Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*

Liyang Sun[†] and Sarah Abraham[‡]

September 22, 2020

Abstract

To estimate the dynamic effects of an absorbing treatment, researchers often use two-way fixed effects regressions that include leads and lags of the treatment. We show that in settings with variation in treatment timing across units, the coefficient on a given lead or lag can be contaminated by effects from other periods, and apparent pretrends can arise solely from treatment effects heterogeneity. We propose an alternative estimator that is free of contamination, and illustrate the relative shortcomings of two-way fixed effects regressions with leads and lags through an empirical application.

Keywords: difference-in-differences, two-way fixed effects, pretrend test

*We are grateful to Isaiah Andrews, Amy Finkelstein, Anna Mikusheva, and Heidi Williams for their guidance and support. We thank Alberto Abadie, Jonathan Cohen, Nathan Hendren, Peter Hull, Guido Imbens, Yunan Ji, Sylvia Klosin, Kevin Kainan Li, Paichen Li, Therese A. McCarty, Whitney Newey, James Poterba, Pedro H. C. Sant'Anna, Gergely Ujhelyi and Helen Willis for helpful discussions. This research was supported by the National Institute on Aging, Grant Number T32-AG000186. A preliminary draft of this paper was circulated on April 16, 2018. Replication code is available at `http://economics.mit.edu/grad/lsun20/`. The companion Stata package `eventstudyweights` is available from the SSC repository.

[†]Department of Economics, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139. Corresponding author; lsun20@mit.edu.

[‡]Cornerstone Research, 699 Boylston St, Boston, MA 02116. The views expressed herein are solely those of the author, who is responsible for the content, and do not necessarily represent the views of Cornerstone Research.

1 Introduction

Rich panel data has fueled a growing literature estimating treatment effects with two-way fixed effects regressions. This body of applied work has prompted a corresponding econometrics literature investigating the assumptions required for these regressions to yield causally interpretable estimates. For example, Athey and Imbens (2018), Borusyak and Jaravel (2017), Callaway and Sant'Anna (2020a), de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2018) interpret the coefficient on the treatment status when there is treatment effects heterogeneity and variation in treatment timing. Researchers are often also interested in dynamic treatment effects, which they estimate by the coefficients μ_ℓ associated with indicators for being ℓ periods relative to the treatment, in a specification that resembles the following:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell} \mu_\ell \mathbf{1}\{t - E_i = \ell\} + \upsilon_{i,t}. \tag{1}$$

Here $Y_{i,t}$ is the outcome of interest for unit i at time t, E_i is the time when unit i initially receives the binary absorbing treatment, and α_i and λ_t are the unit and time fixed effects. Units are categorized into different cohorts based on their initial treatment timing. The relative times $\ell = t - E_i$ included in (1) cover most of the possible relative periods, but may still exclude some periods.

The first goal of this paper is to uncover potential pitfalls associated with using the estimates of the relative period coefficients μ_ℓ as "reasonable" measures of dynamic treatment effects. We decompose μ_ℓ to show it can be expressed as a linear combination of cohort-specific effects from both its own relative period ℓ and other relative periods; unless strong assumptions regarding treatment effects homogeneity hold, the terms that include treatment effects from other relative periods will not cancel out and will contaminate the estimate of μ_ℓ. Importantly, this demonstrates that the widespread practice of using estimates of treatment leads in (1) as a way of testing for parallel pretrends is problematic. Roth (2019), in his survey of the applied literature, notes that checking whether $\mu_\ell = 0$ for ℓ leads of treatment is a common test for pretrends. Our decomposition result implies that such a test would be invalid because the estimate of μ_ℓ is affected by both pretrends and treatment effects heterogeneity, thus any test of $\mu_\ell = 0$ cannot accept or reject the existence of pretrends without further assumptions on treatment effects.

We show how to calculate the weights underlying the linear combination of treatment effects in μ_ℓ using an auxiliary regression. This auxiliary regression depends only on the distribution of cohorts and the relative time indicators included in (1). Examining the weights allows researchers to gauge how large the amount of treatment effects heterogeneity needs to be for μ_ℓ to be contaminated by treatment effects from other relative periods. Our publicly-available Stata package `eventstudyweights` automates the estimation of these weights using the panel dataset underlying any given specification of (1).

The second goal of this paper is to propose an alternative regression-based method that is more robust to treatment effects heterogeneity than regression (1). For dynamic treatment effects, researchers are usually interested in estimating some average of treatment effects from ℓ periods relative to the treatment. Our alternative method estimates the shares of cohort as weights. These weights are more interpretable than the weights underlying regression (1) in the presence of treatment effects heterogeneity, and the resulting weighted average of treatment effects extends beyond a convex combination of treatment effects (Słoczyński, forthcoming). As discussed in Section 4.2, using the procedures of Callaway and Sant'Anna (2020a), our alternative method can also accommodate covariates.

We illustrate both our decomposition results and our alternative method via an empirical application, estimating the dynamic effects of a hospitalization. We follow Dobkin et al. (2018) in using the publicly-available dataset, Health and Retirement Study (HRS), to first estimate two-way fixed effects regressions. We then illustrate our alternative estimation method with this example. Among the outcomes studied by Dobkin et al. (2018), we focus on out-of-pocket medical spending and labor earnings. Our alternative method yields similar big-picture findings as the original paper that uses two-way fixed effects regressions: the earnings decline due to hospitalization is substantial compared to the transitory out-of-pocket spending increase. However, the two-way fixed effects estimates sometimes fall outside the convex hull of the underlying effects. In contrast, estimates using our alternative method, by construction, are guaranteed to be easy-to-interpret because they are weighted averages of the underlying effects, with weights corresponding to cohort shares.

The rest of the paper is organized as follows. In the next subsection, we review the theoretical literature. Section 2 formally introduces the event study design and discusses our definition in relation to the applied literature. Section 3 derives the estimands of two-way fixed effects regression, and introduces sufficient assumptions for them to be causally interpretable. Section 4 develops our alternative estimator. Section 5 illustrates our results using an empirical example and Section 6 concludes. All proofs are contained in the Online Appendix.

1.1 Related Literature

This paper makes two main contributions within an active literature on the causal interpretations of two-way fixed effects models in settings with staggered treatment adoption (Athey and Imbens, 2018; Borusyak and Jaravel, 2017; Callaway and Sant'Anna, 2020a; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2018). Our paper is also related to the traditional literature analyzing non-separable panel and treatment effects models e.g. Heckman et al., 1998, 1997; Blundell et al., 2004; Abadie, 2005; Chernozhukov et al., 2013.

The first main contribution of our paper is to interpret estimates from two-way fixed effects specifications when researchers include "dynamic" indicators for time relative to treatment and when treatment effects are heterogeneous across adoption cohorts. We derive our results for a general class of two-way fixed effects specifications where "dynamic" indicators can be flexibly specified as single relative periods ℓ or sets of relative periods g (thus also capturing any "static" specification where all post-treatment indicators are collected in a single set). This class of specifications encompasses all specifications addressed by Athey and Imbens (2018), Borusyak and Jaravel (2017), Callaway and Sant'Anna (2020a), de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2018).

As a building block for the causal interpretation of estimates, we define $CATT_{e,\ell}$, the cohort average treatment effects on the treated as the cohort-specific average difference in outcomes relative to never being treated. Our choice of a "building block" is governed by the counterfactual and the type of heterogeneity of interest. This object coincides with the "group-time average treatment effect" studied by Callaway and Sant'Anna (2020a) and is more granular than the building block used by Goodman-Bacon (2018) that is an average of $CATT_{e,\ell}$ over some relative period range. Athey and Imbens (2018) consider an alternate counterfactual to never being treated: being treated at a different time. Borusyak and Jaravel (2017) implicitly assume away heterogeneity across cohorts within a relative period, so their building block reduces to ATT_{ℓ}. de Chaisemartin and D'Haultfœuille (2020) allow for heterogeneous treatment paths within a cohort over time across "groups", thus their building block is at the group level. We defer a discussion of assumptions underlying the causal interpretation of these building blocks to Section 2.

The second main contribution of our paper is to propose a simple regression-based alternative estimation strategy that produces a more sensible estimand than conventional two-way fixed effects models under heterogeneous treatment effects. Our procedure is most similar to Callaway and Sant'Anna (2020a), but has the following differences. First, in the setting where there is no never-treated group, our method uses the last cohort to be treated as a control group, whereas Callaway and Sant'Anna (2020a) use the set of not-yet-treated cohorts. Our method and theirs thus rely on different, but non-nested parallel trends assumptions. Second, our estimation method can be cast as a regression specification and thus may be more familiar to applied researchers. However, a third difference is that the procedure of Callaway and Sant'Anna (2020a) allows for conditioning on time-varying covariates. de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2018) respectively propose alternative estimators and diagnostic tools for estimation of causal effects in staggered settings, but do not consider the estimation of the dynamic path of treatment effects as we do.

2 Event studies design

In this section we first formalize the "event studies design". As discussed in Section 2.3, based on how this term is deployed in the empirical literature, an *event study* design is a *staggered adoption* design where units are treated at different times, and there may or may not be never treated units. It also nests a *difference-in-differences* design, where units are either first treated at time t_0 or never treated.

Specifically, we consider a setting with a random sample of N units observed over $T+1$ time periods, where T is fixed. For each $i \in \{0, \dots, N\}$ and $t \in \{0, ..., T\}$, we observe the outcome $Y_{i,t}$ and treatment status $D_{i,t} \in \{0, 1\}$: $D_{i,t} = 1$ if i is treated in period t and $D_{i,t} = 0$ if i is not treated in period t. Throughout we assume that the observations $\{Y_{i,t}, D_{i,t}\}_{t=0}^{T}$ are independent and identically distributed (i.i.d.).

In the general case of event studies we focus on an *absorbing* treatment such that the treatment status over time is a non-decreasing sequence of zeros and then ones, i.e. $D_{i,s} \leq D_{i,t}$ for $s < t$. We can thus uniquely characterize a treatment path by the time period of the initial treatment, denoted with $E_i = \min\{t : D_{i,t} = 1\}$. If unit i is never treated i.e. $D_{i,t} = 0$ for all t, we set $E_i = \infty$. Based on when they first receive the treatment, we can also uniquely categorize units into disjoint cohorts e for $e \in \{0, \dots, T, \infty\}$, where units in cohort e are first treated at the same time $\{i : E_i = e\}$.

We define $Y_{i,t}^{e}$ to be the potential outcome in period t when unit i is first treated in time period e. We define $Y_{i,t}^{\infty}$ to be the potential outcome if unit i never receives the treatment, which we call the "baseline outcome". Since the timing of the initial treatment uniquely characterizes one's treatment path, we can represent the observed outcome for unit i as

$$Y_{i,t} = Y_{i,t}^{E_i} = Y_{i,t}^{\infty} + \sum_{0 \le e \le T} (Y_{i,t}^{e} - Y_{i,t}^{\infty}) \cdot \mathbf{1}\{E_i = e\}. \tag{2}$$

For any treatment that is not absorbing, if we replace the treatment status $D_{i,t}$ with an indicator for ever having received the treatment, the new treatment is absorbing by construction. Oftentimes the effect of having ever received the treatment is of interest, as it captures the path of treatment effects even though the treatment itself may be transient. For example, Deryugina (2017) is interested in the fiscal cost for a county that has been hit by a hurricane. While a hurricane itself may be transient, the impact of having had a hurricane may not be transient, hence why Deryugina (2017) codes the year of the first hurricane experienced in a county as E_i.

In the next section, we use the notation developed above to define the treatment effect of an event study design.

2.1 Defining treatment effect of an event study design

In an event study design, we define the unit-level *treatment effect* as the difference between the observed outcome relative to the never-treated counterfactual outcome: $Y_{i,t} - Y_{i,t}^{\infty}$. Recall that $Y_{i,t}^{\infty}$ denotes the potential outcome if unit i never receives the treatment. This particular counterfactual outcome $Y_{i,t}^{\infty}$ is a reasonable "baseline outcome", though other counterfactual outcomes may be of interest as well. For example, Athey and Imbens (2018) also consider the treatment effect relative to the always-treated counterfactual outcome: $Y_{i,t} - Y_{i,t}^{0}$. Sianesi (2004) defines the unit-level treatment effect to be relative to the not-yet-treated counterfactual outcome: $Y_{i,t} - Y_{i,t}^{e}$ for $e > t$.

When dynamic treatment effects are of interest, empirical researchers commonly report the coefficient estimate $\widehat{\mu}_\ell$ associated with indicators for being ℓ periods relative to the treatment in regression (1) as an estimate for the average lagged effect. To assess the causal interpretation of μ_ℓ, we need "building blocks" for its decomposition, which are the average of unit-level treatment effects at a given relative period across units first treated at time $E_i = e$, i.e. units in the same cohort e. We call this average the cohort-specific average treatment effects on the treated, formally defined below. Later in Section 3 we use them as building blocks for the interpretation of the relative period coefficients μ_ℓ from two-way fixed effects regressions.

Definition 1. The cohort-specific average treatment effect on the treated (CATT) ℓ periods from initial treatment is

$$CATT_{e,\ell} = E[Y_{i,e+\ell} - Y^{\infty}_{i,e+\ell} \mid E_i = e]. \tag{3}$$

Each $CATT_{e,\ell}$ represents the average treatment effect ℓ periods from the initial treatment for the cohort of units first treated at time e. We shift from calendar time index t to relative period index ℓ which denotes the periods since treatment; for cohort e, ℓ ranges from $-e$ to $T-e$ because we observe at most e periods before the initial treatment and $T-e$ periods after the initial treatment. Relative periods allow us to compare cohorts while holding their exposure to the treatment constant.

2.2 Identifying assumptions

With the above definitions, we formalize three potential identifying assumptions for outcomes of interest in our event study design. The first assumption is a generalized form of a parallel trends assumption. The second assumption requires no anticipation of the treatment. The third assumption imposes no variation across cohorts. For each assumption, we first discuss its meaning and then compare it with similar assumptions made in the literature interpreting two-way fixed effects regressions. Later in Section 3 we interpret the relative period coefficients μ_ℓ from two-way fixed effects regressions under different combinations of these assumptions.

Assumption 1. (Parallel trends in baseline outcomes.) *For all $s \neq t$, the $E[Y^{\infty}_{i,t} - Y^{\infty}_{i,s} | E_i = e]$ is the same for all $e \in supp(E_i)$.*

If an application includes never-treated units so that $\infty \in supp(E_i)$, we need to especially consider whether these never-treated units satisfy the parallel trends assumption. Never-treated units are likely to differ from ever-treated units in many ways, and may not share the same evolution of baseline outcomes. If the never-treated units are unlikely to satisfy the parallel trends assumption, then we should exclude them from the estimation to avoid violation of this assumption.

While common in the applied literature, the parallel trends assumption is strong and oftentimes violated. For example, Ashenfelter (1978) documented that participants in job training programs experience a decline in earnings prior to the training period (Ashenfelter's dip). The timing of job training is dependent on the evolution of individual's baseline earnings, and this scenario therefore does not satisfy the parallel trends assumption. Proposition 1 is the only result in this paper derived without this assumption, but there is an active literature studying inference under violations of the parallel trends assumption e.g. Rambachan and Roth (2020).

Our parallel trends assumption coincides with that of de Chaisemartin and D'Haultfœuille (2020). One could substitute this assumption with a different identifying assumption that baseline outcomes are mean independent of E_i i.e. at each t, $E[Y^{\infty}_{i,t}|E_i = e]$ is the same for all $e \in supp(E_i)$ and in particular is equal to $E[Y^{\infty}_{i,t}]$. This stronger assumption is plausible when the timing of treatment is indeed randomized, which is the assumption used by Athey and Imbens (2018). By taking the "fully dynamic" specification as their DGP, Borusyak and Jaravel (2017) implicitly assume this version of a parallel trends assumption. Callaway and Sant'Anna (2020a) propose a weaker version that is conditional on covariates. Finally, for a particular estimand, Goodman-Bacon (2018) identifies a weaker version that only requires a weighted average of $E[Y^{\infty}_{i,t} - Y^{\infty}_{i,s}|E_i = e]$ (averaged across cohorts) to be zero.

Assumption 2. (No anticipatory behavior prior to treatment.) *There is no treatment effect in pre-treatment periods i.e.* $E[Y^{e}_{i,e+\ell} - Y^{\infty}_{i,e+\ell} \mid E_i = e] = 0$ *for all* $e \in supp(E_i)$ *and all* $\ell < 0$.

Assumption 2 requires potential outcomes in any ℓ periods before treatment to be equal to the baseline outcome on average as in Malani and Reif (2015) and Botosaru and Gutierrez (2018). This is most plausible if the full treatment paths are not known to units. If they have private knowledge of the future treatment path they may change their behavior in anticipation and thus the potential outcome prior to treatment may not represent baseline outcomes. For example, Hendren (2017) shows that knowledge of future job loss leads to decreases in consumption. If the periods with anticipation behavior are known, then we may consider an alternative version of Assumption 2, which holds for pre-periods in a subset of pre-treatment periods. Depending on the application, it may still be plausible to assume no anticipation until K periods before the treatment.

The no anticipation assumption proposed by Athey and Imbens (2018) is a deterministic condition which stipulates that $Y^{e}_{i,e+\ell} = Y^{\infty}_{i,e+\ell}$ for all units i and e and $\ell < 0$. By taking the "fully dynamic" specification as their DGP, Borusyak and Jaravel (2017) allow anticipation by including pre-trends indicators in the DGP. Callaway and Sant'Anna (2020a) and Goodman-Bacon (2018) implicitly assume no anticipation by using observed outcomes in time periods before the initial treatment as the untreated potential outcomes.

Assumption 3. (Treatment effect homogeneity.) *For each relative period* ℓ, $CATT_{e,\ell}$ *does not depend on cohort* e *and is equal to* ATT_ℓ.

Assumption 3 requires that each cohort experiences the same path of treatment effects. Treatment effects need to be the same across cohorts in every relative period for homogeneity to hold, whereas for heterogeneity to occur, treatment effects only need to differ across cohorts in one relative period. The assumption of treatment effect homogeneity is therefore strong, and in Section 3.4.1, we describe how it can be violated in applied settings.

Our notion of treatment effect homogeneity does not preclude dynamic treatment effects; it only imposes that cohorts share the same path of treatment effects. The related literature sometimes formulates restrictions on the dynamics of treatment effects as another notion of treatment effect homogeneity. Athey and Imbens (2018) propose an assumption that "restricts the heterogeneity of the treatment effects over time," which implies $CATT_{e,\ell}$ can vary over e but not over ℓ. Borusyak and Jaravel (2017) refer to one type of treatment effects heterogeneity as "only across the time horizon," which implies $CATT_{e,\ell}$ can vary over ℓ but not over e. Callaway and Sant'Anna (2020a) allow for "arbitrary treatment effect heterogeneity" when $CATT_{e,\ell}$ varies across cohorts and over time. Similarly, de Chaisemartin and D'Haultfœuille (2020) describe treatment effects that may be "heterogeneous across groups and over time periods." Goodman-Bacon (2018) allow heterogenous effects to either "vary across units but not over time" or "vary over time but not across units." The literature has not converged on a single notion of treatment effects heterogeneity with time-varying treatment. Since researchers are interested in dynamic treatment effects when using a "dynamic" specification, we do not restrict the path of treatment effects but rather use "heterogeneity" to describe variation across cohorts only.

2.3 Relevance in the applied literature

To gauge the empirical relevance of our results, we survey the estimation methods used by the twelve papers collected by Roth (2019) from three leading economics journals that contain the phrase "event study" in their main text.[1] From this sample of applied papers, we learn what specifications empirical researchers are actually using when estimating two-way fixed effects regressions. Four papers in this sample consider the simple setting where units either receive their first treatment at the same time or never receive the treatment. The other eight papers in this sample consider the more complex setting where treated units receive their

[1] We follow the selection criteria in Roth (2019): the original sample consists of 70 total papers, but is further constrained to these twelve papers with publicly available data and code. The data and code are used to determine exactly the specification estimated in these papers.

treatment at various times, and there may or may not be never treated units. This observation suggests an *event study* in the applied literature nests two popular research designs: *difference-in-differences* design, but also the design where units receive their first treatments at various times, which is our focus.[2]

We summarize the main specifications in this sample of twelve applied papers in Table 1.[3] The columns of Table 1 collect key properties of these specifications. In Section 3.1, we introduce a general class of specifications that encompasses all of these specification. The estimates for relative period coefficients μ_ℓ from all these papers therefore fall under our analysis in the next section.

These papers demonstrated that event studies are used to address a broad range of research questions. As an example of this literature, Bailey and Goodman-Bacon (2015) use the rollout of the first Community Health Centers (CHCs) to study the longer-term health effects of increasing access to primary care. As another example, Tewari (2014) uses variation in the timing of deregulation across states to estimate the impact of financial development on homeownership.

3 Estimators from linear two-way fixed effects regression

We consider a two-way fixed effects (FE) regression of the following form, estimated on a panel of $i = 1, \ldots, N$ units for $t = 0, 1, \ldots, T$ *calendar* time periods:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \mu_g \mathbf{1}\{t - E_i \in g\} + \upsilon_{i,t} \tag{4}$$

Here $Y_{i,t}$ is the outcome of interest for unit i at time t, E_i is the time for unit i to initially receive a binary absorbing treatment, and α_i and λ_t are the unit and time fixed effects. The set $\mathcal{G}$ collects disjoint sets g of *relative periods* $\ell \in [-T, T]$. We allow some relative periods to be excluded from the specification and denote the excluded set with $g^{excl} = \{\ell : \ell \notin \bigcup_{g \in \mathcal{G}} g\}$. We denote by μ_g the relative period coefficients from regression (4), i.e. the population regression coefficients. Their corresponding OLS estimators are denoted by $\widehat{\mu}_g$ respectively.

We are interested in the properties of μ_g when there are variations in the initial treatment timing, and

[2]This setup is the same as the *staggered adoption* design proposed by Athey and Imbens (2018), but we keep the term *event study* because it is common in the applied literature.

[3]We focus on the first specification underlying the event study estimates in each paper, which we view as a reasonable proxy for the main specification in the paper.

there may or may not be never-treated units. Below in Section 3.1 we illustrate how the choice of $\mathcal{G}$ coincides with a large number of specifications encountered in practice such as the "fully dynamic" specification. We next decompose μ_g in terms of $CATT_{e,\ell}$ when various combinations of the three identifying assumptions fail. For Propositions 1-3, we state the results in terms of the general specification (4). To specialize these results to the "fully dynamic" specification (7), we note the corresponding decomposition would replace bins with $g = \{\ell\}$ for each relative period ℓ included in the specification and g^{excl} would contain all excluded relative periods. The decomposition remains unchanged though the summation over $\ell \in g$ simplifies since each g is a singleton. For Proposition 4, the decomposition further simplifies for the "fully dynamic" specification as discussed below.

Researchers may assume that μ_g can be interpreted as a convex average of $CATT_{e,\ell}$ for periods $\ell \in g$ from its corresponding set g; they may further assume the underlying weights have policy-relevant interpretation, e.g. weights depending on proportions of cohorts. For example, Bailey and Goodman-Bacon (2015) interpret them as "intention-to-treat effects" of the treatment in a given relative year. However, our results show that μ_g may not represent the parameter of interest without strong assumptions such as treatment effect homogeneity. Section 3.6 provides intuition for these negative results, and demonstrate the weights are actually non-linear functions of proportions of cohorts. Section 3.7 illustrates how treatment effects heterogeneity invalidates the pretrends test for a simple three-period setting.

3.1 Common specifications

Common specifications can be broadly categorized as either "static" or "dynamic". Static specifications estimate a single treatment effect that is time invariant. In contrast, dynamic specifications allow for non-parametric changes in the treatment effects over time. Within dynamic specifications, researchers also need to address issues of multi-collinearity, and may bin or trim distant relative periods. All of these choices can be written as instances of (4) with the correct specification of $\mathcal{G}$, meaning that our results are applicable for a wide range of specifications employed in the empirical literature.

To clarify how to specify $\mathcal{G}$ in regression (4), we define $D^{\ell}_{i,t} := \mathbf{1}\{t - E_i = \ell\}$ to be an indicator for unit i being ℓ periods away from initial treatment at calendar time t. For never-treated units $E_i = \infty$, we set

$D^{\ell}_{i,t} = 0$ for all ℓ and all t. We can represent the relative period bin indicator as

$$\mathbf{1}\{t - E_i \in g\} = \sum_{\ell \in g} \mathbf{1}\{t - E_i = \ell\} = \sum_{\ell \in g} D^{\ell}_{i,t}. \tag{5}$$

Static specification. For a "static" specification $\mathcal{G}$ contains a single element equal to $g = [0, T]$. The indicator $\mathbf{1}\{t - E_i \in g\}$ is equivalent to an indicator for whether unit i has received its initial treatment by t: $\mathbf{1}\{E_i \leq t\}$. The "static" specification thus takes the following form

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_g \sum_{\ell \geq 0} D^{\ell}_{i,t} + \upsilon_{i,t} \tag{6}$$

and the corresponding set of excluded relative periods is $g^{excl} = [-T, -1]$.

Dynamic specification. "Dynamic" specifications encompass any specifications of $\mathcal{G}$ where $\mathcal{G}$ contains more than one element, thus treatment effects are allowed to vary over time non-parametrically. In its most flexible form, a "fully dynamic" specification takes the following form

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell=-K}^{-2} \mu_{\ell} D^{\ell}_{i,t} + \sum_{l=0}^{L} \mu_{\ell} D^{\ell}_{i,t} + \upsilon_{i,t} \tag{7}$$

and the corresponding set of excluded relative periods is $g^{excl} = \{-T, \ldots, -K-1, -1, L+1, \ldots, T\}$.

Excluding some relative periods from the "fully dynamic" specification is necessary to avoid multi-collinearity, either among the relative period indicators $D^{\ell}_{i,t}$, or with the unit and time fixed effects. For example, when there are no never-treated units i.e. $\infty \notin supp(E_i)$ but with a panel balanced in calendar time, we need to exclude at least two relative period indicators in $\mathcal{G}$. These collinearities are discussed by Borusyak and Jaravel (2017): one multi-collinearity comes from the relative period indicators summing to one for every unit $\sum_{\ell \in [-T,T]} D^{\ell}_{i,t} = 1$, and the other multi-collinearity comes from the linear relationship between two-way fixed effects and the relative period indicators, namely $t - E_i = \ell$.

Excluding relative periods close to the initial treatment is common in practice. Normalizing relative to the period prior to treatment is the most common - six out of the eight papers we survey do so, as reflected in the above specification where we drop $D^{-1}_{i,t}$. The remaining two papers exclude $D^{0}_{i,t}$.

Excluding distant relative periods is however less common (only one of the eight papers we survey does so). Instead researchers "bin" or "trim" distant relative periods. For "binning", researchers bin distant relative periods into $[-T,-K)$ and $(L,T]$ and estimate a "binned" specification

$$Y_{i,t} = \alpha_i + \lambda_t + \beta \cdot \sum_{\ell<-K} D_{i,t}^{\ell} + \sum_{\ell=-K}^{-2} \mu_\ell D_{i,t}^{\ell} + \sum_{l=0}^{L} \mu_\ell D_{i,t}^{\ell} + \gamma \cdot \sum_{\ell>L} D_{i,t}^{\ell} + \upsilon_{i,t} \tag{8}$$

without excluding any distant relative periods so that $g^{excl} = \{-1\}$. For "trimming", researchers trim their panel to be balanced in relative periods.

Neither "binning" nor "trimming" resolves the issues of contamination discussed below (i.e. the possibility that treatment effects from other periods affect the estimate for a given μ_g). We show the contamination issue for the general specification (4) encompasses both practices. For a given coefficient in the dynamic specification, "trimming" does however mechanically remove any treatment effects from the relative periods "trimmed" from the specification. For the static specification put forth in Borusyak and Jaravel (2017), they noted that "trimming" also does not resolve the contamination issue they identified with the static specification.

3.2 Interpreting the coefficients under no assumptions

First we show that without any assumptions, we can write μ_g as a linear combination of differences in trends.

Proposition 1. *The population regression coefficient on relative period bin g is a linear combination of differences in trends from its own relative period $\ell \in g$, from relative periods $\ell \in g'$ belonging to other bins $g' \neq g$ but included in the specification, and from relative periods excluded from the specification $\ell \in g^{excl}$:*

$$\mu_g = \sum_{\ell \in g} \sum_{e} \omega_{e,\ell}^{g} \left(E[Y_{i,e+\ell} - Y_{i,0}^{\infty} | E_i = e] - E[Y_{i,e+\ell}^{\infty} - Y_{i,0}^{\infty}] \right) \tag{9}$$

$$+ \sum_{g' \neq g, g' \in \mathcal{G}} \sum_{\ell \in g'} \sum_{e} \omega_{e,\ell}^{g} \left(E[Y_{i,e+\ell} - Y_{i,0}^{\infty} | E_i = e] - E[Y_{i,e+\ell}^{\infty} - Y_{i,0}^{\infty}] \right) \tag{10}$$

$$+ \sum_{\ell \in g^{excl}} \sum_{e} \omega_{e,\ell}^{g} \left(E[Y_{i,e+\ell} - Y_{i,0}^{\infty} | E_i = e] - E[Y_{i,e+\ell}^{\infty} - Y_{i,0}^{\infty}] \right). \tag{11}$$

We use the superscript g to associate the weight $\omega_{e,\ell}^{g}$ with the coefficient μ_g. The weight $\omega_{e,\ell}^{g}$ is equal to the population regression coefficient on $\mathbf{1}\{t - E_i \in g\}$ from regressing $D_{i,t}^{\ell} \cdot \mathbf{1}\{E_i = e\}$ on all bin indicators $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ included in the specification (4) and two-way fixed effects.

The above proposition is a direct result of regression mechanics. We provide an intuitive derivation for the closed-form expressions for the weights using the classical "omitted variables bias formula" in Section 3.6. We defer the formal derivation to the Appendix. Here we mention the following four properties of the weights $\omega^g_{e,\ell}$.

- For relative periods of μ_g's own bin i.e. $\ell \in g$, their associated weights as displayed in (9) sum to one $\sum_{\ell \in g} \sum_e \omega^g_{e,\ell} = 1$.
- For relative periods belonging to some other bin included in (4) i.e. $\ell \in g'$ for $g' \neq g$ and $g' \in \mathcal{G}$, their associated weights as displayed in (10) sum to zero $\sum_{\ell \in g'} \sum_e \omega^g_{e,\ell} = 0$ for each bin g'.
- For relative periods not included in $\mathcal{G}$, their associated weights as displayed in (11) sum to negative one $\sum_{\ell \in g^{excl}} \sum_e \omega^g_{e,\ell} = -1$.
- If there are never-treated units i.e. $\infty \in supp(E_i)$, we have $\omega^g_{\infty,\ell} = 0$ for all g and ℓ.

We can easily estimate the weights $\omega^g_{e,\ell}$ for any given specification of $\mathcal{G}$ using the following auxiliary regression:

$$D^{\ell}_{i,t} \cdot \mathbf{1}\{E_i = e\} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \omega^g_{e,\ell} \mathbf{1}\{t - E_i \in g\} + \upsilon_{i,t} \tag{12}$$

which regresses $D^{\ell}_{i,t} \cdot \mathbf{1}\{E_i = e\}$ on all bin indicators included in regression (4) and two-way fixed effects.

All of the above properties can be extended to a case where covariates are added to regression (4) by partialling out the covariates before proceeding. In other words, the terms in parentheses in (9), (10) and (11) would be replaced by terms for which the covariates are partialled out. The weights can be estimated by controlling for covariates in regression (12) the same way as they are controlled for in the original regression.

However, covariates complicate the interpretations of μ_g in terms of $CATT_{e,\ell}$ as we describe below in Proposition 2-4. Depending on how covariates are controlled for in regression (4), we may need an additional assumption that the counterfactual trends are linear in the time-varying covariates $X_{i,t}$. We leave a full investigation of the introduction of covariates for future work.

3.3 Interpreting the coefficients under parallel trends assumption only

Proposition 2. *Under Assumption 1 (parallel trends) only, the population regression coefficient on the indicator for relative period bin g is a linear combination of $CATT_{e,\ell \in g}$ as well as $CATT_{e,\ell'}$ from other* *relative periods $\ell' \notin g$, with the same weights stated in Proposition 1:*

$$\mu_g = \sum_{\ell \in g} \sum_e \omega^g_{e,\ell} CATT_{e,\ell} + \sum_{g' \neq g, g' \in \mathcal{G}} \sum_{\ell' \in g'} \sum_e \omega^g_{e,\ell'} CATT_{e,\ell'} + \sum_{\ell' \in g^{excl}} \sum_e \omega^g_{e,\ell'} CATT_{e,\ell'}. \tag{13}$$

Under Assumption 1, the terms in Proposition 1 reduce to a linear combination of the causally interpretable building blocks $CATT_{e,\ell}$ as follows:

$$E[Y_{i,e+\ell} - Y^\infty_{i,0} | E_i] - E[Y^\infty_{i,e+\ell} - Y^\infty_{i,0}] = CATT_{e,\ell} + \underbrace{E[Y^\infty_{i,t} - Y^\infty_{i,0} | E_i] - E[Y^\infty_{i,t} - Y^\infty_{i,0}]}_{=0} \tag{14}$$

for $t = e + \ell$. However, two issues for interpretability remain. First, the coefficient μ_g can be written as an average of not only $CATT_{e,\ell}$ from own periods $\ell \in g$, but also $CATT_{e,\ell'}$ from other periods. Second, the weights are still non-linear functions of the distribution of cohorts, same as those in (12), and they are not restricted to lie in $[0,1]$.

The properties of these weights as described following Proposition 1 imply that contamination from other periods wanes once we impose restrictions on treatment effects. In the next two subsections we illustrate how that can happen.

3.4 Interpreting the coefficients under parallel trends and no anticipation assumptions

Proposition 3. *If Assumption 1 (parallel trends) holds and Assumption 2 (no anticipatory behavior in all periods before the initial treatment) holds, the population regression coefficient μ_g is a linear combination of post-treatment $CATT_{e,\ell'}$ for all $\ell' \geq 0$, with the same weights stated in Proposition 1:*

$$\mu_g = \sum_{\ell' \in g, \ell' > 0} \sum_e \omega^g_{e,\ell} CATT_{e,\ell} + \sum_{g' \neq g, g' \in \mathcal{G}} \sum_{\ell' \in g', \ell' > 0} \sum_e \omega^g_{e,\ell'} CATT_{e,\ell'} + \sum_{\ell' \in g^{excl}, \ell' > 0} \sum_e \omega^g_{e,\ell'} CATT_{e,\ell'}. \tag{15}$$

Once we restrict pre-treatment $CATT_{e,\ell \leq 0}$ to be zero under the no anticipatory behavior assumption, the expression for μ_g simplifies as terms involving $CATT_{e,\ell \leq 0}$ drop out. However, the second term in the expression for μ_g remains unless we further impose treatment effect homogeneity for its summands to cancel out each other. Thus, μ_g may be non-zero for pre-treatment periods even if parallel trends holds.

This result immediately implies a shortcoming of using pre-treatment coefficients (i.e. μ_g where g contains only leads to the treatment $\ell < 0$) to test for pretrends. Under the no anticipatory behavior assumption, cohort-specific treatment effects prior to treatment are all zero: $CATT_{e,\ell} = 0$ for all $\ell < 0$. Therefore, any linear combination of these $CATT_{e,\ell}$ is also zero. However, μ_g is a function of post-treatment $CATT_{e,\ell' \geq 0}$ as well, even when g only contains elements with $\ell < 0$. We revisit this implication in greater depth in Section 3.7. Callaway and Sant'Anna (2020a) provides alternative tests for pretrends that do not suffer from this drawback.

3.4.1 Sources of treatment effect heterogeneity

Since treatment effects heterogeneity violates Assumption 3 and can alter how we interpret μ_g, it is important to think through when different cohorts likely experience different paths of treatment effect. Such heterogeneity could arise for many reasons. For example, cohorts may differ in their covariates, which affect how they respond to treatment. We will explore a concrete example in our application: if treatment effects differ with age, and there is variation in age across units first treated at different times, we will have heterogeneous effects (see Section 5 for details). After controlling for covariates, cohorts may still vary in their responses to the treatment if units select their initial treatment timing based on treatment effects. This source of heterogeneity is still compatible with our parallel trends assumption, which only rules out selection in the initial treatment timing based on the evolution of the baseline outcome. In addition to these two sources of heterogeneity, treatment effects may vary across cohorts due to calendar time-varying effects (e.g. macroeconomic conditions could govern the effects on labor market outcomes across cohorts).

3.5 Interpreting the coefficients under parallel trends and treatment effect homogeneity

Proposition 4. *If Assumption 1 (parallel trends) holds and Assumption 3 (treatment effect homogeneity) holds, then $CATT_{e,\ell} = ATT_\ell$ is constant across e for a given ℓ, and the population regression coefficient μ_g is equal to a linear combination of $ATT_{\ell \in g}$, as well as $ATT_{\ell' \notin g}$ from other relative periods:*

$$\mu_g = \sum_{\ell \in g} \omega_\ell^g ATT_\ell + \sum_{g' \neq g} \sum_{\ell' \in g'} \omega_{\ell'}^g ATT_{\ell'} + \sum_{\ell' \in g^{excl}} \omega_{\ell'}^g ATT_{\ell'} \tag{16}$$

The weight $\omega^g_\ell = \sum_e \omega^g_{e,\ell}$ sums over the weights $\omega^g_{e,\ell}$ from Proposition 1, and is equal to the population regression coefficient from the following auxiliary regression:

$$D^\ell_{i,t} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \omega^g_{e,\ell} \mathbf{1}\{t - E_i \in g\} + v_{i,t} \tag{17}$$

which regresses $D^\ell_{i,t}$ on all bin indicators included in regression (4) and two-way fixed effects.

We note that even under treatment effect homogeneity μ_ℓ can still be contaminated by treatment effects from the excluded periods. This contamination, however, can be avoided by adjusting the specification to only exclude periods with zero treatment effect.

For specifications with relative time bins, we note that there can still be contamination from other bins as suggested by the second term of expression (16). A sufficient condition to avoid such contamination would be to group relative periods ℓ' into a bin only when their effects are the same since their weights $\omega^g_{\ell'}$ would sum to zero.

For the "fully dynamic" specification (7) where all g's are singletons of relative time periods, the weight $\omega^\ell_{\ell'}$ is zero for each relative period $\ell' \neq \ell$ that is included in the specification. The decomposition therefore simplifies to

$$\mu_\ell = ATT_\ell + \sum_{\ell' \in g^{excl}} \omega^g_{\ell'} ATT_{\ell'}. \tag{18}$$

3.6 Intuition for contamination

Proposition 3 demonstrates that even under the assumptions of parallel trends and no anticipation, estimates μ_g can still be contaminated by treatment effects from other periods. In this section we explain the intuition behind why this contamination occurs for the "fully dynamic" specification (7). A decomposition of μ_ℓ into a weighted average of $CATT_{e,\ell'}$ demonstrates that contamination is driven by the interaction of two elements: the weights and $CATT_{e,\ell'}$. The weights underlying the contamination are non-linear functions of the distribution of the cohorts. We do not attempt to provide a heuristic for determining the magnitude of the weights, but instead describe how to estimate the weights and later on in Section 4 how to estimate each $CATT_{e,\ell}$. This allows researchers to directly determine the degree of contamination in their application. Our publicly-available Stata package `eventstudyweights` automates the estimation of these weights using the panel dataset underlying any given specification of (1).

We apply the familiar omitted variable bias (OVB) formula to arrive at our decomposition. In an event study where individuals receive the treatment at different times, the panel can never be balanced in both calendar time and time relative to the initial treatment. As a result, the relative time indicators are still correlated even after controlling for unit and time fixed effects in a two-way fixed effects regression. We use the saturated regression and the OVB formula to illustrate how this correlation leads to contamination. We defer its formal derivation to Appendix B.

Under the parallel trends assumption only, the saturated regression is

$$
\begin{aligned}
Y_{i,t} = & \sum_{e} E[Y_{i,0}^{\infty} \mid E_i = e] \cdot \mathbf{1}\{E_i = e\} + \sum_{s} E[Y_{i,s}^{\infty} - Y_{i,0}^{\infty}] \cdot \mathbf{1}\{t = s\} \\
& + \sum_{\ell' \in g^{incl}} \sum_{e \in \mathcal{I}_{\ell'}} CATT_{e,\ell'} \cdot \left(D_{i,t}^{\ell'} \cdot \mathbf{1}\{E_i = e\} \right) \\
& + \sum_{\ell' \in g^{excl}} \sum_{e \in \mathcal{I}_{\ell'}} CATT_{e,\ell'} \cdot \left(D_{i,t}^{\ell'} \cdot \mathbf{1}\{E_i = e\} \right) + \epsilon_{i,t}
\end{aligned}
\tag{19}
$$

where the regressors are cohort fixed effects, time fixed effects, and cohort-specific relative time indicators. Furthermore, let g^{incl} collect the relative time included in (7). The coefficient associated with the cohort-specific relative time indicator $D_{i,t}^{\ell} \cdot \mathbf{1}\{E_i = e\}$ is the cohort-specific average treatment effects $CATT_{e,\ell}$. To decompose the coefficient μ_ℓ from (7) in terms of this saturated regression (19), the OVB formula multiplies the coefficients in the saturated regression (19), $CATT_{e,\ell'}$, with the regression coefficients from (12), $\omega_{e,\ell'}^{\ell}$, which leads to the following decomposition for μ_ℓ as a linear combination of $CATT_{e,\ell'}$:

$$
\mu_\ell = \sum_{e,\ell'} \omega_{e,\ell'}^{\ell} CATT_{e,\ell'}. \tag{20}
$$

Since $\omega_{e,\ell'}^{\ell}$ is equal to a regression coefficient from (12), we can write it as

$$
\omega_{e,\ell'}^{\ell} = (\sigma_{e,\cdot})^{\intercal} \Delta_{e+\ell'} A_{\ell}^{-1}. \tag{21}
$$

Below we briefly comment on each of the three elements in the above expression to highlight how they depend on the distribution of the cohorts. We defer their detailed definitions and derivations to Appendix A.1.

- $\sigma_{e,\cdot}$ is a vector of the covariance between cohort e and the other cohorts, namely $Cov\left(\mathbf{1}\{E_i = e\}, \mathbf{1}\{E_i = e'\}\right)$. This term thus scales quadratically in the share of cohort e, and is small for small cohorts.

- Δ_t is a matrix of demeaned relative time indicators. The entry that corresponds to cohort e' and relative time indicator $D_{i,t}^{\ell}$ is $E[D_{i,t}^{\ell} \mid E_i = e'] - \frac{1}{T+1}\mathbf{1}\{e' \in \mathcal{I}_{\ell}\}$. When T is large, i.e. the panel is long, the second term is small and this entry is therefore approximately equal to the relative time indicator.

- A_{ℓ}^{-1} is the row of A^{-1} that corresponds to the relative time indicator $D_{i,t}^{\ell}$ for A the covariance matrix of demeaned relative time indicators. Specifically, the entry in A that corresponds to the covariance between demeaned $D_{i,t}^{\ell}$ and $D_{i,t}^{\ell'}$ is

$$\sum_t Cov\left(D_{i,t}^{\ell}, D_{i,t}^{\ell'}\right) - \frac{1}{T+1} Cov\left(\mathbf{1}\{E_i \in \mathcal{I}_{\ell}\}, \mathbf{1}\{E_i \in \mathcal{I}_{\ell'}\}\right). \tag{22}$$

  Within any time period $D_{i,t}^{\ell}$ and $D_{i,t}^{\ell'}$ are negatively correlated because no cohorts can be in these two relative times at the same time. The second covariance term is in general also non-zero because being in one cohort predicts (not) being in another cohort. Therefore A is in general not a diagonal matrix and A^{-1} would depend on the distribution of the cohorts non-linearly.

The three elements of (21) demonstrate the weights are non-linear functions of the distribution of the cohorts, and they are in general non-zero. Nonetheless these weights can be estimated easily by the auxiliary regression (12).

3.7 Invalidity of pretrend tests based on pre-period coefficients.

Contamination undermines the practice of testing for pretrends using pre-period coefficients. Proposition 3 implies that when effects are not homogenous across cohorts, it is problematic to interpret non-zero estimates for μ_g as evidence for pretrends, where the set g contains some leads $\ell < 0$. Proposition 4 implies that even with homogeneous treatment effect, if the effects associated with the excluded periods are not zero, then contamination may still occur. Therefore without strong assumptions, pre-period coefficients should not be used to test for pretrends because contamination can lead to estimates that are non-zero in the absence of pretrends or zero in the presence of pre-trends.

Testing for pretrends using pre-period coefficients is commonly used in practice. As an example, He and Wang (2017) mention "the estimated coefficients of the leads of treatments, i.e. δ_k for all $k \leq -2$

are statistically indifferent from zero" as evidence for lack of pretrends. As another example, Chetty et al. (2014) assert "there is no trend toward higher individual pension contributions prior to year 0 ... as one would expect if individuals' tastes for saving were changing around the job switch" based on pre-period coefficient estimates. These tests are only appropriate when the authors are willing to make strong assumptions.

To provide further intuition for why this test is not meaningful without additional assumptions we walk through a simple example of the fully dynamic specification. Consider a balanced panel with $T = 2$ and cohorts $E_i \in \{1, 2\}$. There are at least two multi-colinearities from including all four relative time indicators. To form the fully dynamic specification we include $g^{incl} = \{-2, 0\}$ and exclude $g^{excl} = \{-1, 1\}$:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell \in \{-2,0\}} \mu_\ell D_{i,t}^\ell + \upsilon_{i,t}. \tag{23}$$

The choice of g^{excl} is based on the common practice of normalizing relative to the -1 period and distant lags.

When there are no never treated units, we can express the pre-trend coefficient μ_{-2} in terms of $CATT$s:

$$\mu_{-2} = \underbrace{CATT_{2,-2}}_{\text{own period}} + \underbrace{\frac{1}{2}CATT_{1,0} - \frac{1}{2}CATT_{2,0}}_{\ell' \in g^{incl}, \ell' \neq -2} + \underbrace{\frac{1}{2}CATT_{1,1} - CATT_{1,-1} - \frac{1}{2}CATT_{2,-1}}_{\ell' \in g^{excl}} \tag{24}$$

It is apparent the weights maintain the structure described in Proposition 1. Without any anticipation effect, the effects $CATT_{e,\ell<0}$ are zero and thus we expect μ_{-2} to be zero regardless of the cohort shares. With homogeneous treatment effect, cohorts 1 and 2 experience the same treatment effect at relative time 0 so that $CATT_{1,0}$ and $CATT_{2,0}$ cancel. But even with homogeneous treatment effect, the last term reflects the role of excluded periods as $CATT_{1,-1}$, $CATT_{2,-1}$ and $CATT_{1,1}$ receive non-zero weights. If there is any lagged effect and $CATT_{1,1}$ is non-zero, the coefficient μ_{-2} would be non-zero even without any anticipation effect. Note such behavior is independent of the distribution of the two cohorts.

We can further introduce never treated units to our example to demonstrate how the weight $\omega_{e,\ell'}^{-2}$ can be non-linear in the distribution of cohorts while maintaining the structure described in Proposition 1. In Figure 1 we plot the weights $\omega_{e,\ell'}^{-2}$ as we vary the distribution of cohorts. Specifically, we vary the total share of treated cohorts (shown on the x-axis), holding the shares of cohort 1 and 2 equal to each other and setting the remaining to be the share of never treated units. For any distribution of cohorts, we have $\omega_{2,-2}^{-2} = 1$ (not pictured). Panel (a) shows the weights for the included period $\ell' = 0$ while panel (b) shows the weights for the excluded periods $\ell' = -1$ or 1.

The example shown in Figure 1 provides a visualization of three takeaways regarding the contamination in the pre-trend coefficient μ_{-2}. First, both panels show that weights are a non-linear function of cohort shares. Second, panel (a) confirms the structure for weights associated with $\ell' \neq \ell$ but $\ell' \in g^{incl}$ as described in Proposition 1, namely $\sum_e \omega_{e,\ell'}^{-2} = 0$ for $\ell' \neq -2$. However, these weights have non-zero magnitude. When the effect is homogenous across cohorts, the contaminations are equal to $\sum_e \omega_{e,\ell'}^{-2} ATT_{\ell'}$ and cancel each other out. In contrast, when effects are heterogeneous the different $CATT_{e,\ell'}$ will not necessarily cancel and will contaminate the estimate for μ_{-2}. Third, panel (b) confirms the structure for weights associated with excluded periods as described in Proposition 1, namely $\sum_e \omega_{e,\ell' \in g^{excl}}^{-2} = -1$. In other words, contaminations from excluded periods can be thought of as a type of "normalization": a weighted average of excluded $CATT_{e,\ell'}$ is subtracted off the estimated treatment effect. However because these weights are not contained in $[0,-1]$, this average may lie outside of the convex hull of $CATT_{e,\ell'}$ for excluded periods. The latter issue can be alleviated by an assumption that $CATT_{e,\ell'}$ are the same in all excluded periods; we can fully avoid contamination from excluded period treatment effects by assuming all associated $CATT_{e,\ell'}$ are equal to zero.

4 Alternative estimation method

We propose a new estimation method that is robust to treatment effects heterogeneity. The goal of our method is to estimate a weighted average of $CATT_{e,\ell}$ for $\ell \in g$ with reasonable weights, namely weights that sum to one and are non-negative. In particular, we focus on the following weighted average of $CATT_{e,\ell}$, where the weights are shares of cohorts that experience at least ℓ periods relative to treatment, normalized by the size of g:

$$\nu_g = \frac{1}{|g|} \sum_{\ell \in g} \sum_{e} CATT_{e,\ell} Pr\{E_i = e \mid E_i \in [-\ell, T-\ell]\}. \tag{25}$$

One can aggregate $CATT_{e,\ell}$ to form other parameters of interest, such as those proposed by Callaway and Sant'Anna (2020a). We focus on the above aggregation ν_g since our goal is to improve the non-convex and non-zero weighting in μ_g. The weights in ν_g are guaranteed to be convex and have an interpretation as the representative shares corresponding to each $CATT_{e,\ell}$. Thus, our alternative estimator $\widehat{\nu}_g$ improves upon the two-way fixed effects estimator $\widehat{\mu}_g$ by estimating an interpretable weighted average of $CATT_{e,\ell \in g}$.

Our method proceeds by replacing each component in ν_g with its consistent estimator. We first estimate each $CATT_{e,\ell}$ using an interacted two-way fixed effects regression, then estimate the weight $Pr\{E_i = e \mid E_i \in [-\ell, T-\ell]\}$ using their sample analogs. In the final step, we average over the cohort-specific estimates associated with relative period ℓ. This method has a similar flavor as the method proposed by Gibbons et al. (2019). They first use an interacted model to estimate the treatment effect for each fixed effect group; the resulting group-specific estimates are averaged to provide the ATE. Their method improves fixed effects regressions in a cross-sectional setting, and our method builds on theirs by improving two-way fixed effects regressions in a panel setting. We therefore follow their terminology in calling our alternative estimator an "interaction-weighted" estimator.

4.1 Interaction-weighted estimator

We describe the estimation procedure in three steps (with more detailed definitions stated in Definition 4 of Online Appendix B).

Step 1. We estimate $CATT_{e,\ell}$ using a linear two-way fixed effects specification that interacts relative period indicators with cohort indicators, excluding indicators for cohorts from some set C:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \notin C} \sum_{\ell \neq -1} \delta_{e,\ell}(\mathbf{1}\{E_i = e\} \cdot D^{\ell}_{i,t}) + \epsilon_{i,t}. \tag{26}$$

The exact specification depends on the cohort shares for a given application. If there is a never-treated cohort, i.e. $\infty \in supp\{E_i\}$, then we may set $C = \{\infty\}$ and estimate regression (26) on all observations. If there are no never-treated units, i.e. $\infty \notin supp\{E_i\}$, then we may set $C = \{\max\{E_i\}\}$, i.e. the latest-treated cohort and estimate regression (26) on observations from $t = 0, \ldots, \max\{E_i\} - 1$. Lastly, if there is a cohort that is always treated, i.e. $0 \in supp\{E_i\}$, then we need to exclude this cohort from estimation.

The coefficient estimator $\widehat{\delta}_{e,\ell}$ from regression (26) is a DID estimator for $CATT_{e,\ell}$ with particular choices of pre-periods and control cohorts. As DID is likely a familiar estimator for applied researchers, we separate the more in-depth discussion of its definition, choices of pre-periods, and choices of control cohorts in Section 4.2.

Step 2. We estimate the weights $Pr\{E_i = e \mid E_i \in [-\ell, T-\ell]\}$ by sample shares of each cohort in the relevant period(s) $\ell \in g$.

Step 3. To form our IW estimator, we take a weighted average of estimates for $CATT_{e,\ell}$ from Step 1 with weight estimates from step 2. More formally, the IW estimator is

$$\widehat{\nu}_g = \frac{1}{|g|}\sum_{\ell\in g}\sum_{e}\widehat{\delta}_{e,\ell}\widehat{Pr}\{E_i = e \mid E_i \in [-\ell, T-\ell]\} \tag{27}$$

where $\widehat{\delta}_{e,\ell}$ is returned from step 1 and $\widehat{Pr}\{E_i = e \mid E_i \in [-\ell, T-\ell]\}$ is the estimated weight returned from step 2. We normalize the weights further by the size of g. If g is a singleton, then its size is $|g| = 1$.

Validity of the IW estimator. Under the parallel trends and no anticipation assumptions the coefficient estimator $\widehat{\delta}_{e,\ell}$ from regression (26) is a consistent estimator for $CATT_{e,\ell}$. The sample shares of each cohort are also consistent estimators for the population shares. Thus, the IW estimator is consistent for a weighted average of $CATT_{e,\ell}$ with weights equal to the share of each cohort in the relevant period(s).

With a few standard assumptions (which we present as Assumption 4 in Online Appendix B) on regression (26), we can show that each IW estimator is asymptotically normal and derive its asymptotic variance. The large sample approximation allows us to estimate the variance of IW estimators directly without relying on bootstrapping as in Callaway and Sant'Anna (2020a). However, we only construct pointwise confidence interval valid for a given IW estimator $\widehat{\nu}_g$. The bootstrap-based inference by Callaway and Sant'Anna (2020a) constructs simultaneous confidence intervals that are valid for the entire path of $\widehat{\nu}_g$.

4.2 Difference-in-differences estimator for $CATT_{e,\ell}$

Definition 2. Assume cohort e is non-empty i.e. $\sum_{i=1}^{N}\mathbf{1}\{E_i = e\} > 0$. Assume there exists some pre-period $s < e$ and some set of control cohorts $C \subseteq \{c : e+\ell < c \leq T\}$ that are non-empty i.e. $\sum_{i=1}^{N}\mathbf{1}\{E_i \in C\} > 0$. Using the notion $\mathbb{E}_N$ to abbreviate the symbol $\frac{1}{N}\sum_{i=1}^{N}$, the DID estimator with pre-period s and control cohorts C estimates $CATT_{e,\ell}$ as

$$\hat{\delta}_{e,\ell} = \frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s})\cdot\mathbf{1}\{E_i = e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]} - \frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s})\cdot\mathbf{1}\{E_i \in C\}]}{\mathbb{E}_N[\mathbf{1}\{E_i \in C\}]}. \tag{28}$$

The assumption of non-empty cohort e, existence of pre-period and non-empty control cohorts makes the DID estimator well-defined. For example, DID estimators for cohort 0 are not well-defined because a pre-period does not exist for this cohort, which is why we exclude them from estimating regression (26) in Step 1.

Using the above definition, in regression (26) from Step 1 of our proposed method, the coefficient estimator $\widehat{\delta}_{e,\ell}$ is a DID estimator for $CATT_{e,\ell}$ with pre-period $s = e-1$ (because we exclude relative period $\ell = -1$) and some choice of control cohorts C. If there is a never-treated cohort i.e. $\infty \in supp\{E_i\}$, then we set the control cohort to be never-treated units $C = \{\infty\}$. If there are no never-treated units, i.e. $\infty \notin supp\{E_i\}$, then we set the control cohort $C = \{\max\{E_i\}\}$, i.e. the latest-treated cohort. Among all possible pre-periods, for regression (26) from step 1 of our proposed method we choose $s = e-1$ and $C = \{\infty\}$ with never-treated units (or $\{\max\{E_i\}\}$ without never-treated units) because the resulting specification is a natural extension of the common specifications of two-way fixed effects regression.

Note that without never-treated units we need to drop time periods $t \geq \max\{E_i\}$ from estimating regression (26) because every unit will be treated in these periods. The DID estimators for $CATT_{e,\ell}$ for $e+\ell \geq \max\{E_i\}$ are thus not well-defined as the control cohort is empty $C = \emptyset$. For example, when there are just two cohorts with $E_i \in \{1, T\}$, one treated at $t = 1$ and the other treated in the last period, we then need to omit interaction terms involving the latest-treated cohorts as well as dropping observations from the last period $t = T$ from estimation.

Under some assumptions, this DID estimator $\widehat{\delta}_{e,\ell}$ is an unbiased and consistent estimator for $CATT_{e,\ell}$, a fact that we build on in deriving the probability limit of the IW estimator. We state this in the following proposition.

Proposition 5. *If Assumptions 1 and 2 hold, then the DID estimator using any pre-period $s < e$ and non-empty control cohorts C is an unbiased and consistent estimator for $CATT_{e,\ell}$.*

It is possible to relax the parallel trends assumption to allow the timing of treatment to depend on covariates. One can estimate $CATT_{e,\ell}$ consistently based on the inverse propensity score reweighted estimator proposed by Abadie (2005) and Callaway and Sant'Anna (2020a), the outcome regression approach by Heckman et al. (1997), and the doubly robust estimator recently proposed by Sant'Anna and Zhao (2020). The resulting estimates $\hat{\delta}_{e,l}$ can then be plugged into step 3 to form our IW estimator. In particular, without covariates and for the case with never treated units, our approach coincides with Callaway and Sant'Anna (2020a). Therefore one can use the `did` R package developed by Callaway and Sant'Anna (2020b) to form the IW estimator.

The choice of the pre-period s and the control cohorts C depends on the trade-off between relaxing Assumption 1 or Assumption 2. If Assumption 2 is likely to hold, we can choose any pre-period $s < e$ and include not-yet-treated units in control cohorts $C = \{c : c > e+\ell\}$ as in Callaway and Sant'Anna (2020a). This choice allows us to relax the parallel trends assumption to be just $E[Y^{\infty}_{i,e+\ell} - Y^{\infty}_{i,0} | E_i = e] = E[Y^{\infty}_{i,e+\ell} - Y^{\infty}_{i,0} | E_i > e + \ell]$. If instead we want to relax Assumption 2 to allow some anticipation, we then need the parallel trends assumption to hold. For example, suppose we are only willing to assume away anticipation 2 periods before a unit is treated, we may still be able to recover several $CATT_{e,\ell}$ by appropriately selecting a pre-period and a smaller set of control cohorts. We may choose any $s < e - 2$ for the pre-period, and for control cohorts, we may choose cohorts treated only after at least 2 periods from now $C \subseteq \{c : e + \ell + 2 < c \leq T\}$, or choose the latest-treated cohort $C = \{\max\{E_i\}\}$ as specified in regression (26).

5 Empirical Illustration

We illustrate our findings in the setting of Dobkin et al. (2018). Dobkin et al. (2018) study the economic consequences of hospitalization, which is a large source of economic risk for adults in the United States. To quantify these economic risks, in the first part of their analysis, Dobkin et al. (2018) leverage variation in the timing of hospitalization observed in the publicly-available dataset, Health and Retirement Study (HRS), which we describe in more detail in Section 5.1. Their estimation of the dynamic effects of hospitalization using two-way fixed effects regressions provides a good context for demonstrating our results. As we argue below in Section 5.2, parallel trends and no anticipation assumptions are plausible in this setting. However, the effect of hospitalization is potentially heterogenous across individuals hospitalized in different years. Our findings of non-convex and non-zero weighting would therefore apply to their two-way fixed effects regression estimates, and our alternative estimator could lead to different estimates. Furthermore, this dataset is publicly available, which allows us to provide replication files.

5.1 Data

Our sample selection closely follows Dobkin et al. (2018) but we include a cursory explanation here for completeness with an emphasis on how our final sample differs from their main analysis sample. Our primary source of data is the biennial Health and Retirement Study (HRS). We identify the sample of individuals who appear in two sequential waves of surveys and newly report having a hospital admission over the last two years (the "index" or initial admission) at the second survey. To focus on health "shocks", we restrict attention to non-pregnancy-related hospital admissions as in Dobkin et al. (2018). We also follow Dobkin et al. (2018) by focusing on adults who are hospitalized at ages 50-59.

Unlike Dobkin et al. (2018), we restrict our analysis to a subsample of these individuals who appear throughout waves 7-11 (roughly 2004-2012). Our sample of analysis therefore includes HRS respondents with index hospitalization during waves 8-11. The purpose of this sample restriction is to maintain a balanced panel with a reasonable sample size.

Here i indexes an individual, and t indexes survey wave ($T = 4$) and is normalized to zero for wave 7, the first wave in our sample. Among the outcomes $Y_{i,t}$ studied by Dobkin et al. (2018), we focus on two: out-of-pocket medical spending and labor earnings. They are derived from self-reports, adjusted to 2005 dollars and censored at the 99.95th percentile.

Summary statistics. Table 2 presents basic summary statistics for our analysis sample before hospitalization. We have a slightly lower fraction of white in our sample, but otherwise have a similar sample to Dobkin et al. (2018).

In Panel D, we compare means of the cross-sectional distributions of outcomes for individuals who have not been hospitalized by each wave. The size of the sample conditional on not having been hospitalized strictly decreases with each subsequent wave. There are apparent time trends in our outcomes of interest prior to hospitalization as we observe distributional changes across waves. Out-of-pocket medical spendings fluctuate and earnings decrease with each wave on average as more individuals are retired in each subsequent wave.

5.2 Setting

We illustrate how variations in the timing of hospitalization fit the event studies design proposed in Section 2. We define treatment $D_{i,t}$ to be ever having been hospitalized. In our terminology, we categorize individuals into cohorts based on E_i, which is defined as the survey wave of their initial hospitalization. Since we restrict the sample to individuals who were ever hospitalized in waves 8-11, there are four cohorts $E_i \in \{1, 2, 3, 4\}$. Although hospitalization itself may not be an absorbing state, we are trying to model the impact of having had any hospitalization. Thus, the cohort-specific average treatment effects $CATT_{e,\ell}$ trace out the path of treatment effects for cohort e following a negative health shock (even though the shock itself may be transient), as opposed to never having been hospitalized. Next we discuss whether each of the three identifying assumptions proposed is likely to hold in the context of unexpected hospitalizations.

Parallel trends (Assumption 1). Hospitalization is likely to be earlier among sicker individuals with high out-of-pocket medical spending and low labor earnings. Thus, it is not plausible that the baseline outcome $Y^{\infty}_{i,t}$ is mean independent of the timing of hospitalization. The parallel trends assumption is more plausible as it allows the timing to depend on unobserved time-invariant characteristics such as chronic disease. Furthermore, hospitalized individuals might be on a downward trend for labor earnings already prior to hospitalization compared to individuals who are never hospitalized. To reduce such confounding, we follow Dobkin et al. (2018) to restrict the parallel trends assumption to individuals who were ever hospitalized.

No anticipatory behavior (Assumption 2). It is plausible that there is no anticipatory behavior prior to the hospitalization, given that the treatment is restricted to conditions that are likely unexpected hospitalizations. This assumption may be violated if individuals have private information about the probability of these hospitalizations over time and thus adjust their behavior prior to hospitalization.

Treatment effect heterogeneity (Assumptions 3). For out-of-pocket medical spending, the effect of hospitalization is potentially heterogenous across individuals hospitalized in different waves. Individuals hospitalized in later waves are mechanically older at the time of hospitalization than individuals hospitalized in earlier waves. The effect on out-of-pocket medical spending is largely determined by generosity of health insurance, which may decrease as individuals age into Medicare. The effect on labor earnings is also likely heterogenous as it depends on the labor market condition at the time of hospitalization: for example, individuals hospitalized during the financial crisis may find it more difficult to return to the labor force, and suffer a more grave decrease in earnings.

5.3 Illustrating weights in two-way fixed effects regression

We illustrate our results on two-way fixed effects regression by estimating the following specification with indicators for up to three leads and lags following equation (3) of Dobkin et al. (2018)

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_{-3} D^{-3}_{i,t} + \mu_{-2} D^{-2}_{i,t} + \mu_0 D^0_{i,t} + \mu_1 D^1_{i,t} + \mu_2 D^2_{i,t} + \mu_3 D^3_{i,t} + \upsilon_{i,t}. \tag{29}$$

For their estimation, Dobkin et al. (2018) trim their sample, keeping only observations up to three waves prior to the hospitalization and three waves after the hospitalization, and weight their regression with survey weights. To fully illustrate issues with this specification, we do not trim, but rather use a sample balanced in calendar time for $t \in \{0, \ldots, 4\}$, and do not apply survey weights. With a sample balanced in calendar time, we need to exclude at least two relative period indicators due to multicollinearity. Following Dobkin et al. (2018), we exclude the period right before hospitalization ($\ell = -1$). We also exclude $\ell = -4$. Since we do not trim, note that our results are not directly comparable to Dobkin et al. (2018) even though they are quite similar.

We focus on a single coefficient μ_{-2} that is supposed to test for any pre-trend of hospitalizations. As in Proposition 2, we can decompose μ_{-2} as

$$\sum_{e=1}^{4} \omega_{e,0}^{-2} CATT_{e,0} + \sum_{\ell=-3,\neq -1}^{3} \sum_{e=1}^{4} \omega_{e,\ell}^{-2} CATT_{e,\ell} + \sum_{\ell'=-4,-1} \sum_{e=1}^{4} \omega_{e,\ell'}^{-2} CATT_{e,\ell'}. \tag{30}$$

As discussed in Proposition 1 we can estimate the underlying weights $\omega_{e,\ell}^{-2}$ by regressing $\mathbf{1}\{E_i = e\} \cdot D_{i,t}^{\ell}$ on the relative wave indicators included in specification (29) i.e. $\{D_{i,t}^{\ell}\}_{\ell=-3,\neq -1}^{3}$ and two-way fixed effects. The coefficient estimator of $D_{i,t}^{-2}$ in such regression, $\widehat{\omega}_{e,\ell}^{-2}$, consistently estimates $\omega_{e,\ell}^{-2}$.

Figure 3 plots these estimated weights. As described in our decomposition results, these weights have the following properties: (a) the four weights from relative wave $\ell = -2$ sum to one; (b) the weights for other included relative waves $\ell \in \{-3, 0, 1, 2, 3\}$ sum to zero for each included relative wave; and (c) the weights from excluded relative waves $\ell \in \{-4, -1\}$ sum to negative one across these excluded relative waves.

The weights are non-negative for lags of treatments, which suggest that the FE estimate $\widehat{\mu}_{-2}$ is particularly sensitive to estimates of the dynamic effects of hospitalizations and does not isolate the pre-trends. Specifically, treatment effects heterogeneity in $\ell = -3, 0, 1, 2$ can affect the FE estimate $\widehat{\mu}_{-2}$. Applied researchers can make similar plots to visualize the role of weights in their settings with our publicly-available Stata package `eventstudyweights` by Sun (2020).

5.4 Comparing FE and IW estimates

We illustrate our alternative method (IW estimator) following the three steps outline in Section 4.1. First, we estimate the interacted specification (26) as

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \in \{1,2,3\}} \sum_{\ell=-3,\neq -1}^{2} \delta_{e,\ell} \mathbf{1}\{E_i = e\} \cdot D_{i,t}^{\ell} + \epsilon_{i,t} \tag{31}$$

for $t = 0, \ldots, 3$. Specifically, we estimate $CATT_{e,\ell}$ using a DID estimator $\widehat{\delta}_{e,\ell}$ with pre-period $s = -1$ and control cohort $C = 4$, the cohort hospitalized in the last period. This means we need to drop $t = 4$ from estimation because everyone has been hospitalized by $t = 4$, and a control cohort for estimating $CATT_{e,\ell}$ in $t = 4$ does not exist. Second, we estimate the sample share of each cohort e across cohorts that experience at least l periods relative to hospitalization by its sample analog. Third, we form IW estimates $\widehat{\nu}_\ell$ by taking weighted averages of $\widehat{\delta}_{e,\ell}$ (returned from step one) with sample cohort share (returned from step two) as weights.

In Table 3, we report the FE estimates $\widehat{\mu}_\ell$ and the IW estimates $\widehat{\nu}_\ell$, as well as the underlying $CATT_{e,\ell}$ estimates $\widehat{\delta}_{e,\ell}$. In this application, the FE estimates and IW estimates happen to be very similar in their magnitude. The conclusion from Dobkin et al. (2018) based on FE estimates similar to ours still holds: we find a substantial and persistent decline in the earnings due to hospitalization, and the increase in out-of-pocket spending is transitory and small in comparison.

However, the FE estimates still suffer from non-convex and non-zero weighting as we show in Section 5.3, which can lead to interpretability issues depending on the amount of treatment effect heterogeneity. Even though for both outcomes $\widehat{\mu}_0$ falls in the convex hull of its underlying $CATT_{e,0}$ estimates, $\widehat{\mu}_{-2}$ turns out to be outside the convex hull of its underlying $CATT_{e,-2}$ estimates. In contrast, by construction, the IW estimates $\widehat{\nu}_\ell$ fall within the convex hull of its underlying $CATT_{e,\ell}$ estimates and are unaffected by $CATT_{e,\ell'}$ estimates from other periods $\ell' \neq \ell$. Thus, they have an interpretation as an average effect of the treatment on the treated ℓ periods after initial treatment.

6 Conclusions

This paper analyzes the behavior of relative period coefficients μ_ℓ on the indicator for being ℓ periods away from the treatment from two-way fixed effects regressions in settings with variation in treatment timing and treatment effects heterogeneity. For dynamic treatment effects, researchers are usually interested in estimating some average of treatment effects from ℓ periods relative to the treatment, and it is common to report the coefficient estimate $\widehat{\mu}_\ell$ assuming that interpretation is valid. However, we show that in the presence of heterogenous treatment effects, the coefficient μ_ℓ does not necessarily capture the dynamic treatment effect as it can fall outside the convex hull of $CATT_{e,\ell}$ from its corresponding period ℓ and may pick up spurious terms consisting of treatment effects from periods other than ℓ. This negative result is based on the decomposition of the coefficients μ_ℓ as a linear combination of cohort-specific average treatment effects on the treated $CATT_{e,\ell}$. The weights in the linear combination can be non-convex, and non-zero on relative periods $\ell' \neq \ell$.

Given these negative results on two-way fixed effects regression estimators, we propose "interaction-weighted" (IW) estimators for estimating dynamic treatment effects. The IW estimators are formed by first estimating $CATT_{e,\ell}$ with a regression saturated in cohort and relative period indicators, and then averaging estimates of $CATT_{e,\ell}$ across e at a given ℓ. These $CATT_{e,\ell}$ are identified under parallel trends and no anticipation assumptions. These estimators are easy to implement and robust to heterogenous treatment effects across cohorts; the IW estimator associated with relative period ℓ is guaranteed to estimate a convex average of $CATT_{e,\ell}$ using weights that are sample share of each cohort e.

Finally, we illustrate the empirical relevance of our results by estimating the dynamic effect of hospitalization on the out-of-pocket medical spending and labor earnings using a setup similar to Dobkin et al. (2018). We find non-convex and non-zero weights for two-way fixed effects regression in this example, and show that the resulting estimates indeed sometimes fall outside the convex hull of the underlying *CATT* estimates due to contamination by treatment effects from other relative periods. IW estimates, on the other hand, are weighted averages of the underlying *CATT* estimates with weights representative of cohort share.

More broadly, our paper suggests that researchers can do more in this context to assess the validity of underlying assumptions and how violations may impact their estimates. We demonstrate the sensitivity of two-way fixed effects regressions to underlying assumptions and provide researchers with tools to assess and address these issues. Specifically, for average treatment effects estimated using two-way fixed effects regressions we recommend that empirical researchers directly estimate the underlying weights on cohort-specific average treatment effects using our proposed auxiliary regressions. This exercise allows empirical researchers to assess the degree of potential contamination under any assumed structure of treatment effects heterogeneity. We also recommend empirical researchers consider estimation methods more robust to treatment effects heterogeneity such as the IW estimator we propose. Developing additional tools for applied researchers to identify and estimate interpretable parameters of interest is a promising avenue for future research.

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 1978, *60* (1), 47–57.

Athey, Susan and Guido Imbens, "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption," Working Paper arXiv:1808.05293 [econ.EM], ArXiV August 2018.

Bailey, Martha J. and Andrew Goodman-Bacon, "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans," *American Economic Review*, March 2015, *105* (3), 1067–1104.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and John van Reenen, "Evaluating the Employment Impact of a Mandatory Job Search Program," *Journal of the European Economic Association*, 2004, *2* (4), 569–606.

Borusyak, Kirill and Xavier Jaravel, "Revisiting Event Study Designs," Working Paper, Harvard University May 2017.

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

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences With Multiple Time Periods and an Application on the Minimum Wage and Employment," Working Paper arXiv:1803.09015 [econ.EM], ArXiV March 2020.

_ **and Pedro H.C. Sant'Anna**, *did: Treatment Effects with Multiple Periods and Groups* 2020. R package version 1.2.3.

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey, "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 2013, *81* (2), 535–580.

Chetty, Raj, John N. Friedman, SÃžren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen, "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark," *The Quarterly Journal of Economics*, 2014, *129* (3), 1141–1219.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, September 2020, *110* (9), 2964–96.

Deryugina, Tatyana, "The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance," *American Economic Journal: Economic Policy*, August 2017, *9* (3), 168–98.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo, "The Economic Consequences of Hospital Admissions," *American Economic Review*, February 2018, *108* (2), 308–52.

Gibbons, Charles E., Juan Carlos Suárez Serrato, and Michael B. Urbancic., "Broken or Fixed Effects?," *Journal of Econometric Methods*, January 2019, *8* (1), 1–12.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," Working Paper 25018, National Bureau of Economic Research September 2018.

He, Guojun and Shaoda Wang, "Do College Graduates Serving as Village Officials Help Rural China?," *American Economic Journal: Applied Economics*, October 2017, *9* (4), 186–215.

Heckman, James, Hidehiko Ichimura, and Petra E. Todd, "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 1997, *64* (4), 605–654.

_ , _ , Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, *66* (5), 1017–1098.

Hendren, Nathaniel, "Knowledge of Future Job Loss and Implications for Unemployment Insurance," *American Economic Review*, July 2017, *107* (7), 1778–1823.

Malani, Anup and Julian Reif, "Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform," *Journal of Public Economics*, 2015, *124*, 1 – 17.

Rambachan, Ashesh and Jonathan Roth, "An Honest Approach to Parallel Trends," Working Paper, Harvard University September 2020.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," Working Paper, Harvard University July 2019.

Sant'Anna, Pedro H.C. and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 2020, *219* (1), 101 – 122.

Sianesi, Barbara, "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *The Review of Economics and Statistics*, 2004, *86* (1), 133–155.

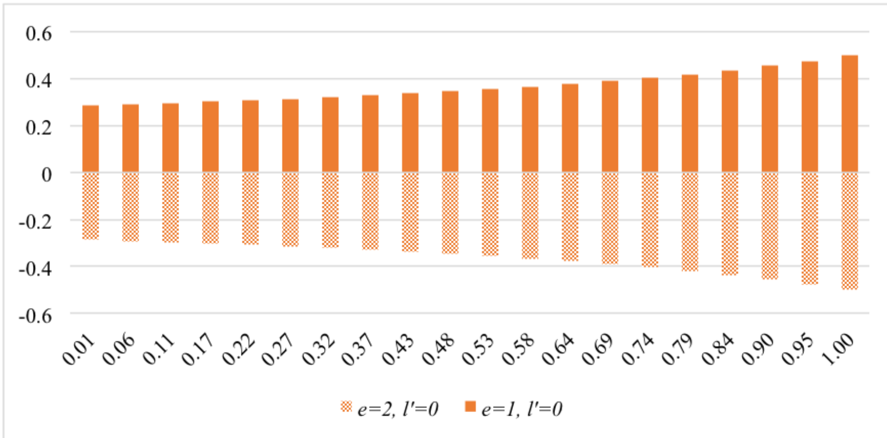
Słoczyński, Tymon, "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *The Review of Economics and Statistics*, forthcoming.

Sun, Liyang, "EVENTSTUDYWEIGHTS: Stata module to estimate the implied weights on the cohort-specific average treatment effects on the treated (CATTs) (event study specifications)," Statistical Software Components, Boston College Department of Economics September 2020.

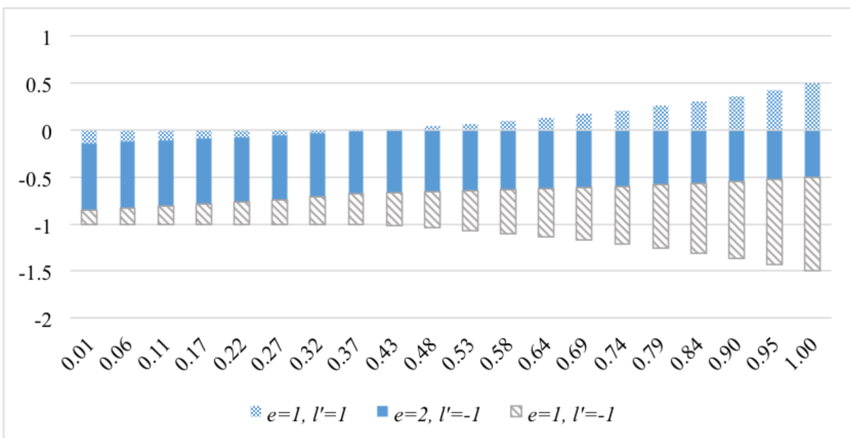
Tewari, Ishani, "The Distributive Impacts of Financial Development: Evidence from Mortgage Markets during US Bank Branch Deregulation," *American Economic Journal: Applied Economics*, October 2014, *6* (4), 175–96.

Figure 1: Weight on $CATT_{e,\ell'}$ for $\ell' \neq -2$ in the Coefficient μ_{-2} from Regression (23)

(a) Weight on $CATT_{e,0}$



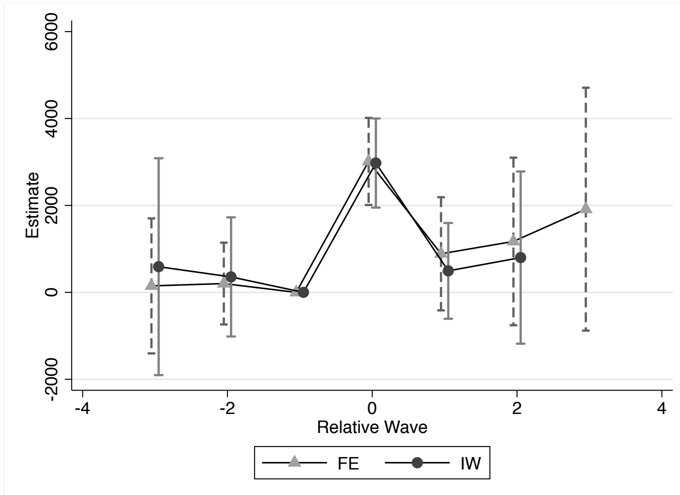
(b) Weight on $CATT_{e,\ell'}$ for $\ell' \in \{-1, 1\}$



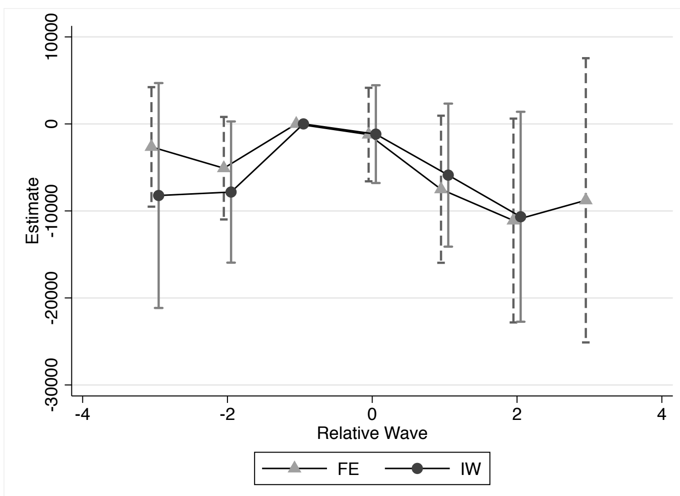
Notes: Panel (a) plots the weight on $CATT_{e,0}$ and panel (b) plots the weight on $CATT_{e,\ell'}$ for $\ell' \in \{-1, 1\}$ under a given distribution of cohorts indexed by the x-axis. Specifically, we construct a panel balanced in calendar time with $T = 2$ and cohorts $E_i \in \{1, 2, \infty\}$. We vary the share of never-treated units $Pr\{E_i = \infty\}$ between $[0, 0.99]$ and divide the rest of the units evenly into cohorts treated at time 1 and 2. The x-axis indexes $1 - Pr\{E_i = \infty\}$.

Figure 2: FE vs IW Estimates of the Effects of Hospitalization on Outcomes

(a) Out-of-pocket Medical Spending

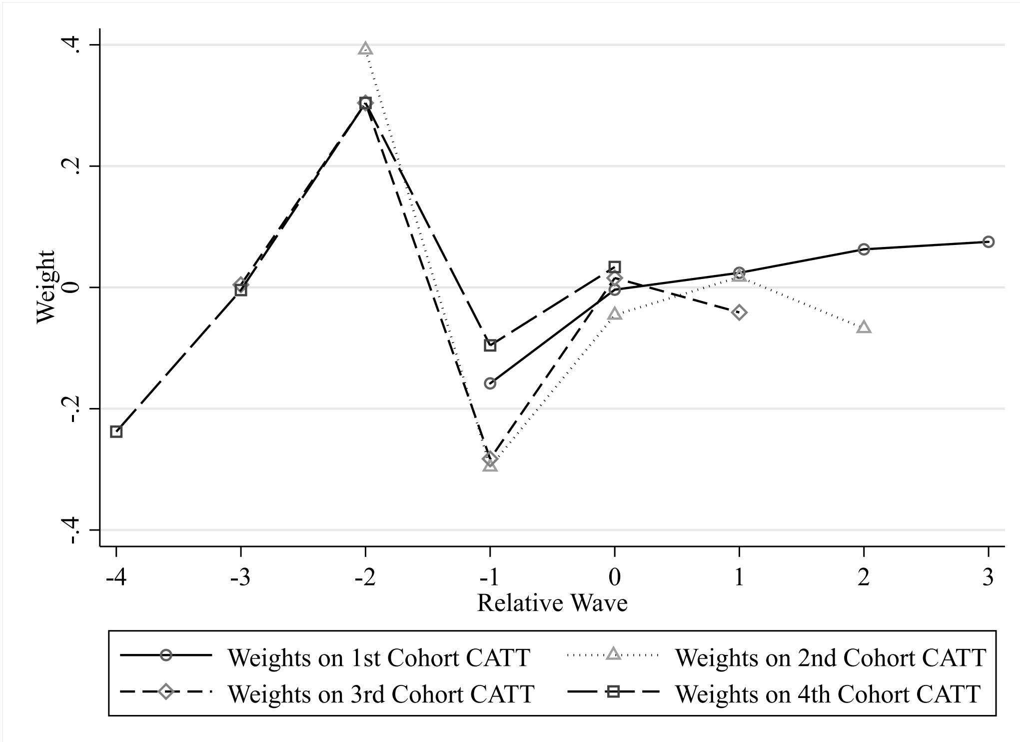


(b) Labor Earnings



Notes: Each figure plots FE estimates $\widehat{\mu}_\ell$ from regression (29) in triangles and IW estimates $\widehat{\nu}_\ell$ from regression (31) in circles against relative wave ℓ, with their respective pointwise 95% confidence intervals. Both are estimates for the effect of hospitalization at relative wave ℓ. The outcome variable is out-of-pocket medical spending in panel (a) and labor earnings in panel (b) respectively.

Figure 3: Estimated Weights $\widehat{\omega}_{e,\ell}^{-2}$ Underlying μ_{-2}



Notes: The FE estimator for the pretrend of hospitalization $\widehat{\mu}_{-2}$ estimates a linear combination of cohort-specific effects $CATT_{e,\ell}$'s from all cohorts e and relative waves ℓ as specified in equation (30). This figure plots the estimated weight $\widehat{\omega}_{e,\ell}^{-2}$ associated with each $CATT_{e,\ell}$ in this linear combination.

Table 1: Survey of Applied Papers

Paper	Binary Absorbing Treatment	Variation in Treatment Timing	Pre-Treatment Relative Period Excluded	Exclude Distant Relative Periods	Bin Distant Relative Periods	Includes Never Treated Units	Panel Balanced in Relative Time
Bosch and Campos-Vazquez (2014)	X						
Fitzpatrick and Lovenheim (2014)							
Gallagher (2014)	X	X	-1		X	X	
Tewari (2014)	X	X	0				X
Ujhelyi (2014)	X	X	-1		X	X	
Bailey and Goodman-Bacon (2015)	X	X	-1		X	X	X
Deryugina (2017)	X	X	-1		X	X	X
Deschenes et al. (2017)	X						
He and Wang (2017)	X	X	-1		X		
Lafortune et al. (2017)	X	X	0	X		X	
Kuziemko et al. (2018	X	X	-1		X		
Markevich and Zhuravskaya (2018)							

Notes: This table consolidates key properties of main event study specifications across a sample of applied papers. We follow the selection criteria in Roth (2019): the original sample consists of 70 total papers, but is further constrained to these twelve papers with publicly available data and code. The data and code are used to determine exactly the specification estimated in these papers. We focus on the first specification underlying the event study estimates in each paper, which we view as a reasonable proxy for the main specification in the paper. Note that two papers (Fitzpatrick and Lovenheim (2014); Markevich and Zhuravskaya (2018)) have none of the attributes listed in the columns.

Table 2: Summary Statistics of the HRS Sample

	N	Mean	Std. Dev
Panel A. Demographics			
Age at admission	656	56	2.29
Male	656	0.456	0.498
Year of admission	656	2,007	2.11
Panel B. Race/ethnicity			
Hispanic	656	0.122	0.327
Black	656	0.151	0.358
White	656	0.742	0.438
Other race	656	0.107	0.309
Panel C. Insurance status			
Medicaid	656	0.05	0.219
Private	656	0.715	0.452
Medicare	656	0.072	0.259
Panel D. Pre-hospitalization outcome			
Out-of-pocket medical spending			
Wave 7 ($E_i \geq 1$)	656	3,302	9,024
Wave 8 ($E_i \geq 2$)	404	2,355	8,132
Wave 9 ($E_i \geq 3$)	228	2,056	3,532
Wave 10 ($E_i = 4$)	65	2,044	4,379
Earnings			
Wave 7 ($E_i \geq 1$)	656	43,810	67,950
Wave 8 ($E_i \geq 2$)	404	38,944	58,601
Wave 9 ($E_i \geq 3$)	228	36,274	56,768
Wave 10 ($E_i = 4$)	65	29,037	46,289

Notes: This table presents summary statistics on our primary analysis sample, taken from the biennial Health and Retirement Survey (HRS). We include the sample of individuals ages 50-59 in waves 7-11 (approximately spanning 2004-2012) who appear in two sequential survey waves and report a recent hospital admission in the second survey. For panel D, the sample corresponding to wave t is conditional on not having hospitalization by wave t.

Table 3: Estimates for the Effect of Hospitalization on Outcomes

(a) Out-of-pocket Medical Spending

	FE	IW		$CATT_{e,\ell}$	
ℓ Wave Relative to Hospitalization	$\widehat{\mu}_\ell$	$\widehat{\nu}_\ell$	$\widehat{\delta}_{1,\ell}$	$\widehat{\delta}_{2,\ell}$	$\widehat{\delta}_{3,\ell}$
-3	149	591	-	-	591
	(792)	(1273)			(1273)
-2	203	353	-	299	411
	(480)	(698)		(967)	(1030)
-1	0	0	0	0	0
0	3,013	2,960	2,826	3,031	3,092
	(511)	(543)	(1038)	(704)	(998)
1	888	530	825	107	-
	(664)	(587)	(912)	(653)	
2	1,172	800	800	-	-
	(983)	(1010)	(1010)		
3	1,914	-	-	-	-
	(1426)				

(b) Labor Earnings

	FE	IW		$CATT_{e,\ell}$	
ℓ Wave Relative to Hospitalization	$\widehat{\mu}_\ell$	$\widehat{\nu}_\ell$	$\widehat{\delta}_{1,\ell}$	$\widehat{\delta}_{2,\ell}$	$\widehat{\delta}_{3,\ell}$
-3	-2,642	-8,228	-	-	-8,228
	(3504)	(6592)			(6592)
-2	-5,089	-7,823	-	-7,691	-7,964
	(3005)	(4197)		(6348)	(4766)
-1	0	0	0	0	0
0	-1,225	-863	455	3,032	-7,107
	(2742)	(2967)	(5593)	(4152)	(5883)
1	-7,508	-5,435	-1,670	-10,826	-
	(4312)	(4348)	(6500)	(4863)	
2	-11,102	-10,670	-10,670	-	-
	(5976)	(6154)	(6154)		
3	-8,780	-	-	-	-
	(8332)				

Notes: This table reports three different sets of estimates for the dynamic effects of hospitalization on out-of-pocket medical spending in panel (a) and labor earnings in panel (b). The first column reports estimates from the FE estimator $\widehat{\mu}_\ell$. The sample includes observations from wave $t = 0, 1, 2, 3, 4$. Columns 3-5 report the estimates for $CATT_{e,\ell}$ from $\widehat{\delta}_{e,\ell}$. The sample includes observations from wave $t = 0, 1, 2, 3$. Column 2 reports the IW estimates which are constructed as the weighted average of $\widehat{\delta}_{e,\ell}$'s across cohorts e who are ℓ periods from hospitalization. Standard errors (clustered on the individual) are shown in parentheses.

Online Appendix

A Notation glossary and properties of double-demeaning

In this section, we provide relevant notations for proofs in Section B. Section A.1 introduces detailed expression for the weight.

We observe a balanced panel of N i.i.d. observations $\{\{Y_{i,t}, D_{i,t}\}_{t=0}^{T}\}_{i=1}^{N}$ where $Y_{i,t} \in \mathbb{R}$ is a real valued outcome variable and $D_{i,t} \in \{0,1\}$ is a binary absorbing treatment status variable: $D_{i,t} = 0$ if i is not treated in period t and $D_{i,t} = 1$ if i is treated in period t.

Since the treatment is absorbing, we can aggregate the treatment path $\{D_{i,t}\}_{t=0}^{T}$ to a single discrete random variable $E_i = \min\{t : D_{i,t} = 1\}$, which is the period of the initial treatment. Additionally, we define $D_{i,t}^{\ell} := \mathbf{1}\{t - E_i = \ell\}$ to be an indicator for being ℓ time periods relative to unit i's initial treatment.

We focus on the following two-way fixed effects regression

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \mu_g \mathbf{1}\{t - E_i \in g\} + \upsilon_{i,t}$$

where g are disjoint sets of relative times, α_i and λ_t are unit and time fixed effects.

Definition 3. For fixed T, consider a random vector $\{X_{i,t}\}_{t=0}^{T} \in \mathbb{R}^{T+1}$, jointly distributed according to P. At each t, let $\ddot{X}_{i,t}$ denote the following random variable:

$$\ddot{X}_{i,t} = X_{i,t} - E[X_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} X_{i,s} + \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}] \tag{32}$$

The expectation $E[X_{i,t}]$ is taken cross-sectionally with respect to P at time $t = 0, \ldots, T$.

Remark 1. Conventionally $\ddot{X}_{i,t}$ is referred to as the double-demeaned version of $X_{i,t}$, as it removes the contemporaneous expectation $E[X_{i,t}]$ as well as the simple average across time for unit i, $\frac{1}{T+1}\sum_{s=0}^{T} X_{i,t}$. It adds back a simple average across time of the expectations so that $\ddot{X}_{i,t}$ has the zero-mean and zero-sum properties as detailed in Lemma 1.

Lemma 1. (Properties of double-demeaning.) *For any $\ddot{X}_{i,t}$ and $\ddot{Z}_{i,t}$, double-demeaned versions of $X_{i,t}$ and $Z_{i,t}$ respectively, we have the following properties:*

- zero-mean and zero-sum: $E[\ddot{X}_{i,t}] = 0$ and $\sum_{t=0}^{T} \ddot{X}_{i,t} = 0$;
- idempotent: $\sum_{t=0}^{T} E[\ddot{X}_{i,t}\ddot{Z}_{i,t}] = \sum_{t=0}^{T} E[X_{i,t}\ddot{Z}_{i,t}] = \sum_{t=0}^{T} E[\ddot{X}_{i,t} Z_{i,t}]$;

- for any time-invariant random variables $Z_{i,t}$ such that $Z_{i,t} = Z_i$, double-demeaning annihilates it: $\ddot{Z}_{i,t} = 0$.

Proof. The zero-mean and zero-sum properties hold by definition of double-demeaning:

$$E[\ddot{X}_{i,t}] = E\left[X_{i,t} - E[X_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} X_{i,s} + \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}]\right] \tag{33}$$

$$= E[X_{i,t}] - E[X_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}] + \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}] = 0 \tag{34}$$

and

$$\sum_{t=0}^{T} \ddot{X}_{i,t} = \sum_{t=0}^{T} X_{i,t} - \sum_{t=0}^{T} E[X_{i,t}] - \sum_{t=0}^{T} \frac{1}{T+1}\sum_{s=0}^{T} X_{i,s} + \sum_{t=0}^{T} \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}] \tag{35}$$

$$= \sum_{t=0}^{T} X_{i,t} - \sum_{t=0}^{T} E[X_{i,t}] - \sum_{s=0}^{T} X_{i,s} + \sum_{s=0}^{T} E[X_{i,s}] = 0 \tag{36}$$

For the idempotent property, first note that by definition of double-demeaning, rearranging terms we can write $E[\ddot{X}_{i,t}\ddot{Z}_{i,t}]$ as

$$= E\left[\left(X_{i,t} - \frac{1}{T+1}\sum_{s=0}^{T} X_{i,s}\right)\ddot{Z}_{i,t}\right] - E\left[\left(E[X_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}]\right)\ddot{Z}_{i,t}\right] \tag{37}$$

$$= E\left[\left(X_{i,t} - \frac{1}{T+1}\sum_{s=0}^{T} X_{i,s}\right)\ddot{Z}_{i,t}\right] - \left(E[X_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} E[X_{i,s}]\right)\cdot E[\ddot{Z}_{i,t}] \tag{38}$$

By the zero-mean property $E[\ddot{Z}_{i,t}] = 0$, the second term of the above expression is zero. Summing the first term over t, we have

$$\sum_{t=0}^{T} E[X_{i,t}\ddot{Z}_{i,t}] - \sum_{t=0}^{T} E\left[\left(\frac{1}{T+1}\sum_{s=0}^{T} X_{i,s}\right)\ddot{Z}_{i,t}\right] = \sum_{t=0}^{T} E[X_{i,t}\ddot{Z}_{i,t}] - E[\left(\frac{1}{T+1}\sum_{s=0}^{T} X_{i,s}\right)\sum_{t=0}^{T} \ddot{Z}_{i,t}] = \sum_{t=0}^{T} E[X_{i,t}\ddot{Z}_{i,t}] \tag{39}$$

where the last equality follows from the zero-sum property $\sum_{t=0}^{T} \ddot{Z}_{i,t} = 0$. This proves $\sum_{t=0}^{T} E[\ddot{X}_{i,t}\ddot{Z}_{i,t}] = \sum_{t=0}^{T} E[X_{i,t}\ddot{Z}_{i,t}]$. Similarly we can show $\sum_{t=0}^{T} E[\ddot{X}_{i,t}\ddot{Z}_{i,t}] = \sum_{t=0}^{T} E[\ddot{X}_{i,t} Z_{i,t}]$.

The annihilating property of double-meaning holds by definition for any time-invariant random variable. Specifically, plugging in $Z_{i,t} = Z_i$ we have

$$\ddot{Z}_{i,t} = Z_{i,t} - E[Z_{i,t}] - \frac{1}{T+1}\sum_{s=0}^{T} Z_{i,s} + \frac{1}{T+1}\sum_{s=0}^{T} E[Z_{i,s}] \tag{40}$$

$$= Z_i - E[Z_i] - Z_i + E[Z_i] = 0 \tag{41}$$

□

A.1 Expression of the weight

In this section, we provide detailed definitions for the three elements in the expression of the weight $\omega^{\ell}_{e,\ell'} = \sigma'_{e,\cdot}\Delta_{e+\ell'}A_{\ell}^{-1}$. Recall $\omega^{\ell}_{e,\ell'}$ is the weight associated with $CATT_{e,\ell'}$ in the linear combination underlying μ_{ℓ}:

$$\begin{aligned}
\omega^{\ell}_{e,\ell'} &= \underbrace{\mathbf{e}_{\ell}^{\intercal}\left(\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}}_{:=(A_{\ell}^{-1})^{\intercal}} E[\ddot{\mathbf{D}}_{i,t}D^{\ell'}_{i,t}\cdot\mathbf{1}\{E_i = e\}] \\
&= (A_{\ell}^{-1})^{\intercal}\sum_{t=0}^{T} Cov(\mathbf{D}_{i,t} - \frac{1}{T+1}\sum_{t=0}^{T}\mathbf{D}_{i,t}, D^{\ell'}_{i,t}\cdot\mathbf{1}\{E_i = e\} - \frac{1}{T+1}\sum_{t=0}^{T} D^{\ell'}_{i,t}\cdot\mathbf{1}\{E_i = e\}) \\
&= (A_{\ell}^{-1})^{\intercal}Pr\{E_i = e\}\cdot\Big((E[\mathbf{D}_{i,e+\ell'} - \frac{1}{T+1}\sum_{t=0}^{T}\mathbf{D}_{i,e+\ell'} \mid E_i = e]) - (E[\mathbf{D}_{i,e+\ell'}] - E[\frac{1}{T+1}\sum_{t=0}^{T}\mathbf{D}_{i,e+\ell'}])\Big) \\
&= (A_{\ell}^{-1})^{\intercal}\sum_{e'} E[\mathbf{D}_{i,e+\ell'} - \frac{1}{T+1}\sum_{t}\mathbf{D}_{i,t} \mid E_i = e']\cdot Cov\left(\mathbf{1}\{E_i = e\}, \mathbf{1}\{E_i = e'\}\right) \\
&:= (\sigma_{e,\cdot})^{\intercal}\Delta_{e+\ell'}A_{\ell}^{-1}
\end{aligned}$$

We note that

- The first term is the vector $\sigma_{e,\cdot} = (\sigma_{e,e'})_{e'\in supp(E_i)}$ of the covariance between cohort indicators. Each entry $\sigma_{e,e'} = Cov\left(\mathbf{1}\{E_i = e\}, \mathbf{1}\{E_i = e'\}\right)$ is the covariance between being in cohort e and being in cohort e'.

- The second term is the matrix $\Delta_{e+\ell'}$ of demeaned relative time indicators. The rows of the matrix are indexed by cohorts $e' \in supp(E_i)$ and the columns of the matrix are indexed by relative time indicators included in the specification (7). The entry that corresponds to cohort e' and relative time indicator $D^{\ell}_{i,t}$ is $E[D^{\ell}_{i,e+\ell'} - \frac{1}{T+1}\sum_t D^{\ell}_{i,t} \mid E_i = e']$. Note that the sum $\sum_t D^{\ell}_{i,t}$ is equal to $\mathbf{1}\{E_i \in \mathcal{I}_{\ell}\}$ the indicator for whether this individual ever experience the relative time ℓ.

- The third term is the row vector of A^{-1} that corresponds to the relative time indicator $D^{\ell}_{i,t}$. Here A is the covariance matrix of demeaned relative time indicators included in the specification (7). Specifically, the entry that corresponds to the covariance between demeaned $D^{\ell}_{i,t}$ and $D^{\ell'}_{i,t}$ is
$$A = \sum_t Cov\left(D^{\ell}_{i,t} - \frac{1}{T+1}\sum_t D^{\ell}_{i,t}, D^{\ell'}_{i,t} - \frac{1}{T+1}\sum_t D^{\ell'}_{i,t}\right).$$
Plugging in $\sum_t D^{\ell}_{i,t} = \mathbf{1}\{E_i \in \mathcal{I}_{\ell}\}$ gives the expression in the main text.

B Proofs of decomposition results

Proof of Proposition 1

Proof. Collect the relative time indicators in a column vector $\mathbf{D}_{i,t} = (\mathbf{1}\{t - E_i \in g\})^{\intercal}_{g\in\mathcal{G}}$. Collect their corresponding coefficients in a column vector $\mu = (\mu_g)^{\intercal}_{g\in\mathcal{G}}$. Partialling out the unit and time fixed effects, Regression (4) is

$$\ddot{Y}_{i,t} = \mu^{\intercal}\ddot{\mathbf{D}}_{i,t} + \upsilon_{i,t} \tag{42}$$

where $\ddot{X}_{i,t}$ is time- and cross-sectional demeaned version of $X_{i,t}$ as defined in Definition 3. By the idempotent property of Lemma 1

$$\mu_g = \mathbf{e}_g^{\intercal}\left(\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}Y_{i,t}] \tag{43}$$

where $\mathbf{e}_g$ is a column vector with one in the entry corresponding to the entry of $\mathbf{1}\{t - E_i \in g\}$ in $\mathbf{D}_{i,t}$, and zero otherwise.

To further develop the expression for the population regression coefficients μ_g, we note that by Lemma 1, we have $\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}Y_{i,0}^{\infty}] = \sum_{t=0}^{T} E[\mathbf{D}_{i,t}\ddot{Y}_{i,0}^{\infty}] = 0$ since $Y_{i,0}^{\infty}$ is time-invariant, and $E[\ddot{\mathbf{D}}_{i,t}] = 0$ by the zero-mean property.

$$\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}Y_{i,t}] = \sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}Y_{i,t}] - \underbrace{\sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}Y_{i,0}^{\infty}]}_{=0} = \sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}\left(Y_{i,t} - Y_{i,0}^{\infty}\right)] = \sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}E[Y_{i,t} - Y_{i,0}^{\infty}|E_i]] \tag{44}$$

$$= \sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}E[Y_{i,t} - Y_{i,0}^{\infty}|E_i]] - \sum_{t=0}^{T}\underbrace{E[\ddot{\mathbf{D}}_{i,t}]}_{=0}E[Y_{i,t}^{\infty} - Y_{i,0}^{\infty}] \tag{45}$$

$$= \sum_{t=0}^{T} E[\ddot{\mathbf{D}}_{i,t}\left(E[Y_{i,t} - Y_{i,0}^{\infty}|E_i] - E[Y_{i,t}^{\infty} - Y_{i,0}^{\infty}]\right)]. \tag{46}$$

We can abbreviate the term in the parentheses as $f(E_i, t)$ to emphasize it is a function of E_i and t. Since E_i and t take on discrete values, we can write

$$E[Y_{i,t} - Y_{i,0}^{\infty}|E_i] - E[Y_{i,t}^{\infty} - Y_{i,0}^{\infty}] =: f(E_i, t) = \sum_{e=0}^{\infty} f(e,t)\cdot\mathbf{1}\{E_i = e\} \tag{47}$$

$$= \sum_{e=0}^{\infty}\sum_{\ell=-T}^{T} f(e, e+\ell)\cdot\mathbf{1}\{t - e = \ell\}\cdot\mathbf{1}\{E_i = e\} \tag{48}$$

$$= \sum_{\ell=-T}^{T}\sum_{e=0}^{\infty} D_{i,t}^{\ell}\mathbf{1}\{E_i = e\}\cdot f(e, e+\ell). \tag{49}$$

Even though we sum over $e \in \{0, \ldots, \infty\}$, if there are no never-treated units (or no units in a particular cohort

e'), then we have no units take on values $E_i = \infty$ (or $E_i = e'$). The range of the summation is therefore still sensible. Note that we replace $\mathbf{1}\{t-e=\ell\}\cdot\mathbf{1}\{E_i=e\}$ with $D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}$ because

$$D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}=\mathbf{1}\{t-E_i=\ell\}\cdot\mathbf{1}\{E_i=e\}=\mathbf{1}\{t-e=\ell\}\cdot\mathbf{1}\{E_i=e\}. \tag{50}$$

Using the above expression, the coefficient μ_g can be written as

$$\mu_g = \mathbf{e}_g^{\intercal}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}\sum_{t=0}^{T}\sum_{\ell=-T}^{T}\sum_{e=0}^{\infty}E[\ddot{\mathbf{D}}_{i,t}D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}]f(e,e+\ell) \tag{51}$$

$$=\sum_{t=0}^{T}\sum_{l\in g}\sum_{e=0}^{\infty}\underbrace{\mathbf{e}_g^{\intercal}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}E[\ddot{\mathbf{D}}_{i,t}D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}]}_{=\omega^{g}_{e,\ell}}f(e,e+\ell) \tag{52}$$

$$+\sum_{t=0}^{T}\sum_{g'\neq g}\sum_{\ell\in g'}\sum_{e=0}^{\infty}\underbrace{\mathbf{e}_g^{\intercal}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}E[\ddot{\mathbf{D}}_{i,t}D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}]}_{=\omega^{g}_{e,\ell}}f(e,e+\ell) \tag{53}$$

$$+\sum_{t=0}^{T}\sum_{\ell\in g^{excl}}\sum_{e=0}^{\infty}\underbrace{\mathbf{e}_g^{\intercal}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^{\intercal}]\right)^{-1}E[\ddot{\mathbf{D}}_{i,t}D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}]}_{=\omega^{g}_{e,\ell}}f(e,e+\ell). \tag{54}$$

The superscript g in $\omega^{g}_{e,\ell}$ indexes μ_g. The subscript ℓ in $\omega^{g}_{e,\ell}$ indexes $f(e,e+\ell)$. The expression above the braces makes it clear that the weight $\omega^{g}_{e,\ell}$ is equal to the population regression coefficient on $\mathbf{1}\{t-E_i\in g\}$ from regressing $D^{\ell}_{i,t}\cdot\mathbf{1}\{E_i=e\}$ on all the bin indicators i.e. $\{\mathbf{1}\{t-E_i\in g\}\}_{g\in\mathcal{G}}$ and two-way fixed effects included in (4). This proves Proposition 1. We next show several properties of these weights in the expression of μ_g:

1. For relative times of μ_g's own bin i.e. $\ell\in g$, the weights sum to one $\sum_{\ell\in g}\sum_e\omega^{g}_{e,\ell}=1$.
2. For relative times belonging to some other bin i.e. $\ell\in g'$ and $g'\neq g$, the weights sum to zero $\sum_{\ell\in g'}\sum_e\omega^{g}_{e,\ell}=0$ for each bin g'.
3. For relative times not contained in $\mathcal{G}$ i.e. $\ell\in g^{excl}$, the weights sum to negative one $\sum_{\ell\in g^{excl}}\sum_e\omega^{g}_{e,\ell}=-1$.
4. If there are never-treated units i.e. $\infty\in supp(E_i)$, we have $\omega^{g}_{\infty,\ell}=0$ for all g and ℓ.

To see that 1) $\sum_{\ell\in g}\sum_e \omega^g_{e,\ell}=1$, note that the sum of weights is equal to

$$\sum_{t=0}^{T}\sum_{\ell\in g}\sum_{e}\omega^g_{e,\ell}=\sum_{t=0}^{T}\sum_{\ell\in g}\sum_{e}\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}E[\ddot{\mathbf{D}}_{i,t}D^\ell_{i,t}\cdot\mathbf{1}\{E_i=e\}] \tag{55}$$

$$=\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\sum_{\ell\in g}D^\ell_{i,t}\cdot\sum_{e}\mathbf{1}\{E_i=e\}] \tag{56}$$

$$=\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{1}\{t-E_i\in g\}]. \tag{57}$$

It is thus the population regression coefficient on $\mathbf{1}\{t-E_i\in g\}$ from regressing $\mathbf{1}\{t-E_i\in g\}$ on $\mathbf{D}_{i,t}$ and the unit and time fixed effects, which is just one. Similarly, for each $g'\neq g$, the sum of weights is the population regression coefficient on $\mathbf{1}\{t-E_i\in g'\}$ from regressing $\mathbf{1}\{t-E_i\in g\}$ on $\mathbf{D}_{i,t}$ and the unit and time fixed effects, which is zero. To see that the weights from all excluded relative time add up to negative one across cohorts, note that the sum of these weights is equal to

$$\sum_{\ell\in g^{excl}}\sum_{e}\omega^g_{e,\ell}=\mathbf{e}_g^\intercal\sum_{\ell\in g^{excl}}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}D^\ell_{i,t}] \tag{58}$$

$$=\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E\left[\ddot{\mathbf{D}}_{i,t}\sum_{\ell\in g^{excl}}D^\ell_{i,t}\right] \tag{59}$$

$$=\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}(1-\sum_{g\in\mathcal{G}}\sum_{\ell\in g}D^\ell_{i,t})] \tag{60}$$

$$=-\mathbf{e}_g^\intercal\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\sum_{g\in\mathcal{G}}\sum_{\ell\in g}D^\ell_{i,t}] \tag{61}$$

$$=-\mathbf{e}_g^\intercal\sum_{g\in\mathcal{G}}\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{1}\{t-E_i\in g\}]=-1 \tag{62}$$

where the third equality follows from $\sum_{\ell\in g^{excl}}D^\ell_{i,t}+\sum_{g\in\mathcal{G}}\sum_{\ell\in g}D^\ell_{i,t}=\sum_{-T\leq\ell\leq T}D^\ell_{i,t}=1$. The fourth equality follows from $\sum_t E[\ddot{\mathbf{D}}_{i,t}]=0$ due to the zero mean property of double-demeaning proved in Lemma 1. The last line simplifies to negative one because each term in the summation

$$\left(\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{D}_{i,t}^\intercal]\right)^{-1}\sum_{t=0}^{T}E[\ddot{\mathbf{D}}_{i,t}\mathbf{1}\{t-E_i\in g\}] \tag{63}$$

equals the population regression coefficient on $\mathbf{D}_{i,t}$ from regressing $\mathbf{1}\{t-E_i\in g\}$ on $\mathbf{D}_{i,t}$ and the unit and time fixed effects, which is a column vector with one in the entry corresponding to $\mathbf{1}\{t-E_i\in g\}$. Summing over g, the above expression is equal to a column vector of ones.

Finally, we note that the weight for the never-treated cohort is always zero $\omega^g_{\infty,\ell}=0$. This is because $D^\ell_{i,t}=0$ for all ℓ when $E_i=\infty$. □

Proof of Proposition 2

Proof. Under the parallel trends assumption, we can replace each term in Proposition 1 with

$$E[Y_{i,e+\ell} - Y^{\infty}_{i,0}|E_i] - E[Y^{\infty}_{i,e+\ell} - Y^{\infty}_{i,0}] = CATT_{e,\ell} + \underbrace{E[Y^{\infty}_{i,t} - Y^{\infty}_{i,0}|E_i] - E[Y^{\infty}_{i,t} - Y^{\infty}_{i,0}]}_{=0} \tag{64}$$

for $t = e + \ell$. □

Proof of Proposition 3

Proof. Under the no anticipation assumption, setting the corresponding terms to zero in the linear combination underlying $\mu_{\underline{g}}$ displayed in Proposition 2. □

Proof of Proposition 4

Proof. Under the assumption of homogeneous treatment effect, we have $CATT_{e,\ell} = ATT_\ell$ for all e. This simplifies Proposition 2 to

$$\sum_{\ell \in g}\left(\sum_{e}\omega^{g}_{e,\ell}\right)ATT_\ell + \sum_{g' \neq g}\sum_{\ell \in g'}\left(\sum_{e}\omega^{g}_{e,\ell}\right)ATT_\ell + \sum_{\ell \in g^{excl}}\left(\sum_{e}\omega^{g}_{e,\ell}\right)ATT_\ell. \tag{65}$$

Recall the weight $\omega^{g}_{e,\ell}$ is equal to the population regression coefficient on $\mathbf{1}\{t - E_i \in g\}$ from regressing $D^{\ell}_{i,t} \cdot \mathbf{1}\{E_i = e\}$ on all the bin indicators, i.e. $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ and two-way fixed effects included in (4). Summing over e would imply the weight $\omega^{g}_{\ell} := \sum_e \omega^{g}_{e,\ell}$ is equal to the population regression coefficient on $\mathbf{1}\{t - E_i \in g\}$ from regressing $D^{\ell}_{i,t}$ on all the bin indicators, i.e. $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ and two-way fixed effects included in (4). □

C Supplementary results related to IW estimator

In this section, we first provide detailed definition for the IW estimator, and then state its asymptotic behavior. We then prove its asymptotic normality and derive its asymptotic variance in Section C.1.

Definition 4. The IW estimator for ν_g, a weighted average of $CATT_{e,\ell \in g}$ is constructed via the following three steps. We focus on the setting without never-treated units, using pre-period $s = -1$ and the last-treated cohort as the control cohort $C = \{\max\{E_i\}\}$. The setting with never-treated units can be defined similarly by modifying the interacted specification.

Step 1 Estimate the $CATT_{e,\ell}$ using the interacted specification.

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \neq \max\{E_i\}} \sum_{\ell \neq -1} \delta_{e,\ell}(\mathbf{1}\{E_i = e\} \cdot D_{i,t}^{\ell}) + \epsilon_{i,t} \tag{66}$$

$$= \alpha_i + \lambda_t + \mathbf{B}_{i,t}^{\intercal} \boldsymbol{\delta} + \epsilon_{i,t} \tag{67}$$

on observations from $t = 0, \ldots, \max\{E_i\} - 1$ and $E_i \neq 0$. We need to drop time period beyond $\max\{E_i\}$ because DID estimators for $CATT_{e,\ell}$ for $e + \ell \geq \max\{E_i\}$do not exist as explained in the main text. We need to exclude cohort 0 from estimation because DID estimators for $CATT_{0,\ell}$ do not exist as explained in the main text. Note that among regressors we exclude interactions with $\mathbf{1}\{E_i = \max\{E_i\}\}$ and we exclude interactions with $D_{i,t}^{-1}$. Here $\mathbf{B}_{i,t}$ is a column vector collecting the interactions $\mathbf{1}\{E_i = e\} \cdot D_{i,t}^{\ell}$. Similarly, $\boldsymbol{\delta}$ is a column vector collecting the coefficients $\delta_{e,\ell}$ on $\mathbf{1}\{E_i = e\} \cdot D_{i,t}^{\ell}$. The matrix notation is used later to derive the asymptotic variance of IW estimators.

Step 2 Estimate the weights, which are cohort shares among cohorts that experience at least ℓ periods of treatment relative to the initial treatment.

Denote by $N_e \coloneqq \sum_{i=1}^{N} \mathbf{1}\{E_i = e\}$ the number of units in cohort e. Denote by $h^{\ell} = \{e : 1 - \ell \leq e \leq \max\{E_i\} - 1 - \ell\}$ to be the set of cohorts that experience at least ℓ periods of treatment relative to the initial treatment. Below $vec(A)$ vectorizes matrix A by stacking its columns.

Define $\widehat{\mathbf{f}}^{\ell}$ to be a matrix with its $(t,e)^{th}$ entry equal to $\mathbf{1}\{t - e = \ell\} \cdot N_e / \sum_{e \in h^{\ell}} N_e$. Here $\mathbf{1}\{t - e = \ell\}$ indicates when cohort e experiences exactly ℓ periods of treatment and $N_e / \sum_{e \in h^{\ell}} N_e$ is equal to the sample share of units in cohort e among units that experience at least l periods of treatment. Denote by $\mathbf{f}^{\ell}$ the probability limit of $\widehat{\mathbf{f}}^{\ell}$, which is a matrix with its $(t,e)^{th}$ entry equal to $\mathbf{1}\{t - e = \ell\} \cdot Pr\{E_i = t - l \mid E_i \in h^{\ell}\}$. For example, with $E_i \in \{0, 1, 2, 3\}$ for $T = 3$ and $\ell = 0$, we have $h^0 = \{1, 2\}$ and thus

$$\widehat{\mathbf{f}}^0 = \begin{pmatrix} \frac{N_1}{N_1 + N_2} & 0 \\ 0 & \frac{N_2}{N_1 + N_2} \end{pmatrix} \tag{68}$$

and its probability limit is

$$\mathbf{f}^0 = \begin{pmatrix} Pr\{E_i = 1 | 1 \leq E_i \leq 2\} & 0 \\ 0 & Pr\{E_i = 2 | 1 \leq E_i \leq 2\} \end{pmatrix}. \tag{69}$$

In proof below, we show that the weight matrix estimator $\widehat{\mathbf{f}}^{\ell}$ is asymptotically normal $\sqrt{N}(vec(\widehat{\mathbf{f}}^{\ell}) - vec(\mathbf{f}^{\ell})) \rightarrow_d N(0, \Sigma_{f^{\ell}})$.

Step 3 Compute the IW estimator as the weighted sum of $\widehat{\delta}_{e,\ell}$ (estimated in Step 1) using weights (estimated in Step 2).

To form an estimator alternative to the dynamic FE estimator $\widehat{\mu}_g$ from Regression (4), we can use

$$\widehat{\nu}_g := \frac{1}{|g|}\sum_{\ell\in g}\sum_{e\in h^\ell}\frac{N_e}{\sum_{e\in h^\ell}N_e}\widehat{\delta}_{e,\ell} = \frac{1}{|g|}\sum_{\ell\in g} vec(\widehat{\mathbf{f}}^\ell)^\intercal\widehat{\boldsymbol{\delta}}. \tag{70}$$

With a few standard assumptions (which we present together below as Assumption 4) on Regression (26), we can show that the IW estimators are asymptotically normal.

Assumption 4. (The saturated regression assumptions).

1. *There are observations from at least two cohorts that are not treated in $t = 0$.*
2. *Independent, identically distributed cross-sectional observations: $\{(E_i,\mathbf{Y}_i) : i = 1,2,\ldots,N\}$ are i.i.d. draws from their joint distribution where $\mathbf{Y}_i = (Y_{i,0},\ldots,Y_{i,T})^\intercal$ is a $T\times 1$ vector.*
3. *Large outliers are unlikely: $(\mathbf{B}_{i,t},\epsilon_{i,t})$ have nonzero finite fourth moments.*
4. *Denote by $\ddot{\mathbf{B}}$ the data matrix, whose rows consist of $\ddot{\mathbf{B}}_{i,t}^\intercal$, double-demeaned version of $\mathbf{B}_{i,t}^\intercal$. Assume $\ddot{\mathbf{B}}$ has full rank. If $\ddot{\mathbf{B}}$ is reduced-rank because cohort e is empty, then discard regressors involving $\mathbf{1}\{E_i = e\}$.*

Denote by $\boldsymbol{\delta}$ the probability limit of $\widehat{\boldsymbol{\delta}}$, which is a vector of $CATT_{e,\ell}$. We next state the asymptotic distribution of the IW estimators specifically for ν_ℓ, the average effect at relative time ℓ. Results for the more general case of ν_g, which is the average effects across relative times $\ell\in g$, can be derived similarly. Note that we use a clustered variance-covariance structure to allow the possibility that $Y_{i,t}$ are dependent across t due to serial correlation.

Proposition 6. (Consistency and asymptotic normality of the *IW* estimators for ν_ℓ). *Under the assumptions of Proposition 5 and Assumption 4, the IW estimator converges in probability to*

$$\widehat{\nu}_\ell \to_p \sum_{e\in h^\ell} Pr\left\{E_i = e \mid E_i\in h^\ell\right\} CATT_{e,\ell} = vec(\mathbf{f}^\ell)^\intercal\boldsymbol{\delta}. \tag{71}$$

The asymptotic distribution of this estimator is

$$\sqrt{N}\left(\widehat{\nu}_\ell - vec\left(\mathbf{f}^\ell\right)^\intercal\boldsymbol{\delta}\right)\to_d N\left(0,\boldsymbol{\delta}^\intercal\Sigma_{f^\ell}\boldsymbol{\delta}+\Sigma_\ell\right) \tag{72}$$

for Σ_{f^ℓ} the asymptotic variance of $\sqrt{N}(vec(\widehat{\mathbf{f}}^\ell) - vec(\mathbf{f}^\ell))$ where $\widehat{\mathbf{f}}^\ell$ is the weight matrix estimator and

$$\boldsymbol{V}_{\ddot{\mathbf{B}}} = \sum_{t=0}^{\max\{E_i\}-1} E[\ddot{\mathbf{B}}_{i,t}\ddot{\mathbf{B}}_{i,t}^\intercal] \quad \Sigma_\ell = vec(\mathbf{f}^\ell)^\intercal\boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1}Var\left(\sum_{t=0}^{\max\{E_i\}-1}\ddot{\mathbf{B}}_{i,t}\ddot{\epsilon}_{i,t}\right)\boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1}vec(\mathbf{f}^\ell). \tag{73}$$

C.1 Proof of the validity of the IW estimator

Proof of Proposition 5

Proof. Provided that the DID estimator is well-defined with pre-period s and control cohorts C, we first show the DID estimator is an unbiased and consistent estimator for $E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+\ell} - Y_{i,s} \mid E_i \in C]$. We prove the first term in the DID estimator $\frac{\mathbb{E}_N[(Y_{i,e+\ell}-Y_{i,s})\cdot\mathbf{1}\{E_i=e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}$ is unbiased and consistent for $E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e$. The argument for the second term in the DID estimator follows similarly.

For unbiasedness, note that by the Law of Iterated Expectations and linearity of $\mathbb{E}_N$, we have

$$E[\frac{\mathbb{E}_N[(Y_{i,e+\ell}-Y_{i,s})\cdot\mathbf{1}\{E_i=e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}] = E\Big[E\Big[\frac{\mathbb{E}_N[(Y_{i,e+\ell}-Y_{i,s})\cdot\mathbf{1}\{E_i=e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]} \mid E_i\Big]\Big] \tag{74}$$

$$= E\Bigg[\frac{\mathbb{E}_N\Big[E[(Y_{i,e+\ell}-Y_{i,s})\cdot\mathbf{1}\{E_i=e\} \mid E_i]\Big]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}\Bigg] \tag{75}$$

$$= E\Bigg[\frac{\mathbb{E}_N\Big[E[Y_{i,e+\ell}-Y_{i,s} \mid E_i=e]\cdot\mathbf{1}\{E_i=e\}\Big]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}\Bigg] \tag{76}$$

$$= E[\frac{E[(Y_{i,e+\ell}-Y_{i,s}) \mid E_i=e]\cdot\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i=e\}]}] \tag{77}$$

$$= E[(Y_{i,e+\ell}-Y_{i,s}) \mid E_i=e]. \tag{78}$$

For consistency, by the Law of Large Numbers the numerator and the denominator converge in probability to $E[(Y_{i,e+\ell} - Y_{i,s})\cdot\mathbf{1}\{E_i = e\}]$ and $Pr\{E_i = e\}$ respectively. By the Law of Iterated Expectations and Slutsky's theorem, it converges in probability to $E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e]$.

To show that the DID estimator is an unbiased and consistent estimator for $CATT_{e,\ell}$, it remains to show $E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+\ell} - Y_{i,s} \mid E_i \in C] = CATT_{e,\ell}$.

Since $s < e$ and $c > e+\ell$, we have

$$E[Y_{i,e+\ell}-Y_{i,s} \mid E_i=e] - E[Y_{i,e+\ell}-Y_{i,s} \mid E_i \in C] \tag{79}$$

$$=E[Y^e_{i,e+\ell}-Y^e_{i,s} \mid E_i=e] - \sum_{c\in C} Pr\{E_i=c \mid E_i\in C\}\, E[Y^c_{i,e+\ell}-Y^c_{i,s} \mid E_i=c] \tag{80}$$

$$=E[Y^e_{i,e+\ell}-Y^\infty_{i,s} \mid E_i=e] - \sum_{c\in C} Pr\{E_i=c \mid E_i\in C\}\, E[Y^\infty_{i,e+\ell}-Y^\infty_{i,s} \mid E_i=c] \tag{81}$$

$$=E[Y^e_{i,e+\ell}-Y^\infty_{i,e+\ell} \mid E_i=e] + E[Y^\infty_{i,e+\ell}-Y^\infty_{i,s} \mid E_i=e] \tag{82}$$

$$-\sum_{c\in C} Pr\{E_i=c \mid E_i\in C\}\, E[Y^\infty_{i,e+\ell}-Y^\infty_{i,s} \mid E_i=c] \tag{83}$$

$$=E[Y^e_{i,e+\ell}-Y^\infty_{i,e+\ell}|E_i=e] + E[Y^\infty_{i,e+\ell}-Y^\infty_{i,s}] - E[Y^\infty_{i,e+\ell}-Y^\infty_{i,s}] \tag{84}$$

$$=E[Y^e_{i,e+\ell}-Y^\infty_{i,e+\ell}|E_i=e] \tag{85}$$

where the second equality follows from Assumption 2 and the fourth equality follows from Assumption 1. □

Proof of Proposition 6

Proof. We first show the consistency and asymptotic normality of the weight estimators. The weight estimators $\widehat{\mathbf{f}}^{\ell}$ are consistent since $\frac{N_e}{\sum_{e \in h^{\ell}} N_e} \to_p \frac{Pr\{E_i = e\}}{Pr\{E_i \in h^{\ell}\}} = Pr\left\{E_i = e \mid E_i \in h^{\ell}\right\}$ by the Law of Large Numbers and Slutsky's theorem. Note that $\frac{N_e}{\sum_{e \in h^{\ell}} N_e}$ is also the regression coefficient estimator from the following cross-sectional regression

$$\mathbf{1}\{E_i = e\} = \beta \mathbf{1}\left\{E_i \in h^{\ell}\right\} + \eta_i(e) \tag{86}$$

with population regression coefficient equal to $\beta = Pr\left\{E_i = e \mid E_i \in h^{\ell}\right\}$. Then by OLS asymptotics which holds as E_i are iid by assumption and $\eta_i(e)$ is bounded, we have

$$\sqrt{N}\left(\frac{N_e}{\sum_{e \in h^{\ell}} N_e} - Pr\left\{E_i = e \mid E_i \in h^{\ell}\right\}\right) \to_d N\left(0, \frac{E[\mathbf{1}\left\{E_i \in h^{\ell}\right\}^2 \eta_i^2(e)]}{E[\mathbf{1}\{E_i \in h^{\ell}\}^2]^2}\right). \tag{87}$$

Note that $\mathbf{1}\left\{E_i \in h^{\ell}\right\}^2 = \mathbf{1}\left\{E_i \in h^{\ell}\right\}$ so the asymptotic variance is equal to

$$\frac{E[\eta_i^2(e) \mid E_i \in h^{\ell}] Pr\left\{E_i \in h^{\ell}\right\}}{Pr\{E_i \in h^{\ell}\}^2} = \frac{E[\eta_i^2(e) \mid E_i \in h^{\ell}]}{Pr\{E_i \in h^{\ell}\}}. \tag{88}$$

Similarly, for a pair of cohorts with $e \neq e'$, $\frac{N_e}{\sum_{e \in h^{\ell}} N_e}$ and $\frac{N_{e'}}{\sum_{e \in h^{\ell}} N_e}$ are asymptotically correlated with covariance $E[\eta_i(e)\eta_i(e') \mid E_i \in h^{\ell}]/Pr\left\{E_i \in h^{\ell}\right\}$. Thus, $vec\left(\widehat{\mathbf{f}}^{\ell}\right)$ has asymptotic distribution

$$\sqrt{N}\left(vec\left(\widehat{\mathbf{f}}^{\ell}\right) - vec\left(\mathbf{f}^{\ell}\right)\right) \to_d N\left(0, \Sigma_{f^{\ell}}\right). \tag{89}$$

Here $\Sigma_{f^{\ell}}$ is a matrix with diagonal entries equal to $\frac{E[\eta_i^2(e) \mid E_i \in h^{\ell}]}{Pr\{E_i \in h^{\ell}\}}$, and off-diagonal entries equal to $\frac{E[\eta_i(e)\eta_i(e') \mid E_i \in h^{\ell}]}{Pr\{E_i \in h^{\ell}\}}$.

By consistency of the weight estimators and Proposition 5, we prove the consistency of the IW estimator, the first part of Proposition 6.

We next show the asymptotic normality of the $\widehat{\delta}_{e,\ell}$. The standard OLS asymptotics applies because by assumption after double demeaning, the data $(\ddot{\mathbf{B}}_{i,t}, \ddot{\epsilon}_{i,t})$ is iid across i and has nonzero finite fourth moments. The asymptotic distribution of this estimator is thus

$$\sqrt{N}\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \to_d N\left(0, \boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1} Var\left(\sum_{t=0}^{T-1} \ddot{\mathbf{B}}_{i,t} \ddot{\epsilon}_{i,t}\right) \boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1}\right) \tag{90}$$

where $\boldsymbol{V}_{\ddot{\mathbf{B}}} = \sum_{t=0}^{T-1} E[\ddot{\mathbf{B}}_{i,t} \ddot{\mathbf{B}}_{i,t}^{\intercal}]$.

Lastly, by the delta method, we have

$$\sqrt{N}\left(vec\left(\widehat{\mathbf{f}}^{\ell}\right)^{\intercal} \widehat{\boldsymbol{\delta}} - vec\left(\mathbf{f}^{\ell}\right)^{\intercal} \boldsymbol{\delta}\right) \to_d N\left(0, \boldsymbol{\delta}^{\intercal} \Sigma_{f^{\ell}} \boldsymbol{\delta} + \Sigma_{\ell}\right) \tag{91}$$

where $\Sigma_\ell = vec(\mathbf{f}^\ell)^\intercal \boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1} Var\left(\sum_{t=0}^{T-1} \ddot{\mathbf{B}}_{i,t}\ddot{\epsilon}_{i,t}^2\ddot{\mathbf{B}}_{i,t}^\intercal\right)\boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1} vec(\mathbf{f}^\ell)$. This follows because $vec\left(\widehat{\mathbf{f}^\ell}\right)$ and $\widehat{\boldsymbol{\delta}}$ are uncorrelated: the asymptotic covariance between $\frac{N_e}{\sum_{e\in h^\ell} N_e}$ and $\widehat{\boldsymbol{\delta}}$ is equal to

$$\frac{\boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1} Cov(\mathbf{1}\left\{E_i \in h^\ell\right\}\eta_i(e), \ddot{\mathbf{B}}_{i,t}\ddot{\epsilon}_{i,t})}{E[\mathbf{1}\{E_i \in h^\ell\}^2]^2} = \frac{\boldsymbol{V}_{\ddot{\mathbf{B}}}^{-1} E[\ddot{\mathbf{B}}\eta_i(e)\ddot{\epsilon}_{i,t} \mid E_i \in h^\ell]}{E[\mathbf{1}\{E_i \in h^\ell\}^2]^2}. \tag{92}$$

Since $\ddot{\mathbf{B}}$ and $\eta_i(e)$ are functions of E_i, we have $E[\ddot{\mathbf{B}}\eta_i(e)\ddot{\epsilon}_{i,t} \mid E_i \in h^\ell] = E[\ddot{\mathbf{B}}\eta_i(e)E[\ddot{\epsilon}_{i,t} \mid E_i, E_i \in h^\ell]]$. Furthermore, specification (26) is saturated in E_i and relative time so $E[\ddot{\epsilon}_{i,t} \mid E_i] = 0$. This proves the asymptotic asymptotic normality of the IW estimators, the second part of Proposition 6. □