'\underline{t}} + Y_{i't'} \right]$$

Using our definition of $\bar{Y}_{g,t}$, we have:

$$\widehat{ATT}_{g^*}(\underline{t}) = \left[\frac{N_\infty}{N} \frac{(g^*-2)}{(g^*-1)} \right] \times \frac{1}{g^*-2} \sum_{\substack{t'=1 \\ t' \neq \underline{t}}}^{g^*-1} \left[\bar{Y}_{g^*,\underline{t}} - \bar{Y}_{g^*,t'} - \bar{Y}_{\infty,\underline{t}} + \bar{Y}_{\infty,t'} \right]$$

We end up with an expression for the in-sample imputation estimator in terms of an average over every difference-in-differences between the period of interest $\underline{t}$ and every *other* pre-treatment period t' and a shrinkage factor $\frac{N_\infty}{N} \frac{(g^*-2)}{(g^*-1)}$. This factor is strictly less than 1 since $N_\infty = N - N_{g^*} < N$. As a result, this average over the 2×2 differences-in-differences is attenuated towards zero.

Contrast the in-sample fit with the expression for the imputation estimator for a held-out (post-treatment) period $\bar{t}$:

$$\hat{Y}_{i\bar{t}} = \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} Y_{it'} + \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} Y_{i'\bar{t}} - \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} \hat{\gamma}_{t'} - \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} \hat{\alpha}_{i'}$$

Adding zero

$$\hat{Y}_{i\bar{t}} = \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} Y_{it'} + \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} Y_{i'\bar{t}} - \frac{1}{N_\infty(g^*-1)} \sum_{i':G_{i'}=\infty} \sum_{t'=1}^{g^*-1} Y_{i't'}$$
$$+ \frac{1}{N_\infty(g^*-1)} \sum_{i':G_{i'}=\infty} \sum_{t'=1}^{g^*-1} Y_{i't'} - \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} \hat{\gamma}_{t'} - \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} \hat{\alpha}_{i'}$$

Writing the second line as a sum over residuals

$$\hat{Y}_{i\bar{t}} = \frac{1}{g^*-1}\sum_{t'=1}^{g^*-1} Y_{it'} + \frac{1}{N_\infty}\sum_{i':G_{i'}=\infty} Y_{i'\bar{t}} - \frac{1}{N_\infty(g^*-1)}\sum_{i':G_{i'}=\infty}\sum_{t'=1}^{g^*-1} Y_{i't'} + \frac{1}{N_\infty(g^*-1)}\sum_{i':G_{i'}=\infty}\sum_{t'=1}^{g^*-1}\left[Y_{i't'} - \hat{Y}_{i't'}\right]$$

By the properties of OLS, the sum of the regression residuals equals zero. Since the regression is fit over all observations with $D_{it} = 0$, we can write the sum of the residuals as a sum of three terms.

$$0 = \sum_{i':G_{i'}=\infty}\sum_{t'=1}^{g^*-1}\left[Y_{i't'} - \hat{Y}_{i't'}\right] + \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1}\left[Y_{i't'} - \hat{Y}_{i't'}\right] + \sum_{i':G_{i'}=\infty}\sum_{t'=g^*}^{T}\left[Y_{i't'} - \hat{Y}_{i't'}\right]$$

We can show that the latter two terms are equal to 0

$$\begin{aligned}
\sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1}\left[Y_{i't'} - \hat{Y}_{i't'}\right] &= \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1} Y_{i't'} - (g^*-1)\sum_{i':G_{i'}=g^*}\hat{\alpha}_{i'} - N_{g^*}\sum_{t'=1}^{g^*-1}\hat{\gamma}_{t'} \\
&= \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1} Y_{i't'} - \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1} Y_{i't'} + \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1}\hat{\gamma}_{t'} - N_{g^*}\sum_{t'=1}^{g^*-1}\hat{\gamma}_{t'} \\
&= \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1} Y_{i't'} - \sum_{i':G_{i'}=g^*}\sum_{t'=1}^{g^*-1} Y_{i't'} + N_{g^*}\sum_{t'=1}^{g^*-1}\hat{\gamma}_{t'} - N_{g^*}\sum_{t'=1}^{g^*-1}\hat{\gamma}_{t'} \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i':G_{i'}=\infty}\sum_{t'=g^*}^{T}\left[Y_{i't'} - \hat{Y}_{i't'}\right] &= \sum_{i':G_{i'}=\infty}\sum_{t'=g^*}^{T} Y_{i't'} - (T-g^*+1)\sum_{i':G_{i'}=\infty}\hat{\alpha}_i - N_\infty\sum_{t'=g^*}^{T}\hat{\gamma}_{t'} \\
&= \sum_{i':G_{i'}=\infty}\sum_{t'=g^*}^{T} Y_{i't'} - (T-g^*+1)\sum_{i':G_{i'}=\infty}\hat{\alpha}_i - \sum_{t'=g^*}^{T}\sum_{t'=g^*}^{T} Y_{i't'} + \sum_{t'=g^*}^{T}\sum_{i':G_{i'}=\infty}\hat{\alpha}_i \\
&= -(T-g^*+1)\sum_{i':G_{i'}=\infty}\hat{\alpha}_i + (T-g^*+1)\sum_{i':G_{i'}=\infty}\hat{\alpha}_i \\
&= 0
\end{aligned}$$

Therefore

$$0 = \sum_{i':G_{i'}=\infty} \sum_{t'=1}^{g^*-1} \left[Y_{i't'} - \hat{Y}_{i't'} \right]$$

and

$$\hat{Y}_{i\bar{t}} = \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} Y_{it'} + \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} Y_{i'\bar{t}} - \frac{1}{N_\infty(g^*-1)} \sum_{i':G_{i'}=\infty} \sum_{t'=1}^{g^*-1} Y_{i't'}$$

Which yields an expression for the out-of-sample imputation estimator as an average over 2×2 difference-in-differences between the period of interest $\bar{t}$ and *all other* pre-treatment baselines t'.

$$\begin{aligned}
\widehat{ATT}_{g^*}(\bar{t}) &= \frac{1}{N_{g^*}} \sum_{i:G_i=g^*} \left\{ Y_{i\bar{t}} - \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} Y_{it'} - \frac{1}{N_\infty} \sum_{i':G_{i'}=\infty} Y_{i'\bar{t}} + \frac{1}{N_\infty(g^*-1)} \sum_{i':G_{i'}=\infty} \sum_{t'=1}^{g^*-1} Y_{i't'} \right\} \\
&= \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} \frac{1}{N_{g^*} N_\infty} \sum_{i:G_i=g^*} \sum_{i':G_{i'}=\infty} \left[Y_{i\bar{t}} - Y_{i'\bar{t}} - Y_{it'} + Y_{i't'} \right] \\
&= \frac{1}{g^*-1} \sum_{t'=1}^{g^*-1} \left[\bar{Y}_{g^*,\bar{t}} - \bar{Y}_{g^*,t'} - \bar{Y}_{\infty,\bar{t}} + \bar{Y}_{\infty,t'} \right]
\end{aligned}$$

By ensuring that the time period for which the imputations are generated is *not* included in the regression fit, we avoid the aforementioned attenuation bias. This can be accomplished for the pre-treatment placebo effects in one of the two ways described in Liu, Wang, and Xu (2022), either holding out *just* the pre-treatment period of interest and estimating a separate placebo regression for each (the "leave-one-out" estimator) or holding out some pre-defined number of periods prior to treatment initiation and using all other non-held-out pre-treatment periods as the baselines (the "placebo test"). These two approaches are implemented in the main `fect()` estimation function in the `fect` R package but must be enabled by the user explicitly by setting either `loo = TRUE` for the leave-one-out estimator or `placeboTest = TRUE` for the placebo test. We strongly recommend that researchers not use the default settings and elect to enable one of these two approaches.

We illustrate the magnitude of the bias in the in-sample imputation placebos compared to the leave-one-out approach using a simple Monte Carlo simulation. We compare the performance

of the leave-one-out imputation estimator to the in-sample imputation estimator across different scenarios, varying the magnitude of the parallel trends violation, the proportion of treated units, and the number of pre-treatment periods. For simplicity, we do not stagger treatment adoption (keeping with the theoretical setting described above). We implement the in-sample imputation estimator by fitting the two-way fixed effects regression once to the not-yet-treated and never-treated units and time periods and generating the predicted counterfactual for the period of interest. The "leave-one-out" estimator omits observations for the period of interest for "treated" units from the two-way fixed effects fit. For each estimator, we use the cluster bootstrap to construct our confidence intervals and present the rejection rate of the conventional hypothesis test under the null that the placebo ATE parameter is zero.

The data-generating process for our outcome models an "anticipation" effect in the period prior to treatment adoption. Formally, we construct the outcome for each iteration of the simulation with the TWFE structure:

$$Y_{it} = v \times \alpha_i \times D_{it}^{(-1)} + \alpha_i + \gamma_t + \varepsilon_{it}$$

$D_{it}^{(-1)}$ is an indicator for whether unit i at time t is one period before initiating treatment. v is the trend-violation parameter, α_i is a unit-specific shock, γ_t is a common time shock, and ε_{it} is a mean-zero error term. With no staggering, treated units initiate treatment at a common time $G_i = g^*$ and control units ($G_i = \infty$) never initiate treatment. The components are generated as follows:

$$\begin{aligned}
&D_i = \text{Bernoulli}(\pi) \\
&G_i = g^* \text{ if } D_i = 1, \infty \text{ otherwise} \\
&\gamma_t = t \\
&\alpha_i \sim \mathcal{N}(\mathbf{1}(G_i = g^*), 1) \\
&D_{it}^{(-1)} = \begin{cases} 1 & \text{if } G_i = g^* \text{ and } t = g^* - 1 \\ 0 & \text{otherwise} \end{cases} \\
&\varepsilon_{it} \sim \mathcal{N}(0, 1)
\end{aligned}$$

Note that the treatment effect is 0 in our outcome model. We run the simulation for 1,000 iterations with $N = 1000$ units and $T = 10$ periods. Standard errors are estimated using a cluster bootstrap procedure as in `fect`.[19] We vary the expected proportion of treated units $\pi \in \{0.25, 0.5, 0.75\}$, the treatment initiation time $g^* \in \{3, 5, 7\}$ (which correspond to $2, 4, 6$, pre-treatment periods respectively), and the magnitude of the parallel trends violation $v \in \{0, 0.1, 0.2\}$.

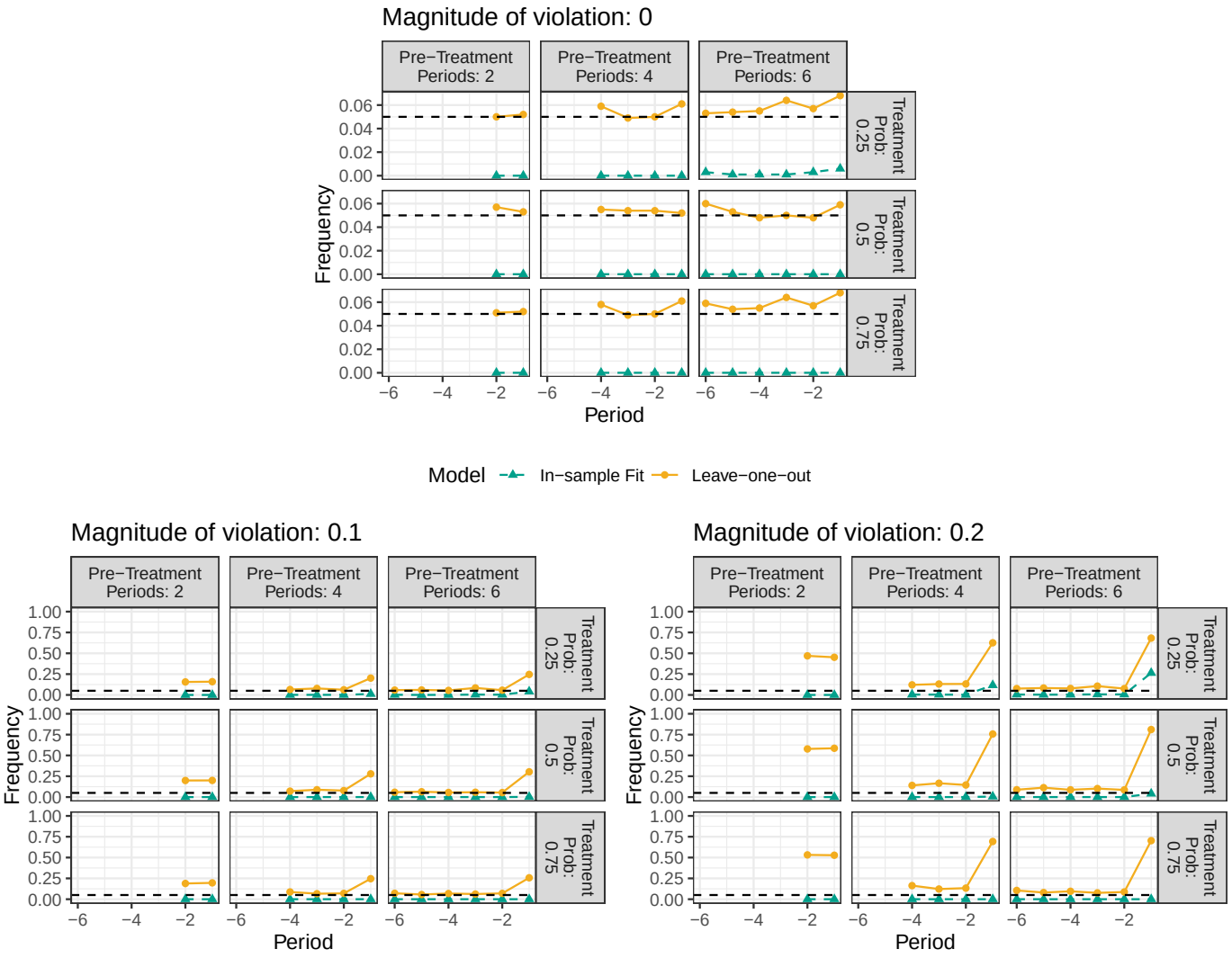

Figure A1: Rejection rates under different trend violation magnitudes: leave-one-out vs. in-sample imputation

Figure A1 shows the results of this simulation. The y-axis shows the rejection rate under a 5%

19. Because of the fixed effects parameters in the regression, we use a cluster Bayesian bootstrap (Rubin 1981) for ease of implementation. Instead of resampling observations, this approach re-weights each unit by a random continuous weight drawn from a Dirichlet distribution.

significance level. The top plot presents the case where there is no trend violation and illustrates the *under-rejection* problem created by the in-sample placebo test. When the null is true, the leave-one-out estimator's rejection rate is close to the nominal level of 5%, as would be expected. The in-sample imputation estimator, however, has a rejection rate that is much lower—nearly zero for most of the simulations. The one case where we start to see a slight increase in the rejection rate is when we have 6 pre-treatment periods and a larger number of control units (only 1/4 treated). Even here, we see that the in-sample imputation estimator is far too conservative.

When we allow the trend violation to be non-zero (the bottom two plots), we find that the power of the leave-one-out placebo test estimator in detecting the violation for period -1 is *much* higher. Rejection rates for the in-sample imputation estimator are still far too low. This is most clearly illustrated in the case for our largest "anticipation effect" magnitude. When the treatment-control split is equal and we have six pre-treatment periods, our simulations show the leave-one-out estimator rejecting the null at a rate of over 75 percent while the in-sample imputation estimator's rejection rate still hovers at about 5 percent. Consistent with our earlier results, the performance of the in-sample imputation estimator is best when the proportion of treated units is small (N_∞/N is large) and the number of pre-treatment periods is large ($(g^* - 2)/(g^* - 1)$ is close to 1). Even here, we see markedly better performance for the leave-one-out estimator. Because the in-sample imputation estimator, by construction, leans heavily towards detecting no violation even when notable pre-trends violations exist, we strongly recommend against its use when conducting pre-trends tests.

B Proofs of results

B.1 Proof of Proposition 1

Write the expectation of $\hat{\tau}_{gt}(g',t')$ conditional on the design $\mathcal{D}$. By consistency,

$$\begin{aligned}
E[\hat{\tau}_{gt}(g',t')|\mathcal{D}] &= E[\bar{Y}_{g,t}|\mathcal{D}] - E[\bar{Y}_{g',t}|\mathcal{D}] - E[\bar{Y}_{g,t'}|\mathcal{D}] + E[\bar{Y}_{g,t'}|\mathcal{D}] \\
&= \frac{1}{N_g}\sum_{i:G_i=g} E[Y_{it}|\mathcal{D}] - \frac{1}{N_{g'}}\sum_{i:G_i=g'} E[Y_{it}|\mathcal{D}] - \frac{1}{N_g}\sum_{i:G_i=g} E[Y_{it'}|\mathcal{D}] + \frac{1}{N_{g'}}\sum_{i:G_i=g'} E[Y_{it'}|\mathcal{D}] \\
&= \frac{1}{N_g}\sum_{i:G_i=g} E[Y_{it}(g)|G_i=g] - \frac{1}{N_{g'}}\sum_{i:G_i=g'} E[Y_{it}(g')|G_i=g'] \\
&\quad - \frac{1}{N_g}\sum_{i:G_i=g} E[Y_{it'}(g)|G_i=g] + \frac{1}{N_{g'}}\sum_{i:G_i=g'} E[Y_{it'}(g')|G_i=g'] \\
&= E[Y_{it}(g)|G_i=g] - E[Y_{it}(g')|G_i=g'] - E[Y_{it'}(g)|G_i=g] + E[Y_{it'}(g')|G_i=g']
\end{aligned}$$

Adding/subtracting $E[Y_{it}(\infty)|G_i=g]$ and $E[Y_{it}(\infty)|G_i=g']$

$$\begin{aligned}
E[\hat{\tau}_{gt}(g',t')|\mathcal{D}] &= E[Y_{it}(g) - Y_{it}(\infty)|G_i=g] - E[Y_{it}(g) - Y_{it}(\infty)|G_i=g'] \\
&\quad + E[Y_{it}(\infty)|G_i=g] - E[Y_{it}(\infty)|G_i=g'] - E[Y_{it'}(g)|G_i=g] + E[Y_{it'}(g')|G_i=g']
\end{aligned}$$

Under parallel trends (Assumption 2)

$$E[Y_{it}(\infty)|G_i=g] - E[Y_{it}(\infty)|G_i=g'] = E[Y_{it'}(\infty)|G_i=g] - E[Y_{it'}(\infty)|G_i=g']$$

Therefore,

$$\begin{aligned}
E[\hat{\tau}_{gt}(g',t')|\mathcal{D}] &= E[Y_{it}(g) - Y_{it}(\infty)|G_i=g] - E[Y_{it}(g) - Y_{it}(\infty)|G_i=g'] \\
&\quad + E[Y_{it'}(\infty)|G_i=g] - E[Y_{it'}(\infty)|G_i=g'] - E[Y_{it'}(g)|G_i=g] + E[Y_{it'}(g')|G_i=g'] \\
&= E[Y_{it}(g) - Y_{it}(\infty)|G_i=g] - E[Y_{it}(g) - Y_{it}(\infty)|G_i=g'] \\
&\quad - E[Y_{it'}(g) - Y_{it'}(\infty)|G_i=g] + E[Y_{it'}(g') - Y_{it'}(\infty)|G_i=g']
\end{aligned}$$

By our definition of the group-time ATT,

$$E[\hat{\tau}_{gt}(g', t')|\mathcal{D}] = ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t')$$

B.2 Proof of Lemma 1.1

Consider the static two-way fixed effects regression

$$Y_{it} = \tau D_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

Using Frisch-Waugh-Lovell, we can write the OLS coefficient $\hat{\tau}$ as the bivariate regression coefficient

$$\hat{\tau} = \frac{\sum_{it} Y_{it}(D_{it} - \hat{D}_{it})}{\sum_{it}(D_{it} - \hat{D}_{it})^2}$$

where $\hat{D}_{it}$ are the fitted values from an auxiliary regression of D_{it} on the two-way fixed effects.

$$\hat{D}_{it} = \hat{\tilde{\alpha}}_i + \hat{\tilde{\gamma}}_t$$

Solving the least squares minimization problem for the auxiliary fixed effects.

$$\begin{aligned}
0 &= \sum_t \left(\hat{\tilde{\alpha}}_i + \hat{\tilde{\gamma}}_t - D_{it} \right) \\
T\hat{\tilde{\alpha}}_i &= \sum_t D_{it} - \sum_t \hat{\tilde{\gamma}}_t \\
\hat{\tilde{\alpha}}_i &= \bar{D}_i - \frac{1}{T} \sum_t \hat{\tilde{\gamma}}_t
\end{aligned}$$

and

$$
\begin{aligned}
0 &= \sum_i \left(\hat{\hat{\alpha}}_i + \hat{\hat{\gamma}}_t - D_{it} \right) \\
N\hat{\hat{\gamma}}_t &= \sum_i D_{it} - \sum_i \hat{\hat{\alpha}}_i \\
\hat{\hat{\gamma}}_t &= \bar{D}_t - \frac{1}{N} \sum_i \hat{\hat{\alpha}}_i
\end{aligned}
$$

Substituting:

$$
\begin{aligned}
\hat{\hat{\alpha}}_i &= \bar{D}_i - \frac{1}{T} \sum_t \bar{D}_t + \frac{1}{T} \sum_t \frac{1}{N} \sum_i \hat{\hat{\alpha}}_i \\
&= \bar{D}_i - \bar{\bar{D}} + \frac{1}{N} \sum_i \hat{\hat{\alpha}}_i
\end{aligned}
$$

And finally, substituting into the expression for $\hat{D}_{it}$ yields

$$
\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{\bar{D}}
$$

Therefore, the two-way fixed-effects regression is equivalent to the regression of Y_{it} on the double de-meaned D_{it} treatment indicator

$$
\hat{\tau} = \frac{\sum_{it} Y_{it}(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}})}{\sum_{it}(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}})^2}
$$

Re-writing the numerator:

$$
\begin{aligned}
&\sum_{it} Y_{it}(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}}) \\
&= \sum_{it} Y_{it}D_{it} - \frac{1}{N}\sum_{it}\sum_{i'} Y_{it}D_{i't} - \frac{1}{T}\sum_{it}\sum_{t'} Y_{it}D_{it'} + \frac{1}{NT}\sum_{it}\sum_{i't'} Y_{it}D_{i't'} \\
&= \sum_{it} Y_{it}D_{it} - \frac{1}{N}\sum_{it}\sum_{i'} Y_{i't}D_{it} - \frac{1}{T}\sum_{it}\sum_{t'} Y_{it'}D_{it} + \frac{1}{NT}\sum_{it}\sum_{i't'} Y_{i't'}D_{it} \\
&= \sum_{it} D_{it}\left(Y_{it} - \frac{1}{N}\sum_{i'} Y_{i't} - \frac{1}{T}\sum_{t'} Y_{it'} + \frac{1}{NT}\sum_{i't'} Y_{i't'}\right) \\
&= \sum_{it} D_{it}\left(Y_{it} - \bar{Y}_t - \bar{Y}_i + \bar{\bar{Y}}\right) \\
&= \frac{1}{NT}\sum_{it}\sum_{i't'} D_{it}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)
\end{aligned}
$$

Re-writing the one-way and two-way means over D yields the expression for $\hat{\tau}$ in terms of the component "difference-in-differences"

$$
\hat{\tau} = \frac{\sum_{it}\sum_{i't'} D_{it}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)}{NT\sum_{it}\left(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}}\right)^2}
$$

B.3 Proof of Proposition 2

Starting with the numerator of Lemma 1.1, if $i = i'$ or $t = t'$, then $Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} = 0$, therefore

$$
\hat{\tau} = \frac{\sum_{it}\sum_{i'\neq i, t'\neq t} D_{it}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)}{NT\sum_{it}\left(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}}\right)^2}
$$

Since treatment is binary, we can split the sum into eight terms which denote each possible combination of treatment values for each of the four unit-time combinations that comprise the

difference-in-difference term.

$$
\begin{aligned}
\hat{\tau} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Under staggered adoption, certain treatment/control combinations cannot appear in the data. If a unit i is treated at time t and untreated at time t', this implies that $t' < t$. Therefore, there cannot be another unit i' that is untreated at t and treated at time t'. This eliminates one

combination from the above expression.

$$
\begin{aligned}
\hat{\tau} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Swapping indices i and i' for the third term flips the sign on the differences-in-differences and yields a term that cancels with the fourth term.

$$
\begin{aligned}
\hat{\tau} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Split some of the difference-in-differences terms

$$
\begin{aligned}
\hat{\tau} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} D_{it'} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} \Big) - \Big(D_{it} D_{i't} D_{it'} D_{i't'} \Big) \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} \Big) - \Big(D_{it} (1 - D_{i't}) D_{it'} (1 - D_{i't'}) \Big) \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} D_{i't} (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{it'} \Big) - \Big(D_{it} D_{i't} (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{i't} - Y_{i't'} \Big) \\
& + \sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Swapping indices t and t' for the second part of line 1 and the second part of line 3, and swapping i and i' for the second part of line 4 yields expressions that cancel with the first part of each respective line

This leaves only two remaining sets of terms, those where all other unit-time combinations besides it are under control and those where observation $i't$ is under control and the other unit-time combinations are treated.

$$
\begin{aligned}
\hat{\tau} = & \frac{\sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{i't}) (1 - D_{it'}) (1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)}{NT \sum_{it} \Big(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}} \Big)^2} \\
& + \frac{\sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it} (1 - D_{it'}) D_{i't} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)}{NT \sum_{it} \Big(D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}} \Big)^2}
\end{aligned}
$$

Expanding the denominator:

$$
\begin{aligned}
&NT\sum_{it}(D_{it}-\bar{D}_i-\bar{D}_t+\bar{\bar{D}})^2 \\
&= NT\Bigg[\sum_{it}D_{it}^2+\sum_{it}(\bar{D}_i)^2+\sum_{it}(\bar{D}_t)^2+\sum_{it}(\bar{\bar{D}})^2 \\
&-2\sum_{it}D_{it}\bar{D}_i-2\sum_{it}D_{it}\bar{D}_t+2\sum_{it}D_{it}\bar{\bar{D}}+2\sum_{it}\bar{D}_i\bar{D}_t-2\sum_{it}\bar{D}_i\bar{\bar{D}}-2\sum_{it}\bar{D}_t\bar{\bar{D}}\Bigg] \\
&= NT\Bigg[NT\bar{\bar{D}}+T\sum_{i}(\bar{D}_i)^2+N\sum_{t}(\bar{D}_t)^2+NT(\bar{\bar{D}})^2 \\
&-2T\sum_{i}(\bar{D}_i)^2-2N\sum_{t}(\bar{D}_t)^2+2NT(\bar{\bar{D}})^2+2NT(\bar{\bar{D}})^2-2NT(\bar{\bar{D}})^2-2NT(\bar{\bar{D}})^2\Bigg] \\
&= (NT)^2\bar{\bar{D}}-NT^2\sum_{i}(\bar{D}_i)^2-N^2T\sum_{t}(\bar{D}_t)^2+(NT)^2(\bar{\bar{D}})^2
\end{aligned}
$$

Re-writing the single and double means over D_{it}

$$
\begin{aligned}
&(NT)^2\bar{\bar{D}}-NT^2\sum_{i}(\bar{D}_i)^2-N^2T\sum_{t}(\bar{D}_t)^2+(NT)^2(\bar{\bar{D}})^2 \\
&= NT\Bigg(\sum_{it}D_{it}\Bigg)-N\sum_{i}\Bigg(\sum_{t}D_{it}\Bigg)^2-T\sum_{t}\Bigg(\sum_{i}D_{it}\Bigg)^2+\Bigg(\sum_{it}D_{it}\Bigg)^2 \\
&= \Bigg(\sum_{it}D_{it}\Bigg)\Bigg(NT+\sum_{it}D_{it}\Bigg)-N\sum_{i}\Bigg(\sum_{t}D_{it}\Bigg)^2-T\sum_{t}\Bigg(\sum_{i}D_{it}\Bigg)^2
\end{aligned}
$$

We then re-write sums over i and t in terms of sums over the treatment timing groups g. Under staggered adoption and no treatment reversal, $\sum_t D_{it} = (T+1-G_i)$ or the number of treatment of periods between initiation of treatment (G_i) and the end of the panel (T). Since treatment timing groups are mutually exclusive, we can write the sum over units i as a sum over treatment

timing groups g

$$\begin{aligned}\sum_{it} D_{it} &= \sum_{i}(T+1-G_i) \\ &= \sum_{g=1}^{G}\sum_{i:G_i=g}(T+1-G_i) \\ &= \sum_{g=1}^{G} N_g(T+1-g)\end{aligned}$$

And, similarly

$$\sum_{i}\left(\sum_{t} D_{it}\right)^2 = \sum_{g=1}^{G} N_g(T+1-g)^2$$

Under staggered adoption, $\sum_i D_{it} = \sum_{g\leq t} N_g$ since the total number of units under treatment at time t is equivalent to the number of timing groups that start treatment on or prior to time t.

$$\begin{aligned}\sum_{t}\left(\sum_{i} D_{it}\right)^2 &= \sum_{t}\left(\sum_{g\leq t} N_g\right)^2 \\ &= \sum_{g=1}^{G}\left(\sum_{g'\leq g} N_{g'}\right)^2 \\ &= \sum_{g=1}^{G} N_g^2(T+1-g) + \sum_{g=1}^{G}\sum_{g'<g} N_g N_{g'}(T+1-g) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(T+1-g')\end{aligned}$$

Substituting back into the denominator and combining terms yields

$$
\begin{aligned}
&\Big(\sum_{it} D_{it}\Big)\Big(NT + \sum_{it} D_{it}\Big) - N\sum_{i}\Big(\sum_{t} D_{it}\Big)^2 - T\sum_{t}\Big(\sum_{i} D_{it}\Big)^2 \\
&= \sum_{g=1}^{G} N_g(T+1-g)\Big(NT + \sum_{g'=1}^{G} N_{g'}(T+1-g')\Big) - N\sum_{g=1}^{G} N_g(T+1-g)^2 - T\sum_{g=1}^{G}\Big(\sum_{g'\le g} N_{g'}\Big)^2 \\
&= \sum_{g=1}^{G} N_g(T+1-g)\Big(NT + \sum_{g'=1}^{G} N_{g'}(T+1-g') - NT + N(g-1)\Big) - T\sum_{g=1}^{G}\Big(\sum_{g'\le g} N_{g'}\Big)^2 \\
&= \sum_{g=1}^{G} N_g(T+1-g)\Big(N_\infty(g-1) + \sum_{g'=1}^{G} N_{g'}(T+g-g')\Big) - T\sum_{g=1}^{G}\Big(\sum_{g'\le g} N_{g'}\Big)^2 \\
&= \sum_{g=1}^{G} N_g(T+1-g)\Big(N_\infty(g-1) + \sum_{g'\neq g} N_{g'}(T+g-g')\Big) - \sum_{g=1}^{G}\sum_{g'<g} N_g N_{g'}(T+1-g)T \\
&\qquad - \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(T+1-g')T \\
&= \sum_{g=1}^{G} N_g N_\infty(T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'<g} N_g N_{g'}(T+1-g)(g-g') \\
&\qquad + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}\Big[(T+1-g)(T+1-g') - (T+1-g')T\Big] \\
&= \sum_{g=1}^{G} N_g N_\infty(T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'<g} N_g N_{g'}(T+1-g)(g-g') \\
&\qquad + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}\Big[-g^2 + gg' + g - g'\Big] \\
&= \sum_{g=1}^{G} N_g N_\infty(T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(T+1-g')(g'-g) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(g-1) \\
&= \sum_{g=1}^{G} N_g N_\infty(T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T+1-g') \\
&= \sum_{g=1}^{G} N_g N_\infty(T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)
\end{aligned}
$$

The first term captures all comparisons between each treatment timing group g and the "never-treated" group with $G_i = \infty$. The second captures comparisons between each treatment timing group and the "not-yet-treated" groups relative to that timing group ($g' > g$) that use the $g-1$

pre-treatment periods to de-bias the cross-sectional comparisons. It also includes the "forbidden comparisons" between g and the not-yet-treated periods of g' that use *future* periods as de-biasing second differences.

Turning to the numerator, we start by re-writing the sums indexed by i as sums over different treatment timing groups. Define $\bar{Y}_{g,t} = \frac{1}{N_g} \sum_{i:G_i=g} Y_{it}$ as the mean outcome at time t among units in treatment timing group g.

$$
\begin{aligned}
&\sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it}(1 - D_{i't})(1 - D_{it'})(1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
&= \sum_{g=1}^{G} \sum_{i:G_i=g} \sum_{i':G_{i'}=\infty} \sum_{t,t' \neq t} \Big(D_{it}(1 - D_{i't})(1 - D_{it'})(1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
&\qquad + \sum_{g=1}^{G} \sum_{g' \neq g} \sum_{i:G_i=g} \sum_{i':G_{i'}=g'} \sum_{t,t' \neq t} \Big(D_{it}(1 - D_{i't})(1 - D_{it'})(1 - D_{i't'}) \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
&= \sum_{g=1}^{G} \sum_{t=g}^{T} \sum_{t'=1}^{g-1} N_g N_\infty \Big(\bar{Y}_{g,t} - \bar{Y}_{\infty,t} - \bar{Y}_{g,t'} + \bar{Y}_{\infty,t'} \Big) + \sum_{g=1}^{G} \sum_{g'>g} \sum_{t=g}^{g'-1} \sum_{t'=1}^{g-1} N_g N_{g'} \Big(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'} \Big)
\end{aligned}
$$

where the last line follows from the fact that "never-treated" observations are always under control and observations with $G_i = g' > g$ are under control between periods g and $g'-1$. Under staggered adoption, the only time periods where units in timing group g and units in timing group $g' > g$ are both under control are from 1 to $g-1$.

Similarly, the "forbidden comparisons" can be written as:

$$
\begin{aligned}
&\sum_{it} \sum_{i' \neq i, t' \neq t} \Big(D_{it}(1 - D_{it'}) D_{i't} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
&= \sum_{g=1}^{G} \sum_{g' \neq g} \sum_{i:G_i=g} \sum_{i':G_{i'}=g'} \sum_{t,t' \neq t} \Big(D_{it}(1 - D_{it'}) D_{i't} D_{i't'} \Big) \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
&= \sum_{g=1}^{G} \sum_{g'>g} \sum_{t=g}^{g'-1} \sum_{t'=g'}^{T} N_g N_{g'} \Big(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'} \Big)
\end{aligned}
$$

with the last line again following from the fact that units with treatment timing g and $g' > g$ are *both* under treatment from time periods g' to T.

Which yields a final expression for the static two-way fixed effects estimator as an average over

2×2 difference-in-differences for each treatment timing group that use either the "never-treated" group or a "not-yet-treated" group as the cross-sectional comparison.

$$\hat{\tau} = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\sum_{t'=1}^{g-1} N_g N_\infty \left(\bar{Y}_{g,t} - \bar{Y}_{\infty,t} - \bar{Y}_{g,t'} + \bar{Y}_{\infty,t'}\right) + \sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=1}^{g-1} N_g N_{g'} \left(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=g'}^{T} N_g N_{g'} \left(\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}$$

B.4 Proof of Proposition 3

Conditioning on the distribution of treatment and taking the expectation of $\hat{\tau}$:

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\sum_{t'=1}^{g-1} N_g N_\infty \left(E[\bar{Y}_{g,t} - \bar{Y}_{\infty,t} - \bar{Y}_{g,t'} + \bar{Y}_{\infty,t'}|\mathcal{D}]\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=1}^{g-1} N_g N_{g'} \left(E[\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}|\mathcal{D}]\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=g'}^{T} N_g N_{g'} \left(E[\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}|\mathcal{D}]\right)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}$$

Under parallel trends, each difference-in-difference term identifies a combination of group-time ATTs

$$E[\bar{Y}_{g,t} - \bar{Y}_{\infty,t} - \bar{Y}_{g,t'} + \bar{Y}_{\infty,t'}|\mathcal{D}] = ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t')$$

By definition, the group-time ATT for the "never-treated" history is zero in every time period.

$$
\begin{aligned}
E[\hat{\tau}|\mathcal{D}] =& \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\sum_{t'=1}^{g-1} N_g N_\infty \Big(ATT_g(t) - ATT_g(t') \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)} \\
&+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=1}^{g-1} N_g N_{g'} \Big(ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t') \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)} \\
&+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=g'}^{T} N_g N_{g'} \Big(ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t') \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}
\end{aligned}
$$

No anticipation implies also that $ATT_g(t) = 0$ for all $t < g$. In the first term, all t' in the summation are less than g. In the second, all $t < g'$ and all $t' < g$, $t' < g'$. However, in the third term, $t < g'$, but $t'g$, $t' > g$

$$
\begin{aligned}
E[\hat{\tau}|\mathcal{D}] =& \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\sum_{t'=1}^{g-1} N_g N_\infty \Big(ATT_g(t) \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)} \\
&+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=1}^{g-1} N_g N_{g'} \Big(ATT_g(t) \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)} \\
&+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1}\sum_{t'=g'}^{T} N_g N_{g'} \Big(ATT_g(t) + ATT_{g'}(t') - ATT_g(t') \Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'} (g'-g)(T-g'+g)}
\end{aligned}
$$

There are two constant effects assumptions that allow the static TWFE estimator to identify an average of group-time ATTs. The first is an assumption of constant effects across time within treatment-timing group, $ATT_g(t) = ATT_g(t')$ if $t \geq g$, $t' \geq g$. Under this assumption, treatment effects are not permitted to vary across time but *are* allowed to be heterogeneous across timing

group. Then, the TWFE estimator can be written as:

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T} N_g N_\infty (g-1)\Big(ATT_g(t)\Big) + \sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1} N_g N_{g'}(g-1)\Big(ATT_g(t)\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'>g}\sum_{t'=g'}^{T} N_g N_{g'}(g'-g)\Big(ATT_{g'}(t')\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

Re-arranging terms

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T} N_g N_\infty (g-1)\Big(ATT_g(t)\Big) + \sum_{g=1}^{G}\sum_{g'>g}\sum_{t=g}^{g'-1} N_g N_{g'}(g-1)\Big(ATT_g(t)\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$
$$+ \frac{\sum_{g=1}^{G}\sum_{g'<g}\sum_{t=g}^{T} N_g N_{g'}(g-g')\Big(ATT_g(t)\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

Combining yields an expression of weighted group-time ATTs each with a non-negative weight that sum to 1.

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\left(N_g(g-1)\Big[N_\infty + \sum_{g'>g} N_{g'}\mathbf{1}(t<g')\Big] + N_g\Big[\sum_{g'<g}(g-g')N_{g'}\Big]\right)\times\Big(ATT_g(t)\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

Alternatively, under the assumption of constant effects by calendar time, $ATT_g(t) = ATT_{g'}(t)$ for $t \geq g, t \geq g'$, we have:

$$E[\hat{\tau}|\mathcal{D}] = \frac{\sum_{g=1}^{G}\sum_{t=g}^{T}\left(N_g N_\infty (g-1) + \sum_{g'>g} N_g N_{g'}(T-g'+g)\mathbf{1}(t<g')\right)\times\Big(ATT_g(t)\Big)}{\sum_{g=1}^{G} N_g N_\infty (T+1-g)(g-1) + \sum_{g=1}^{G}\sum_{g'>g} N_g N_{g'}(g'-g)(T-g'+g)}$$

B.5 Proof of Lemma 3.1

Write the dynamic regression in terms of the relative treatment time indicator for q', the relative time period interest, and all other relative treatment times not excluded from the regression $q \neq q', q \neq -1$.

$$Y_{it} = \tau^{(q')} D_{it}^{(q')} + \sum_{q \notin \{q', -1\}} \tau^{(q)} D_{it}^{(q)} + \alpha_i + \gamma_t + \varepsilon_{it}$$

By Frisch-Waugh-Lovell, we can write the OLS estimator of $\tau^{(q')}$ in terms of an auxiliary regression of Y_{it} on $D_{it}^{(q')} - \widehat{D_{it}^{(q')}}$, the residuals from a regression of $D_{it}^{(q')}$ on the other included indicators $D_{it}^{(q)}$ and the two-way fixed effects.

$$\hat{\tau}^{(q')} = \frac{\sum_{it} Y_{it}(D_{it} - \hat{D_{it}^{(q')}})}{\sum_{it}(D_{it} - \hat{D_{it}^{(q')}})^2}$$

where

$$\hat{D_{it}^{(q')}} = \sum_{q \notin \{q', -1\}} \omega_q D_{it}^{(q)} + \hat{\bar{\alpha}}_i + \hat{\bar{\gamma}}_t$$

As in the static regression, write the normal equations for the OLS minimization problem and solve for $\hat{\bar{\alpha}}_i$

$$\begin{aligned}
0 &= 2 \sum_t \left(\hat{\bar{\alpha}}_i + \hat{\bar{\gamma}}_t - D_{it}^{(q')} + \sum_{q \notin \{q', -1\}} \omega_q D_{it}^{(q)} \right) \\
T\hat{\bar{\alpha}}_i &= \sum_t D_{it}^{(q')} - \sum_t \hat{\bar{\gamma}}_t - \sum_{q \notin \{q', -1\}} \omega_q \sum_t D_{it}^{(q)} \\
\hat{\bar{\alpha}}_i &= \bar{D}_i^{(q')} - \frac{1}{T} \sum_t \hat{\bar{\gamma}}_t - \sum_{q \notin \{q', -1\}} \omega_q \bar{D}_i^{(q)}
\end{aligned}$$

And solve for $\hat{\hat{\gamma}}_t$:

$$
\begin{aligned}
0 &= 2\sum_i \left(\hat{\hat{\gamma}}_t + \hat{\hat{\alpha}}_i - D_{it}^{(q')} + \sum_{q \notin \{q',-1\}} \omega_q D_{it}^{(q)} \right) \\
N\hat{\hat{\gamma}}_t &= \sum_i D_{it}^{(q')} - \sum_i \hat{\hat{\alpha}}_i - \sum_{q \notin \{q',-1\}} \omega_q \sum_i D_{it}^{(q)} \\
\hat{\hat{\gamma}}_t &= \bar{D}_t^{(q')} - \frac{1}{N}\sum_i \hat{\hat{\alpha}}_i - \sum_{q \notin \{q',-1\}} \omega_q \bar{D}_t^{(q)}
\end{aligned}
$$

where $\bar{D}_i^{(q)} = \frac{1}{T}\sum_t D_{it}^{(q)}$, $\bar{D}_i^{(q')} = \frac{1}{T}\sum_t D_{it}^{(q')}$, $\bar{D}_t^{(q)} = \frac{1}{N}\sum_i D_{it}^{(q)}$, $\bar{D}_t^{(q')} = \frac{1}{N}\sum_i D_{it}^{(q')}$.

Substituting $\hat{\hat{\alpha}}_i$ in the expression for $\hat{\hat{\gamma}}_t$, we get

$$
\begin{aligned}
\hat{\hat{\gamma}}_t &= \bar{D}_t^{(q')} - \frac{1}{N}\sum_i \left[\bar{D}_i^{(q')} - \frac{1}{T}\sum_t \hat{\hat{\gamma}}_t - \sum_{q \notin \{q',-1\}} \omega_q \bar{D}_i^{(q)} \right] - \sum_{q \notin \{q',-1\}} \omega_q \bar{D}_t^{(q)} \\
&= \bar{D}_t^{(q')} - \bar{\bar{D}}^{(q')} + \frac{1}{NT}\sum_{it} \hat{\hat{\gamma}}_t + \sum_{q \notin \{q',-1\}} \omega_q \bar{\bar{D}}^{(q)} - \sum_{q \notin \{q',-1\}} \omega_q \bar{D}_t^{(q)} \\
&= \bar{D}_t^{(q')} - \bar{\bar{D}}^{(q')} + \frac{1}{T}\sum_t \hat{\hat{\gamma}}_t + \sum_{q \notin \{q',-1\}} \omega_q \bar{\bar{D}}^{(q)} - \sum_{q \notin \{q',-1\}} \omega_q \bar{D}_t^{(q)}
\end{aligned}
$$

where $\bar{\bar{D}}^{(q)} = \frac{1}{NT}\sum_{it} D_{it}^{(q)}$ and $\bar{\bar{D}}^{(q')} = \frac{1}{NT}\sum_{it} D_{it}^{(q')}$.

Further substituting the expressions for $\hat{\hat{\alpha}}_i$ and $\hat{\hat{\gamma}}_t$ into the residuals $D_{it}^{(q')} - \hat{D_{it}^{(q')}}$, we get

$$
D_{it}^{(q')} - \hat{D_{it}^{(q')}} = \left(D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')} \right) - \sum_{q \notin \{q',-1\}} \omega_q \left(D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)} \right)
$$

Finally, substituting this into the FWL bivariate regression, we obtain an expression for $\hat{\tau}^{(q')}$ in terms of a weighted average of "static" two-way fixed effects regression estimators.

$$
\hat{\tau}^{(q')} = \frac{\sum_{it} Y_{it} \left(D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')} \right) - \sum_{q \notin \{q',-1\}} \omega_q \sum_{it} Y_{it} \left(D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)} \right)}{\sum_{it} \left((D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')}) - \sum_{q \neq (q',-1)} \omega_q \left(D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)} \right) \right)^2}
$$

Expanding $\bar{D}_t^{(q')}, \bar{D}_i^{(q')}$ and $\bar{\bar{D}}^{(q')}$ and swapping indices:

$$\begin{aligned}
&\sum_{it} Y_{it}\left(D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')}\right) \\
=&\sum_{it} Y_{it} D_{it}^{(q')} - \frac{1}{N}\sum_{it}\sum_{i'} Y_{it} D_{i't}^{(q')} - \frac{1}{T}\sum_{it}\sum_{t'} Y_{it} D_{it'}^{(q')} + \frac{1}{NT}\sum_{it}\sum_{i't'} Y_{it} D_{i't'}^{(q')} \\
=&\sum_{it} Y_{it} D_{it}^{(q')} - \frac{1}{N}\sum_{it}\sum_{i'} Y_{i't} D_{it}^{(q')} - \frac{1}{T}\sum_{it}\sum_{t'} Y_{it'} D_{it}^{(q')} + \frac{1}{NT}\sum_{it}\sum_{i't'} Y_{i't'} D_{it}^{(q')} \\
=&\sum_{it} D_{it}^{(q')}\left(Y_{it} - \frac{1}{N}\sum_{i'} Y_{i't} - \frac{1}{T}\sum_{t'} Y_{it'} + \frac{1}{NT}\sum_{i't'} Y_{i't'}\right) \\
=&\sum_{it} D_{it}^{(q')}\left(Y_{it} - \bar{Y}_t - \bar{Y}_i + \bar{\bar{Y}}\right) \\
=&\frac{1}{NT}\sum_{it}\sum_{i't'} D_{it}^{(q')}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)
\end{aligned}$$

where the single and double bars again denote single and double averages of Y_{it}.

Substituting back for q' and q in the expression for $\hat{\tau}^{(q')}$ yields our expression for the event study coefficient

$$\hat{\tau}^{(q')} = \frac{\sum_{it}\sum_{i't'} D_{it}^{(q')}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right) - \sum_{q\notin\{q',-1\}} \omega_q \sum_{it}\sum_{i't'} D_{it}^{(q)}\left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\right)}{NT\sum_{it}\left((D_{it}^{(q')} - \bar{D}_t^{(q')} - \bar{D}_i^{(q')} + \bar{\bar{D}}^{(q')}) - \sum_{q\notin\{q',-1\}} \omega_q\left(D_{it}^{(q)} - \bar{D}_t^{(q)} - \bar{D}_i^{(q)} + \bar{\bar{D}}^{(q)}\right)\right)^2}$$

B.6 Proof of Lemma 3.2

Start with the auxiliary regression

$$\widehat{D_{it}^{q'}} = \sum_{q\notin\{q',-1\}} \omega_q D_{it}^{(q)} + \hat{\bar{\alpha}}_i + \hat{\bar{\bar{\gamma}}}_t$$

As shown in our proof in B.5, by FWL, this is equivalent to the regression

$$\widehat{D_{it}^{q'}} = \sum_{q\notin\{q',-1\}} \omega_q\left(D_{it}^{(q)} - \bar{D}_i^{(q)} - \bar{D}_t^{(q)} + \bar{\bar{D}}^{(q)}\right)$$

Write the normal equation for ω_q in terms of all other weights ω_q^*

$$\omega_q = \frac{\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q)}}{\sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)}} - \sum_{q^* \notin \{q', q, -1\}} \omega_{q^*} \frac{\sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q)}}{\sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)}}$$

Write

$$\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q)} = \sum_{it} \left(D_{it}^{(q')} - \bar{D}_i{}^{(q')} - \bar{D}_t{}^{(q')} + \bar{\bar{D}}^{(q')} \right) \left(D_{it}^{(q)} - \bar{D}_i{}^{(q)} - \bar{D}_t{}^{(q)} + \bar{\bar{D}}^{(q)} \right)$$

Expanding

$$\begin{aligned}
&\sum_{it} D_{it}^{(q')} D_{it}^{(q)} - \sum_{it} D_{it}^{(q')} \bar{D}_i^{(q)} - \sum_{it} D_{it}^{(q')} \bar{D}_t^{(q)} + \sum_{it} D_{it}^{(q')} \bar{\bar{D}}^{(q)} \\
&- \sum_{it} \bar{D}_i^{(q')} D_{it}^{(q)} + \sum_{it} \bar{D}_i^{(q')} \bar{D}_i^{(q)} + \sum_{it} \bar{D}_i^{(q')} \bar{D}_t^{(q)} - \sum_{it} \bar{D}_i^{(q')} \bar{\bar{D}}^{(q)} \\
&- \sum_{it} \bar{D}_t^{(q')} D_{it}^{(q)} + \sum_{it} \bar{D}_t^{(q')} \bar{D}_i^{(q)} + \sum_{it} \bar{D}_t^{(q')} \bar{D}_t^{(q)} - \sum_{it} \bar{D}_t^{(q')} \bar{\bar{D}}^{(q)} \\
&+ \sum_{it} \bar{\bar{D}}^{(q')} D_{it}^{(q)} - \sum_{it} \bar{\bar{D}}^{(q')} \bar{D}_i^{(q)} - \sum_{it} \bar{\bar{D}}^{(q')} \bar{D}_t^{(q)} + \sum_{it} \bar{\bar{D}}^{(q')} \bar{\bar{D}}^{(q)}
\end{aligned}$$

Simplifying for $q \neq q'$

$$\begin{aligned}
&- \sum_{i} \bar{D}_i^{(q)} \sum_{t} D_{it}^{(q')} - \sum_{t} \bar{D}_t^{(q)} \sum_{i} D_{it}^{(q')} + \bar{\bar{D}}^{(q)} \sum_{it} D_{it}^{(q')} \\
&- \sum_{i} \bar{D}_i^{(q')} \sum_{t} D_{it}^{(q)} + T \sum_{i} \bar{D}_i^{(q')} \bar{D}_i^{(q)} + \sum_{i} \bar{D}_i^{(q')} \sum_{t} \bar{D}_t^{(q)} - \bar{\bar{D}}^{(q)} T \sum_{i} \bar{D}_i^{(q')} \\
&- \sum_{t} \bar{D}_t^{(q')} \sum_{i} D_{it}^{(q)} + \sum_{t} \bar{D}_t^{(q')} \sum_{i} \bar{D}_i^{(q)} + N \sum_{t} \bar{D}_t^{(q')} \bar{D}_t^{(q)} - \bar{\bar{D}}^{(q)} N \sum_{t} \bar{D}_t^{(q')} \\
&+ \bar{\bar{D}}^{(q')} \sum_{it} D_{it}^{(q)} - \bar{\bar{D}}^{(q')} T \sum_{i} \bar{D}_i^{(q)} - \bar{\bar{D}}^{(q')} N \sum_{t} \bar{D}_t^{(q)} + NT \bar{\bar{D}}^{(q')} \bar{\bar{D}}^{(q)}
\end{aligned}$$

Definitions of the one-way and two-way means

$$
\begin{aligned}
&- T\sum_i \bar{D}_i^{(q)}\bar{D}_i^{(q')} - N\sum_t \bar{D}_t^{(q)}\bar{D}_t^{(q')} + NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} \\
&- T\sum_i \bar{D}_i^{(q')}\bar{D}_i^{(q)} + T\sum_i \bar{D}_i^{(q')}\bar{D}_i^{(q)} + NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} - NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} \\
&- N\sum_t \bar{D}_t^{(q')}\bar{D}_t^{(q)} + NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} + N\sum_t \bar{D}_t^{(q')}\bar{D}_t^{(q)} - NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} \\
&+ NT\bar{\bar{D}}^{(q')}\bar{\bar{D}}^{(q)} - NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} - NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')} + NT\bar{\bar{D}}^{(q')}\bar{\bar{D}}^{(q)}
\end{aligned}
$$

Collecting terms

$$
\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q)} = -T\sum_i \bar{D}_i^{(q)}\bar{D}_i^{(q')} - N\sum_t \bar{D}_t^{(q)}\bar{D}_t^{(q')} + NT\bar{\bar{D}}^{(q)}\bar{\bar{D}}^{(q')}
$$

Since $\bar{\bar{D}}^{(q)} = \frac{\sum_{g\in\mathcal{G}_q} N_g}{NT}$

$$
\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q)} = -T\sum_i \bar{D}_i^{(q)}\bar{D}_i^{(q')} - N\sum_t \bar{D}_t^{(q)}\bar{D}_t^{(q')} + \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q'}} N_g\right)}{NT}
$$

Next, we re-write $\sum_i \bar{D}_i^{(q)}\bar{D}_i^{(q')}$ in terms of the number of observations in each timing group N_g. $\bar{D}_i^{(q)}\bar{D}_i^{(q')}$ is non-zero only if both q and q' are in the set of relative time indicators $\mathcal{Q}_g$ for treatment timing group g and otherwise takes on a value of $\frac{1}{T^2}$ since each relative time indicator is equal to 1 for only a single time period.

$$
\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q)} = -\frac{\sum_{g\in G_q\cap G_{q'}} N_g}{T} - N\sum_t \bar{D}_t^{(q)}\bar{D}_t^{(q')} + \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q'}} N_g\right)}{NT}
$$

Let $N_{g+q-q'} = 0$ for $g+q-q' > T$ or $g+q-q' < 1$ as there are no observations where the treatment timing group is greater than T (aside from the pure controls) or less than 1.

$$
\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q)} = -\frac{\sum_{g\in G_q\cap G_{q'}} N_g}{T} - \frac{\sum_{g\in G_q} N_g N_{g+q-q'}}{N} + \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q'}} N_g\right)}{NT}
$$

Take out the common denominator:

$$\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q)} = \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q'}} N_g\right) - T\sum_{g\in G_q} N_g N_{g+q-q'} - N\sum_{g\in G_q\cap G_{q'}} N_g}{NT}$$

For the case where $q' = q$:

$$\begin{aligned}
\sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)} &= \frac{NT\sum_{g\in G_q} N_g + \left(\sum_{g\in\mathcal{G}_q} N_g\right)^2 - T\sum_{g\in\mathcal{G}_q} N_g^2 - N\sum_{g\in G_q} N_g}{NT} \\
&= \frac{\left(\sum_{g\in\mathcal{G}_q} N_g - N + NT\right)\left(\sum_{g\in\mathcal{G}_q} N_g\right) - T\sum_{g\in\mathcal{G}_q} N_g^2}{NT} \\
&= \frac{\left(\sum_{g\in\mathcal{G}_q} N_g + N(T-1)\right)\left(\sum_{g\in\mathcal{G}_q} N_g\right) - T\sum_{g\in\mathcal{G}_q} N_g^2}{NT} \\
&= \frac{(T-1)\left[\sum_{g\in G_q} N_g(N-N_g)\right] + \left(\sum_{g\in\mathcal{G}_q} N_g\right)^2 - \sum_{g\in\mathcal{G}_q} N_g^2}{NT}
\end{aligned}$$

Substituting into the expression for ω_q

$$\begin{aligned}
\omega_q =& \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q'}} N_g\right) - T\sum_{g\in G_q} N_g N_{g+q-q'} - N\sum_{g\in G_q\cap G_{q'}} N_g}{(T-1)\left[\sum_{g\in G_q} N_g(N-N_g)\right] + \left(\sum_{g\in\mathcal{G}_q} N_g\right)^2 - \sum_{g\in\mathcal{G}_q} N_g^2} \\
&- \sum_{q^*\notin\{q',q,-1\}} \omega_{q^*} \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q^*}} N_g\right) - T\sum_{g\in G_q} N_g N_{g+q-q^*} - N\sum_{g\in G_q\cap G_{q^*}} N_g}{(T-1)\left[\sum_{g\in G_q} N_g(N-N_g)\right] + \left(\sum_{g\in\mathcal{G}_q} N_g\right)^2 - \sum_{g\in\mathcal{G}_q} N_g^2}
\end{aligned}$$

With $\omega_{q'} = -1$, $\omega_{-1} = 0$

$$\omega_q = -\sum_{q^*\neq q} \omega_{q^*} \frac{\left(\sum_{g\in\mathcal{G}_q} N_g\right)\left(\sum_{g\in\mathcal{G}_{q^*}} N_g\right) - T\sum_{g\in G_q} N_g N_{g+q-q^*} - N\sum_{g\in G_q\cap G_{q^*}} N_g}{(T-1)\left[\sum_{g\in G_q} N_g(N-N_g)\right] + \left(\sum_{g\in\mathcal{G}_q} N_g\right)^2 - \sum_{g\in\mathcal{G}_q} N_g^2}$$

Or equivalently

$$\omega_q\Bigg((T-1)\Big[\sum_{g\in G_q} N_g(N-N_g)\Big]+\Big(\sum_{g\in\mathcal{G}_q} N_g\Big)^2-\sum_{g\in\mathcal{G}_q} N_g^2\Bigg)=$$
$$-\sum_{q^*\neq q}\omega_{q^*}\Bigg(\Big(\sum_{g\in\mathcal{G}_q} N_g\Big)\Big(\sum_{g\in\mathcal{G}_{q^*}} N_g\Big)-T\sum_{g\in G_q} N_g N_{g+q-q^*}-N\sum_{g\in G_q\cap G_{q^*}} N_g\Bigg)$$

B.7 Proof of Lemma 3.3

First, we expand the square in the denominator of $\hat{\tau}^{(q')}$

$$NT\sum_{it}\Big(\tilde{D}_{it}^{(q')}-\sum_{q\notin\{q',-1\}}\omega_q\tilde{D}_{it}^{(q)}\Big)^2=NT\Big[\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q')}-2\sum_{q\notin\{q',-1\}}\omega_q\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q')}+\sum_{it}\Big(\sum_{q\notin\{q',-1\}}\omega_q\tilde{D}_{it}^{(q)}\Big)^2\Big]$$

The third term in the square brackets can be further expanded as:

$$\sum_{it}\Big(\sum_{q\notin\{q',-1\}}\omega_q\sum_{it}\tilde{D}_{it}^{(q)}\Big)^2=\sum_{q\notin\{q',-1\}}\omega_q\omega_q\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q)}+\sum_{q\notin\{q',-1\}}\sum_{q^*\notin\{q,q',-1\}}\omega_q\omega_{q^*}\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q^*)}$$

Using our result from the normal equations in Appendix B.6

$$\omega_q\Big(\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q)}\Big)=\Big(\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q')}\Big)-\sum_{q^*\notin\{q,q',-1\}}\omega_{q^*}\Big(\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q^*)}\Big)$$

Substituting and simplifying

$$\sum_{it}\Big(\sum_{q\notin\{q',-1\}}\omega_q\sum_{it}\tilde{D}_{it}^{(q)}\Big)^2=\sum_{q\notin\{q',-1\}}\omega_q\Big(\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q')}\Big)$$

And substituting back into the denominator, we obtain

$$NT\Big[\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q')}-\sum_{q\notin\{q',-1\}}\omega_q\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q')}\Big]$$

Using our expressions for $\sum_{it}\tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q')}$ and $\sum_{it}\tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q')}$ from Appendix B.6, we can also

write the denominator as:

$$\Big\{(T-1)\Big[\sum_{g\in\mathcal{G}_{q'}} N_g(N-N_g)\Big] + \Big(\sum_{g\in\mathcal{G}_{q'}} N_g\Big)^2 - \sum_{g\in\mathcal{G}_{q'}} N_g^2\Big\} -$$
$$\sum_{q\notin\{q',-1\}} \omega_q \Big\{\Big(\sum_{g\in\mathcal{G}_q} N_g\Big)\Big(\sum_{g\in\mathcal{G}_{q'}} N_g\Big) - T\sum_{g\in\mathcal{G}_q} N_g N_{g+q-q'} - N\sum_{g\in\mathcal{G}_q,\mathcal{G}_{q'}} N_g\Big\}$$

B.8 Proof of Proposition 4

Start with the numerator from the result of Lemma 3.1:

$$\hat{\tau}^{(q')} \propto \sum_{it}\sum_{i't'} D_{it}^{(q')}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) - \sum_{q\notin\{q',-1\}} \omega_q \sum_{it}\sum_{i't'} D_{it}^{(q)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big)$$

If $i = i'$, or $t = t'$, then $Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} = 0$, therefore we can write

$$\hat{\tau}^{(q')} \propto \sum_{it}\sum_{i'\neq i,t'\neq t} D_{it}^{(q')}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) - \sum_{q\notin\{q',-1\}} \omega_q \sum_{it}\sum_{i'\neq i,t'\neq t} D_{it}^{(q)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big)$$

In the conventional event study regression where each relative treatment time indicator $D_{it}^{(q)}$ only takes on a value of 1 for a single time period, $D_{it}^{(q)} = 1$ implies $D_{it'}^{(q)} = 0$ for all $t' \neq t$.

As in the the proof for Lemma 2, we can split each sum along the other treatment indicators $D_{it}^{(q)}$

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it}\sum_{i'\neq i,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& + \sum_{q\notin\{q',-1\}}\sum_{it}\sum_{i'\neq i,t'\neq t} D_{it}^{(q')} D_{it'}^{(q)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& - \sum_{q\notin\{q',-1\}}\sum_{it}\sum_{i'\neq i,t'\neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& - \sum_{q\notin\{q',-1\}}\sum_{it}\sum_{i'\neq i,t'\neq t} \omega_q D_{it}^{(q)} D_{it}^{(q')}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& - \sum_{q\notin\{q',-1\}}\sum_{q^*\notin\{q',q,-1\}}\sum_{it}\sum_{i'\neq i,t'\neq t} \omega_q D_{it}^{(q)} D_{it'}^{(q^*)}\Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big)
\end{aligned}$$

Combining terms, we have four differences-in-differences comparisons.

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& + \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} (1 + \omega_q) D_{it}^{(q')} D_{it'}^{(q)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& - \sum_{q \notin \{q', -1\}} \sum_{q^* \notin \{q', q, -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} \omega_q D_{it}^{(q)} D_{it'}^{(q^*)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right)
\end{aligned}$$

Split the last term and swap indices (re-include the zero terms where $t' = t$ and $q^* = q$)

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& + \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} (1 + \omega_q) D_{it}^{(q')} D_{it'}^{(q)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \left(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \right) \\
& - \sum_{q \notin \{q', -1\}} \sum_{q^* \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t'} \omega_q D_{it}^{(q)} D_{it'}^{(q^*)} \left(Y_{it} - Y_{i't} \right) \\
& + \sum_{q \notin \{q', -1\}} \sum_{q^* \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t'} \omega_{q^*} D_{it'}^{(q^*)} D_{it}^{(q)} \left(Y_{it} - Y_{i't} \right)
\end{aligned}$$

Factoring

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} (1 + \omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i} D_{it}^{(q)} \Big(Y_{it} - Y_{i't} \Big) \omega_q \sum_{q^* \notin \{q', -1\}} \sum_{t'} D_{it'}^{(q^*)} \\
& + \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i} D_{it}^{(q)} \Big(Y_{it} - Y_{i't} \Big) \sum_{q^* \notin \{q', -1\}} \omega_{q^*} \sum_{t'} D_{it'}^{(q^*)}
\end{aligned}$$

Combine terms

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i' \neq i, t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} (1 + \omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i, t' \neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{q \notin \{q', -1\}} \sum_{it} \sum_{i' \neq i} D_{it}^{(q)} \Big(Y_{it} - Y_{i't} \Big) \Big(\omega_q \sum_{q^* \notin \{q', -1\}} \sum_{t'} D_{it'}^{(q^*)} - \sum_{q^* \notin \{q', -1\}} \omega_{q^*} \sum_{t'} D_{it'}^{(q^*)} \Big)
\end{aligned}$$

Define $\Omega_g = \sum_{q \notin \{q', -1\}} \omega_q \mathbf{1}(q \in \mathcal{Q}_g)$ as the sum over the weights on each of the other relative time periods in treatment timing group g aside from the period of interest q' (if present in g) and

the baseline period -1:

$$\begin{aligned}\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i'\neq i, t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i} D_{it}^{(q)} \Big(Y_{it} - Y_{i't} \Big) \Big(\omega_q (T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})) - \Omega_{G_i} \Big)\end{aligned}$$

Since $D_{it'}^{(-1)}$ takes on a value of 1 for exactly one time period t across any unit i, we can multiply any expression by $\sum_{t'} D_{it'}^{(-1)} = 1$. Multiplying by 1 and adding zero:

$$\begin{aligned}\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i'\neq i, t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} \omega_q D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} \Big(\omega_q (T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})) - \Omega_{G_i} \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} \Big(\omega_q (T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})) - \Omega_{G_i} \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it'} + Y_{i't'} \Big)\end{aligned}$$

Swap indices and sum through

$$\begin{aligned}\hat{\tau}^{(q')} \propto & \sum_{it} \sum_{i'\neq i, t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i, t'\neq t} \Big(\Omega_{G_i} - \omega_q [T - \mathbf{1}(q' \in \mathcal{Q}_{G_i})] \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\ & - \sum_{it} \sum_{i'\neq i} \Big((T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})) \sum_{q\notin\{q',-1\}} \sum_{t'} \omega_q D_{it'}^{(q)} - \Omega_{G_i} \sum_{q\notin\{q',-1\}} \sum_{t'} D_{it'}^{(q)} \Big) D_{it}^{(-1)} \Big(Y_{it} + Y_{i't} \Big)\end{aligned}$$

By the definition of Ω_{G_i}, the last term is equal to zero since $\sum_{g\notin\{q',-1\}}\sum_{t'}\omega_q D_{it'}^{(q)} = \Omega_{G_i}$ and $\sum_{g\notin\{q',-1\}}\sum_{t'} D_{it'}^{(q)} = T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})$

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{it}\sum_{i'\neq i,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i,t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{q\notin\{q',-1\}} \sum_{it} \sum_{i'\neq i,t'\neq t} \Big(\Omega_{G_i} - \omega_q [T - \mathbf{1}(q' \in \mathcal{Q}_{G_i})] \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Split the last sum between units i that contain q' in $\mathcal{Q}_{G_i}$ and those that do not.

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - \omega_q T + \omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\notin\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Split the third term

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - \omega_q T + \omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} \Big) \\
& - \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - \omega_q T + \omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{i:q'\notin\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Again using $\sum_{q \notin q',-1} \sum_t D_{it}^{(q)} = T - 1 - \mathbf{1}(q' \in \mathcal{Q}_{G_i})$ and multiplying by 1

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{i' \neq i} \sum_{t,t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(\Omega_{G_i} - \omega_q T + \omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} \Big) \\
& - \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(- \Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{i:q' \notin \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Swapping indices and dividing/multiplying by $\sum_t D_{it}^{(-1)} = \sum_t D_{it}^{(q')} = 1$

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{i' \neq i} \sum_{t,t' \neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} (1+\omega_q) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(- \Omega_{G_i} + \omega_q T - \omega_q \Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& - \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(- \Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{i:q' \notin \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i' \neq i} \sum_{t,t' \neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Adding 0 and combining terms again

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(1 - \Omega_{G_i} + T\omega_q \Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& - \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(-\Omega_{G_i} + \omega_q T - \omega_q \Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} \Big) \\
& - \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(-\Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{i:q'\notin\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Sum over q and multiply by $\sum_i D_{it}^{(-1)} = 1$ again

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(1 - \Omega_{G_i} + T\omega_q \Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(-\Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} \Big) \\
& - \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(-\Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it'} - Y_{i't'} \Big) \\
& + \sum_{i:q'\notin\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

Combining terms yields an expression in terms of three sets of DiDs:

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(1 - \Omega_{G_i} \Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\in\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(1 - \Omega_{G_i} + T\omega_q \Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big) \\
& + \sum_{i:q'\notin\mathcal{Q}_{G_i}} \sum_{q\notin\{q',-1\}} \sum_{i'\neq i} \sum_{t,t'\neq t} \Big(\Omega_{G_i} - T\omega_q \Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'} \Big)
\end{aligned}
$$

All terms where $G_{i'} = G_i$ cancel due to symmetry since each timing group has the same calendar time/relative time pattern. Therefore each i' matched to i is either never-treated or has a different timing group $G_{i'} \neq G_i$

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{i':G_{i'} \neq G_i} \sum_{t,t' \neq t} \Big(1 - \Omega_{G_i}\Big) D_{it}^{(q')} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& + \sum_{i:q' \in \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i':G_{i'} \neq G_i} \sum_{t,t' \neq t} \Big(1 - \Omega_{G_i} + T\omega_q\Big) D_{it}^{(q')} D_{it'}^{(q)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big) \\
& + \sum_{i:q' \notin \mathcal{Q}_{G_i}} \sum_{q \notin \{q',-1\}} \sum_{i':G_{i'} \neq G_i} \sum_{t,t' \neq t} \Big(\Omega_{G_i} - T\omega_q\Big) D_{it}^{(q)} D_{it'}^{(-1)} \Big(Y_{it} - Y_{i't} - Y_{it'} + Y_{i't'}\Big)
\end{aligned}$$

Re-write in terms of sums over treatment timing groups g and g' and recognizing that the product of relative treatment time indicators equals 1 for only a single pair of t and t'

$$\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{g \in \mathcal{G}_{q'}} \sum_{i:G_i=g} \sum_{i':G_{i'}=\infty} \Big(1 - \Omega_g\Big) \Big(Y_{i(g+q')} - Y_{i'(g+q')} - Y_{i(g-1)} + Y_{i'(g-1)}\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \sum_{i:G_i=g} \sum_{i':G_{i'}=g'} \Big(1 - \Omega_g\Big) \Big(Y_{i(g+q')} - Y_{i'(g+q')} - Y_{i(g-1)} + Y_{i'(g-1)}\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{i:G_i=g} \sum_{i':G_{i'}=\infty} \Big(1 - \Omega_g + T\omega_q\Big) \Big(Y_{i(g+q')} - Y_{i'(g+q')} - Y_{i(g+q)} + Y_{i'(g+q)}\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \sum_{i:G_i=g} \sum_{i':G_{i'}=g'} \Big(1 - \Omega_g + T\omega_q\Big) \Big(Y_{i(g+q')} - Y_{i'(g+q')} - Y_{i(g+q)} + Y_{i'(g+q)}\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{i:G_i=g} \sum_{i':G_{i'}=\infty} \Big(\Omega_g - T\omega_q\Big) \Big(Y_{i(g+q)} - Y_{i'(g+q)} - Y_{i(g-1)} + Y_{i'(g-1)}\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \sum_{i:G_i=g} \sum_{i':G_{i'}=g'} \Big(\Omega_g - T\omega_q\Big) \Big(Y_{i(g+q)} - Y_{i'(g+q)} - Y_{i(g-1)} + Y_{i'(g-1)}\Big)
\end{aligned}$$

Re-write in terms of averages of the outcome within each timing group $\bar{Y}_{g,t} = \frac{1}{N_g} \sum_{i:G_i=g} Y_{it}$

$$
\begin{aligned}
\hat{\tau}^{(q')} \propto & \sum_{g \in \mathcal{G}_{q'}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g + T\omega_q \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{\infty,g+q} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g + T\omega_q \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{g',g+q} \Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty \Big) \Big(\Omega_g - T\omega_q \Big) \Big(\bar{Y}_{g,g+q} - \bar{Y}_{\infty,g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(\Omega_g - T\omega_q \Big) \Big(\bar{Y}_{g,g+q} - \bar{Y}_{g',g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big)
\end{aligned}
$$

Including the denominator from Lemma 3.3 yields the final decomposition in terms of six

difference-in-difference terms

$$
\begin{aligned}
\hat{\tau}^{(q')} = \Bigg[& \sum_{g \in \mathcal{G}_{q'}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty \Big) \Big(1 - \Omega_g + T\omega_q \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{\infty,g+q} \Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(1 - \Omega_g + T\omega_q \Big) \Big(\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{g',g+q} \Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty \Big) \Big(\Omega_g - T\omega_q \Big) \Big(\bar{Y}_{g,g+q} - \bar{Y}_{\infty,g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1} \Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'} \Big) \Big(\Omega_g - T\omega_q \Big) \Big(\bar{Y}_{g,g+q} - \bar{Y}_{g',g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1} \Big) \Bigg] \times \\
& \Bigg[(T-1) \Bigg[\sum_{g \in \mathcal{G}_{q'}} N_g (N - N_g) \Bigg] + \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g \Bigg)^2 - \sum_{g \in \mathcal{G}_{q'}} N_g^2 \\
& - \sum_{q \notin \{q',-1\}} \omega_q \Bigg\{ \Bigg(\sum_{g \in \mathcal{G}_q} N_g \Bigg) \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g \Bigg) - T \sum_{g \in \mathcal{G}_q} N_g N_{g+q-q'} - N \sum_{g \in \mathcal{G}_q \cap \mathcal{G}_{q'}} N_g \Bigg\} \Bigg]^{-1}
\end{aligned}
$$

B.9 Proof of Lemma 4.1

From Lemma 3.2, we have

$$
\omega_q = - \sum_{q^* \neq q} \omega_{q^*} \frac{\Big(\sum_{g \in \mathcal{G}_q} N_g \Big) \Big(\sum_{g \in \mathcal{G}_{q^*}} N_g \Big) - T \sum_{g \in G_q} N_g N_{g+q-q^*} - N \sum_{g \in G_q \cap G_{q^*}} N_g}{(T-1) \Big[\sum_{g \in G_q} N_g (N - N_g) \Big] + \Big(\sum_{g \in \mathcal{G}_q} N_g \Big)^2 - \sum_{g \in \mathcal{G}_q} N_g^2}
$$

In the absence of staggering, $\mathcal{G}_q = \mathcal{G}_{q^*} = \{g^*\}$ for all $q, q^* \neq \infty$. Likewise, $N_{g+q-q^*} = 0$ for $q \neq q^*$ as there is only one treatment timing group. This yields:

$$
\omega_q = - \sum_{q^* \neq q} \omega_{q^*} \frac{N_{g^*}^2 - N N_{g^*}}{\Big(T - 1 \Big) \Big(N_{g^*} (N - N_{g^*}) \Big)}
$$

Factoring the numerator:

$$\omega_q = \sum_{q^* \neq q} \omega_{q^*} \frac{\Big(N_{g^*}(N - N_{g^*})\Big)}{\Big(T-1\Big)\Big(N_{g^*}(N - N_{g^*})\Big)}$$

Cancelling yields:

$$\omega_q = \frac{1}{T-1} \sum_{q^* \neq q} \omega_{q^*}$$

Since we fix $\omega_{q'} = -1$, $\omega_{-1} = 0$:

$$\omega_q = \frac{1}{T-1} \sum_{q^* \notin \{q, q', -1\}} \omega_{q^*} - \frac{1}{T-1}$$

By symmetry, for $q, q^* \notin \{q', -1\}$, we have $\omega_q = \omega_q^*$. Additionally, under no staggering, there are T total relative time periods, thus

$$\omega_q = \frac{T-3}{T-1}\omega_q - \frac{1}{T-1}$$

Solving ω_q, we have:

$$\begin{aligned} \frac{1}{T-1} &= \frac{T-3-T+1}{T-1}\omega_q \\ \frac{1}{T-1} &= \frac{-2}{T-1}\omega_q \\ \omega_q &= -\frac{1}{2} \end{aligned}$$

We can use this result to simplify the expression of the denominator in Lemma 3.3. Note that $N_g = 0$ for $g \neq g^*$ in the case of no staggering ($G_i \in \{g^*, \infty\}$). In this setting, the denominator reduces to:

$$\frac{T N_g^* N_\infty}{2}$$

B.10 Proof of Proposition 5

Without treatment staggering, $G_i = g^*$ or $G_i = \infty$. Therefore, we can write $\hat{\tau}^{(q')}$ as

$$\begin{aligned}\hat{\tau}^{(q')} = &\Bigg[\Bigg(N_{g^*}N_\infty\Bigg)\Bigg(1-\Omega_{g^*}\Bigg)\Bigg(\bar{Y}_{g^*,g^*+q'} - \bar{Y}_{\infty,g^*+q'} - \bar{Y}_{g^*,g^*-1} + \bar{Y}_{\infty,g^*-1}\Bigg) \\ &+ \sum_{\substack{q\in\mathcal{Q}_{g^*} \\ q\notin\{q',-1\}}} \Bigg(N_{g^*}N_\infty\Bigg)\Bigg(1-\Omega_{g^*}+T\omega_q\Bigg)\Bigg(\bar{Y}_{g^*,g^*+q'} - \bar{Y}_{\infty,g^*+q'} - \bar{Y}_{g^*,g^*+q} + \bar{Y}_{\infty,g^*+q}\Bigg)\Bigg]\times \\ &\Bigg[(T-1)\Bigg[N_{g^*}N_\infty\Bigg] - \sum_{q\notin\{q',-1\}}\omega_q\Bigg(N_{g^*}^2 - NN_{g^*}\Bigg)\Bigg]^{-1}\end{aligned}$$

From Lemma 4.1, $\omega_q = -\frac{1}{2},\ \forall q \notin \{q', -1\}$. By extension, $\Omega_{g^*} = -\frac{(T-2)}{2}$ and $1 - \Omega_{g^*} + T\omega_q = 0$

$$\begin{aligned}\hat{\tau}^{(q')} = &\Bigg(N_{g^*}N_\infty\Bigg)\Bigg(\frac{T}{2}\Bigg)\Bigg(\bar{Y}_{g^*,g^*+q'} - \bar{Y}_{\infty,g^*+q'} - \bar{Y}_{g^*,g^*-1} + \bar{Y}_{\infty,g^*-1}\Bigg)\Bigg]\times \\ &\Bigg[(T-1)\Bigg[N_{g^*}N_\infty\Bigg] + \Omega_{g^*}\Bigg(N_{g^*}N_\infty\Bigg)\Bigg]^{-1}\end{aligned}$$

Simplifying the denominator and cancelling yields the standard non-parametric DiD estimator

$$\begin{aligned}\hat{\tau}^{(q')} &= \Bigg(N_{g^*}N_\infty\Bigg)\Bigg(\frac{T}{2}\Bigg)\Bigg(\bar{Y}_{g^*,g^*+q'} - \bar{Y}_{\infty,g^*+q'} - \bar{Y}_{g^*,g^*-1} + \bar{Y}_{\infty,g^*-1}\Bigg)\Bigg] \times \Bigg[\frac{T}{2}\Bigg(N_{g^*}N_\infty\Bigg)\Bigg]^{-1} \\ &= \bar{Y}_{g^*,g^*+q'} - \bar{Y}_{\infty,g^*+q'} - \bar{Y}_{g^*,g^*-1} + \bar{Y}_{\infty,g^*-1}\end{aligned}$$

Or equivalently, using the definition of $\bar{Y}_{g,t}$

$$\hat{\tau}^{(q')} = \Bigg[\frac{1}{N_{g^*}}\sum_{i:G_i=g^*} Y_{i(g^*+q')} - \frac{1}{N_{g^*}}\sum_{i:G_i=g^*} Y_{i(g^*-1)}\Bigg] - \Bigg[\frac{1}{N_\infty}\sum_{i:G_i=\infty} Y_{i(g^*+q')} - \frac{1}{N_\infty}\sum_{i:G_i=\infty} Y_{i(g^*-1)}\Bigg]$$

B.11 Proof of Proposition 6

Start with the expression for $\hat{\tau}^{(q')}$ in terms of difference-in-differences. We'll focus on the numerator as the denominator contains only treatment indicators and is treated as a constant

when we take the expectation conditional on $\mathcal{D}$

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] = \Bigg[& \sum_{g \in \mathcal{G}_{q'}} \Big(N_g N_\infty\Big)\Big(1-\Omega_g\Big)\Big(E[\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1}|\mathcal{D}]\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(1-\Omega_g\Big)\Big(E[\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1}|\mathcal{D}]\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty\Big)\Big(1-\Omega_g+T\omega_q\Big)\Big(E[\bar{Y}_{g,g+q'} - \bar{Y}_{\infty,g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{\infty,g+q}|\mathcal{D}]\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(1-\Omega_g+T\omega_q\Big)\Big(E[\bar{Y}_{g,g+q'} - \bar{Y}_{g',g+q'} - \bar{Y}_{g,g+q} + \bar{Y}_{g',g+q}|\mathcal{D}]\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty\Big)\Big(\Omega_g-T\omega_q\Big)\Big(E[\bar{Y}_{g,g+q} - \bar{Y}_{\infty,g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{\infty,g-1}|\mathcal{D}]\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(\Omega_g-T\omega_q\Big)\Big(E[\bar{Y}_{g,g+q} - \bar{Y}_{g',g+q} - \bar{Y}_{g,g-1} + \bar{Y}_{g',g-1}|\mathcal{D}]\Big)\Bigg] \times \\
& \Bigg[(T-1)\Bigg[\sum_{g \in \mathcal{G}_{q'}} N_g(N-N_g)\Bigg] + \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g\Bigg)^2 - \sum_{g \in \mathcal{G}_{q'}} N_g^2 \\
& - \sum_{q \notin \{q',-1\}} \omega_q \Bigg\{\Bigg(\sum_{g \in \mathcal{G}_q} N_g\Bigg)\Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g\Bigg) - T\sum_{g \in \mathcal{G}_q} N_g N_{g+q-q'} - N\sum_{g \in \mathcal{G}_q \cap \mathcal{G}_{q'}} N_g\Bigg\}\Bigg]^{-1}
\end{aligned}
$$

From Proposition 1, each DiD comparison identifies a combination of group-time ATTs under parallel trends.

$$
E[\bar{Y}_{g,t} - \bar{Y}_{g',t} - \bar{Y}_{g,t'} + \bar{Y}_{g',t'}|\mathcal{D}] = ATT_g(t) - ATT_{g'}(t) - ATT_g(t') + ATT_{g'}(t')
$$

The group-time ATT of being assigned to the "never-treated" group is, by construction, zero. Additionally, under the no-anticipation assumption with respect to the held-out (baseline) period,

$ATT_g(g-1) = 0$. Substituting these effects into the decomposition

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] = \Bigg[& \sum_{g \in \mathcal{G}_{q'}} \Big(N_g N_\infty\Big)\Big(1-\Omega_g\Big)\Big(ATT_g(g+q')\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(1-\Omega_g\Big)\Big(ATT_g(g+q') - ATT_{g'}(g+q') + ATT_{g'}(g-1)\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty\Big)\Big(1-\Omega_g+T\omega_q\Big)\Big(ATT_g(g+q') - ATT_g(g+q)\Big) \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(1-\Omega_g+T\omega_q\Big)\Big(ATT_g(g+q') - ATT_{g'}(g+q') - ATT_g(g+q) + ATT_{g'}(g+q)\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \Big(N_g N_\infty\Big)\Big(\Omega_g-T\omega_q\Big)\Big(ATT_g(g+q)\Big) \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q',-1\}}} \sum_{g' \neq g} \Big(N_g N_{g'}\Big)\Big(\Omega_g-T\omega_q\Big)\Big(ATT_g(g+q)\Big) - ATT_{g'}(g+q) + ATT_{g'}(g-1)\Bigg] \times \\
& \Bigg[(T-1)\Bigg[\sum_{g \in \mathcal{G}_{q'}} N_g(N-N_g)\Bigg] + \Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g\Bigg)^2 - \sum_{g \in \mathcal{G}_{q'}} N_g^2 \\
& - \sum_{q \notin \{q',-1\}} \omega_q \Bigg\{\Bigg(\sum_{g \in \mathcal{G}_q} N_g\Bigg)\Bigg(\sum_{g \in \mathcal{G}_{q'}} N_g\Bigg) - T\sum_{g \in \mathcal{G}_q} N_g N_{g+q-q'} - N\sum_{g \in \mathcal{G}_q \cap \mathcal{G}_{q'}} N_g\Bigg\}\Bigg]^{-1}
\end{aligned}
$$

Omitting the denominator for ease of expression, we can collect the group-time treatment

effects that appear in multiple DiD terms.

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{g\in\mathcal{G}_{q'}} ATT_g(g+q')\Big[N_g\Big(\sum_{g'\neq g} N_{g'}+N_\infty\Big)\Big]\Big[(T-1)(1-\Omega_g)+T\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\omega_q\Big] \\
& + \sum_{g\in\mathcal{G}_{q'}}\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\Big(-ATT_g(g+q)\Big)\Big[N_g\Big(\sum_{g'\neq g} N_{g'}+N_\infty\Big)\Big]\Big[1-\Omega_g+T\omega_q\Big] \\
& + \sum_{g\notin\mathcal{G}_{q'}}\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\Big(-ATT_g(g+q)\Big)\Big[N_g\Big(\sum_{g'\neq g} N_{g'}+N_\infty\Big)\Big]\Big[-\Omega_g+T\omega_q\Big] \\
& + \sum_{g\in\mathcal{G}_{q'}}\sum_{g'\neq g}\Big(-ATT_{g'}(g+q')\Big)\Big[N_gN_{g'}\Big]\Big[(T-1)(1-\Omega_g)+T\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\omega_q\Big] \\
& + \sum_{g\in\mathcal{G}_{q'}}\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\sum_{g'\neq g}\Big(ATT_{g'}(g+q)\Big)\Big[N_gN_{g'}\Big]\Big[1-\Omega_g+T\omega_q\Big] \\
& + \sum_{g\notin\mathcal{G}_{q'}}\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\sum_{g'\neq g}\Big(ATT_{g'}(g+q)\Big)\Big[N_gN_{g'}\Big]\Big[-\Omega_g+T\omega_q\Big] \\
& + \sum_{g\in\mathcal{G}_{q'}}\sum_{g'\neq g}\Big(ATT_{g'}(g-1)\Big)\Big[N_gN_{g'}\Big]\Big[(1-\Omega_g)\Big] \\
& + \sum_{g\notin\mathcal{G}_{q'}}\sum_{g'\neq g}\Big(ATT_{g'}(g-1)\Big)\Big[N_gN_{g'}\Big]\Big[(T-1)\Omega_g-T\sum_{\substack{q\in\mathcal{Q}_g\\ q\notin\{q',-1\}}}\omega_q\Big]
\end{aligned}
$$

Using the definition of Ω_g

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{g \in \mathcal{G}_{q'}} ATT_g(g+q') \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[(T-1)(1-\Omega_g) + T\Omega_g \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(- ATT_g(g+q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[1 - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(- ATT_g(g+q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[- \Omega_g + T\omega_q \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(- ATT_{g'}(g+q') \Big) \Big[N_g N_{g'} \Big] \Big[(T-1)(1-\Omega_g) + T\Omega_g \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \sum_{g' \neq g} \Big(ATT_{g'}(g+q) \Big) \Big[N_g N_{g'} \Big] \Big[1 - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \sum_{g' \neq g} \Big(ATT_{g'}(g+q) \Big) \Big[N_g N_{g'} \Big] \Big[- \Omega_g + T\omega_q \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(ATT_{g'}(g-1) \Big) \Big[N_g N_{g'} \Big] \Big[(1-\Omega_g) \Big] \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(ATT_{g'}(g-1) \Big) \Big[N_g N_{g'} \Big] \Big[(T-1)\Omega_g - T\Omega_g \Big]
\end{aligned}
$$

To more easily collect terms in the summation and write sums over all relative time periods

including both -1 and q', we define $\omega_{-1} = 0$ and $\omega_{q'} = -1$

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{g \in \mathcal{G}_{q'}} \Big(- ATT_g(g+q') \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[1 - \Omega_g - T \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(- ATT_g(g+q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[1 - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \notin \{q', -1\}}} \Big(- ATT_g(g+q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[- \Omega_g + T\omega_q \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{g' \neq g} \Big(ATT_{g'}(g+q') \Big) \Big[N_g N_{g'} \Big] \Big[1 - \Omega_g - T \Big] \\
& + \sum_{g \in \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \neq q'}} \sum_{g' \neq g} \Big(ATT_{g'}(g+q) \Big) \Big[N_g N_{g'} \Big] \Big[1 - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \notin \mathcal{G}_{q'}} \sum_{\substack{q \in \mathcal{Q}_g \\ q \neq q'}} \sum_{g' \neq g} \Big(ATT_{g'}(g+q) \Big) \Big[N_g N_{g'} \Big] \Big[- \Omega_g + T\omega_q \Big]
\end{aligned}
$$

Combining sums over q and q' yields two sets of sums over all of the relative treatment time effects of timing group g and timing group g' (and again imposing $ATT_g(g-1) = 0$)

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{g \in \mathcal{G}} \sum_{q \in \mathcal{Q}_g} \Big(- ATT_g(g+q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \in \mathcal{G}} \sum_{q \in \mathcal{Q}_g} \sum_{g' \neq g} \Big(ATT_{g'}(g+q) \Big) \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big]
\end{aligned}
$$

Under the assumption of homogeneous relative treatment time effects, we have:

$$
ATT_g(g+q) = ATT_{g'}(g'+q) = RTT(q)
$$

Substituting

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{g \in \mathcal{G}} \sum_{q \in \mathcal{Q}_g} \Big(- RTT(q) \Big) \Big[N_g \Big(\sum_{g' \neq g} N_{g'} + N_\infty \Big) \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{g \in \mathcal{G}} \sum_{q \in \mathcal{Q}_g} \sum_{g' \neq g} \Big(RTT(g - g' + q) \Big) \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big]
\end{aligned}
$$

Switching the order of summation for g and q and factoring out the common relative treatment time effects:

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \Big[N_g N_\infty \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big(RTT(g-g'+q) \Big) \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big]
\end{aligned}
$$

We can re-write the sum over relative-treatment time effects $RTT(g-g'+q)$ as a sum over $RTT(q)$ for q that are associated with treatment timing group g'.

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \Big[N_g N_\infty \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(RTT(q) \Big) \sum_{g'\in\mathcal{G}_q} \sum_{g\neq g'} \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_{(g'-g+q)} \Big]
\end{aligned}
$$

Flip g and g' for the last term

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \Big[N_g N_\infty \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g \in \mathcal{G}_{q'}) - \Omega_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big[N_g N_{g'} \Big] \Big[\mathbf{1}(g' \in \mathcal{G}_{q'}) - \Omega_{g'} + T\omega_{(g-g'+q)} \Big]
\end{aligned}
$$

Define $\Omega^*_g = \sum_{q\in\mathcal{G}_q} \omega_q$ (including the $\omega_{-1} = 0$ and $\omega_{q'} = -1$ terms). Combine terms again.

$$
\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] \propto & \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \Big[N_g N_\infty \Big] \Big[- \Omega^*_g + T\omega_q \Big] \\
& + \sum_{q\in\mathcal{Q}} \Big(- RTT(q) \Big) \sum_{g\in\mathcal{G}_q} \sum_{g'\neq g} \Big[N_g N_{g'} \Big] \Big[- \Omega^*_g + \Omega^*_{g'} + T\omega_q - T\omega_{(g-g'+q)} \Big]
\end{aligned}
$$

Combining terms and factoring out N_g

$$E[\hat{\tau}^{(q')}|\mathcal{D}] \propto \sum_{q\in\mathcal{Q}} \Big(- RTT(q)\Big) \sum_{g\in\mathcal{G}_q} \Big[N_g\Big] \Big[N_\infty \Big(- \Omega_g^* + T\omega_q\Big) + \sum_{g'\neq g} N_{g'} \Big(- \Omega_g^* + \Omega_{g'}^* + T\omega_q - T\omega_{(g-g'+q)}\Big)\Big]$$

Re-arranging again

$$E[\hat{\tau}^{(q')}|\mathcal{D}] \propto \sum_{q\in\mathcal{Q}} \Big(RTT(q)\Big) \sum_{g\in\mathcal{G}_q} \Big[N_g\Big] \Big[(N-N_g)\Omega_g^* - (N-N_g)T\omega_q - \sum_{g'\neq g} N_{g'}\Omega_{g'}^* + T\sum_{g'\neq g} N_{g'}\omega_{(g-g'+q)}\Big]$$

Re-write as a sum of ω_q and $\sum_{q^*\neq q}\omega_{q^*}$ (including $\omega_{q'} = -1$ and $\omega_{-1} = 0$).

$$\begin{aligned} E[\hat{\tau}^{(q')}|\mathcal{D}] \propto \sum_{q\in\mathcal{Q}} \Big(RTT(q)\Big) \Big[&- \omega_q \Big((T-1)\sum_{g\in\mathcal{G}_q} N_g(N-N_g) + \Big(\sum_{g\in\mathcal{G}_q} N_g\Big)^2 - \sum_{g\in\mathcal{G}_q} N_g^2\Big) \\ &- \sum_{q^*\neq q} \omega_{q^*} \Big(\Big(\sum_{g\in\mathcal{G}_q} N_g\Big)\Big(\sum_{g\in\mathcal{G}_{q^*}} N_g\Big) - T\sum_{g\in G_q} N_g N_{g+q-q^*} - N\sum_{g\in\mathcal{G}_q\cap\mathcal{G}_{q^*}} N_g\Big)\Big] \end{aligned}$$

We recognize from Lemma 3.2 that this is equivalent to

$$E[\hat{\tau}^{(q')}|\mathcal{D}] \propto \sum_{q\in\mathcal{Q}} RTT(q) \times NT \Big[- \omega_q \sum_{it} \tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q)} - \sum_{q^*\neq q}\omega_q \sum_{it} \tilde{D}_{it}^{(q^*)}\tilde{D}_{it}^{(q)}\Big]$$

Returning the denominator:

$$E[\hat{\tau}^{(q')}|\mathcal{D}] = \sum_{q\in\mathcal{Q}} RTT(q) \frac{\Big[- \omega_q \sum_{it} \tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q)} - \sum_{q^*\neq q}\omega_q \sum_{it} \tilde{D}_{it}^{(q^*)}\tilde{D}_{it}^{(q)}\Big]}{\Big[\sum_{it} \tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q')} - \sum_{q^*\notin\{q',-1\}}\omega_{q^*}\sum_{it} \tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q^*)}\Big]}$$

Imposing $RTT(-1) = 0$:

$$E[\hat{\tau}^{(q')}|\mathcal{D}] = \sum_{\substack{q\in\mathcal{Q}\\ q\neq -1}} RTT(q) \frac{\Big[- \omega_q \sum_{it} \tilde{D}_{it}^{(q)}\tilde{D}_{it}^{(q)} - \sum_{q^*\neq q}\omega_q \sum_{it} \tilde{D}_{it}^{(q^*)}\tilde{D}_{it}^{(q)}\Big]}{\Big[\sum_{it} \tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q')} - \sum_{q^*\notin\{q',-1\}}\omega_{q^*}\sum_{it} \tilde{D}_{it}^{(q')}\tilde{D}_{it}^{(q^*)}\Big]}$$

From Lemma 3.2, the coefficient on $RTT(q)$ equals 0 for $q \notin \{q', -1\}$ since

$$-\omega_q \sum_{it} \tilde{D}_{it}^{(q)} \tilde{D}_{it}^{(q)} = \sum_{q^* \neq q} \omega_q \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q)}$$

Since $\omega_{q'} = -1$ and $\omega_{-1} = 0$

$$\begin{aligned}
E[\hat{\tau}^{(q')}|\mathcal{D}] &= RTT(q') \frac{\left[-\omega_{q'} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q'} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q')} \right]}{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q', -1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q^*)} \right]} \\
&= RTT(q') \frac{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q', -1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q^*)} \tilde{D}_{it}^{(q')} \right]}{\left[\sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q')} - \sum_{q^* \neq q', -1} \omega_{q^*} \sum_{it} \tilde{D}_{it}^{(q')} \tilde{D}_{it}^{(q^*)} \right]} \\
&= RTT(q')
\end{aligned}$$