

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

Liyang Sun[†] and Sarah Abraham[‡]

February 14, 2020

Abstract

To estimate the dynamic effect 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 effect heterogeneity. We propose an alternative estimator that is free of contamination, and illustrate the shortcomings of two-way fixed effects regression with leads and lags in comparison to our proposed estimator 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, Helen Willis and Jeffrey Wong for helpful discussions. We are grateful to the a editor and four anonymous referees for valuable and constructive comments. This research was supported by the National Institute on Aging, Grant Number T32-AG000186. This draft is a work in progress and comments are welcome; all errors are our own. A preliminary draft of this paper was circulated on April 16, 2018. Replication code is available at <http://economics.mit.edu/grad/lsun20/>.

[†]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 (2018), de Chaisemartin and D’Haultfœuille (2019) and Goodman-Bacon (2018) interpret the coefficient on the treatment status when there is treatment effect 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. These relative period coefficients μ_ℓ are still not well-understood in the presence of treatment effect heterogeneity and variation in treatment timing. This paper builds on the existing literature in two ways. First, we demonstrate that with treatment effect heterogeneity and variation in treatment timing, the relative period coefficients μ_ℓ from two-way fixed effects regressions may be contaminated by effects from periods other than ℓ . Second, we improve two-way fixed effects regressions by proposing a simple regression-based method that estimates an interpretable analog of the relative period coefficient μ_ℓ , namely an average of treatment effects from ℓ periods relative to the treatment.

Two-way fixed effects regression estimates are loosely referred to as *event study* estimates 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 We learn what specifications empirical researchers are actually using when estimating two-way fixed effects regressions from this sample of applied papers. 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

¹We 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.

sample consider the more complex setting with variation in treatment timing 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 for the former setting, but also the latter setting where units receive their first treatments at various times, which is our focus.² These papers all report some estimates for relative period coefficients μ_ℓ from two-way fixed effects regression, which fall under our analysis.

We begin our analysis by setting up a potential outcome framework. The potential outcome framework we adopt is a special case of the dynamic potential outcome framework used by Robins (1986, 1997). We focus on an *absorbing* treatment such that the treatment status over time can be uniquely characterized by the time period of the initial treatment E_i . We consider a particular counterfactual outcome $Y_{i,t}^\infty$, the potential outcome had the unit never received the treatment, as opposed to all possible counterfactual outcomes had the unit first received the treatment at some time other than observed time E_i . This allows us to define the cohort-specific average treatment effects (where units in the same cohort e share the same $E_i = e$), which we call $CATT_{e,\ell}$.

We use these cohort-specific average treatment effects $CATT_{e,\ell}$ as building blocks to decompose the relative periods coefficients μ_ℓ from two-way fixed effects regression. While each μ_ℓ is meant to capture an average of treatment effects $CATT_{e,\ell}$ from its own relative period ℓ , under the familiar parallel trends assumption only, we find that is not the case in the presence of treatment effect heterogeneity. The parallel trends assumption alone allows us to write μ_ℓ as a linear combination of $CATT_{e,\ell}$. The weights in this linear combination, derived from our decomposition result, can be non-convex. They can also be non-zero for relative periods not from μ_ℓ 's own relative period ℓ , so that in this linear combination underlying μ_ℓ , $CATT_{e,\ell'}$ may receive non-zero weight even for $\ell' \neq \ell$. These weights are unlikely what researchers have in mind with two-way fixed effects regression, as the coefficient on a given lead or lag can be contaminated by effects from other periods.

²This 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.

This result has practical consequences: under treatment effect heterogeneity, we can have a spurious non-zero positive lead coefficient even when there is no pretrend. In his survey of the applied literature, Roth (2019) notes that it is common to test for pretrends, specifically checking whether $\mu_\ell = 0$ for ℓ leads of treatment. Our decomposition result implies such test would be invalid as under the null hypothesis of no pretrend, we can still have $\mu_\ell \neq 0$ due to treatment effect heterogeneity.

The extent to which the relative period coefficient μ_ℓ can be contaminated by treatment effects from other relative periods ℓ' depends on the magnitude of the weights on $CATT_{e,\ell'}$ in the linear combination decomposition we derive for μ_ℓ . The weights are functions of the cohort composition, and we illustrate via numerical examples how we can easily calculate the weights using the cohort composition alone for any given application of two-way fixed effects regression. Examining the weights lets researchers gauge how treatment effect heterogeneity would interact with potential non-convex and non-zero weighting in μ_ℓ .

In the second part of this paper, we propose a simple regression-based solution to recover an easy-to-interpret causal parameter that can accommodate treatment effect heterogeneity. Our alternative method flexibly estimates the dynamic treatment effect for each cohort, and then calculates the average of these cohort-specific estimates, with weights representative of the cohort share. We derive a consistent variance estimator for our alternative estimator, with which researchers can easily construct a pointwise confidence interval for the average dynamic treatment effects.

In practice two-way fixed effects regressions are almost always estimated with covariates. We note that our negative decomposition result holds with or without covariates, and briefly discuss how our alternative method accommodates covariates in Section 4.2.

We illustrate both our decomposition results and our alternative method via an empirical application. The publicly-available dataset, Health and Retirement Study (HRS), includes variations in the timing of hospitalization. Dobkin et al. (2018) use two-way fixed-effects regressions on this dataset to estimate the effect of a hospitalization. We illustrate that the cohort composition in this dataset would result in non-convex and non-zero weighting in μ_ℓ , so that μ_ℓ can include treatment

effects from its own relative period as well as other relative periods.

We then illustrate our alternative method with this example. Among the outcomes Dobkin et al. (2018) study, we focus on out-of-pocket medical spending and labor earnings. For these two outcomes, our alternative method yields similar big-picture findings as the original paper: the earnings decline due to hospitalization is substantial compared to the transitory out-of-pocket spending increase. 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 convex weighted averages of the underlying effects.

Related Literature. Our paper is related to the traditional analysis of nonseparable panel and treatment effects models (e.g. Heckman et al. (1998, 1997); Blundell et al. (2004); Abadie (2005); Chernozhukov et al. (2013)), but more closely related to the recent literature on the interpretation of two-way fixed-effects regressions when the timing of treatment varies.

The recent literature on two-way fixed effects regressions has a similar structure to the first part of our paper. We all first provide decomposition results for coefficients from some specification of two-way fixed effects regression, and then derive assumptions for the regression coefficients to be causally interpretable.

These papers focus on a few selected specifications of two-way fixed effects regression that differ on the included regressors other than the two-way fixed effects. In contrast, our decomposition results apply to a general class of specifications. Athey and Imbens (2018), Goodman-Bacon (2018), and a special case (staggered adoption design) of de Chaisemartin and D’Haultfœuille (2019) focus on a “static” specification where the regressor is an indicator for having received treatment. Borusyak and Jaravel (2017) discuss several other specifications: “fully-dynamic” where the each regressor is an indicator for a single relative period $\{\ell\}_{\ell=-T}^T$, “semi-dynamic” where the regressors are single relative period indicators, but only post-treatment $\{\ell\}_{\ell=0}^T$, and “capped” where post-treatment periods beyond period B are collected in one indicator. However, they only illustrate non-convex weighting from the “capped” specification for a specific cohort composition.

While our derivations all rely on the double-demeaning properties of the two-way fixed effects, we consider a flexible specification of these regressors, allowing indicators for both single relative periods ℓ as well as a set g of relative periods. Our specification thus encompasses the aforementioned specifications, as well as other specifications that can be expressed them with the appropriate choice of relative period indicators.

Our causal interpretation of two-way fixed-effects regression is also closely related to the decomposition results in this literature. The main difference is due to varying definitions of treatment effects, which results in different building blocks for the interpretation. We interpret the regression coefficient as a linear combination of $CATT_{e,\ell}$, which is the cohort-specific average change in outcome relative to never being treated. Athey and Imbens (2018) define the treatment effect as the cohort-specific average change in outcome due to change in the timing of initial treatment from e to e' . This is different from our building block, $CATT_{e,\ell}$, which is the cohort-specific average change in outcome relative to never being treated. Borusyak and Jaravel (2017) assume the “fully-dynamic” specification as their DGP, implicitly assuming $CATT_{e,\ell}$ is constant across e and use that as their building block. So our negative result on the “dynamic” two-way fixed effects regression nests Borusyak and Jaravel (2017). de Chaisemartin and D’Haultfœuille (2019) allow the presence of “leavers” (units switching from being treated to untreated). So their setup is more complicated as the same cohort can experience different evolutions of treatment, constituting groups. Their building block is thus the group-time average difference in outcome with and without treatment. Goodman-Bacon (2018) uses an average of $CATT_{e,\ell}$ over some relative period range as his building block.

To reach any causal interpretation, we all use some potential outcomes framework along with some identifying assumptions. In Section 2 of the paper, for each of the identifying assumptions we propose, we clarify how they relate to the assumptions made in these papers.

Unlike previous work, we provide estimators for corrected analogs to the coefficients from two-way fixed effects regressions, offering a simple regression-based alternative method that estimates what μ_g ’s are supposed to estimate, namely an average of treatment effects from its own relative

periods. This alternative estimator yields μ_g 's, free of contamination from relative periods other than those included in g , which estimate a convex average of the underlying treatment effect. Athey and Imbens (2018) study design-based inference for the estimand of the “static” specification using the random adoption date assumption. Callaway and Sant’Anna (2018) propose inverse propensity score weighted estimator for “group-time average treatment effect”, which states $CATT_{e,\ell}$ in calendar time but otherwise identical. Specifically, they use the notation $ATT(g, t)$ where g corresponds to our cohort and t denotes calendar time, which can be translated using the relationship $t = e + \ell$. Then they discuss a number of different ways to aggregate them into parameters of interest. de Chaisemartin and D’Haultfœuille (2019) propose robustness measures that assess how much non-convex weighting can affect the coefficient from the “static” specification and an alternative estimator for its corrected analog. Goodman-Bacon (2018) identifies a weakened parallel trends assumption for the original two-way fixed effects regressions and proposes an associated balance test.

Relevance in the Applied Literature. In addition to surveying the theoretical econometrics literature, we examine what specifications empirical researchers are actually using when estimating two-way fixed effects regressions. For that purpose, we summarize the main specifications in our sample of twelve applied papers in Table 1.³ These papers use a wide variety of specifications to study a broad range of 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. The columns of Table 1 collect key properties of these specifications. We later discuss the implications of these properties in the context of our decomposition results, and reference Table 1 for the empirical relevance of these properties.

In the next section, we formally introduce the event study design. Section 3 derives the esti-

³We 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.

mands 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.

2 Event studies design

We consider a setting with a random sample of N units observed over $T + 1$ time periods, where both T and N are fixed. Specifically, for each $i \in \{0, \dots, N\}$ and $t \in \{0, \dots, T\}$, 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 $\{Y_{i,t}, D_{i,t}\}_{t=0}^T$ independent and identically distributed (i.i.d.).

In the general case, at each time t there are two possible treatment status $D_{i,t} \in \{0, 1\}$, over $T + 1$ time periods, and the path of treatment status $\{D_{i,t}\}_{t=0}^T$ can take on 2^{T+1} possible values. Since the number of possible paths scales exponentially with T , we restrict the paths of interest to make the analysis manageable. Specifically, for 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 $\{i : E_i = e\}$.

As noted in the introduction, based on the empirical papers we survey an *event study* design nests a *difference-in-differences* design, where units are either first treated at time t_0 or never treated so that $E_i \in \{t_0, \infty\}$. An *event study* design is a *staggered adoption* design when E_i takes on more than one values from $\{0, \dots, T\}$, and there may or may not be never treated units with $E_i = \infty$.

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

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

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, we would have a new treatment being absorbing by construction. The effect of having ever received the treatment is oftentimes of interest, as it captures the path of treatment effect, even though the treatment itself may be transient. For example, Deryugina (2017) is interested the fiscal cost for a county to be hit with a hurricane. While a hurricane itself may be transient, we can still study the impact of having had any hurricane. This is exactly the approach taken by Deryugina (2017), where she 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 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 *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 considers the treatment effect relative to the always-treated counterfactual outcome: $Y_{i,t} - Y_{i,t}^0$. Sianesi (2004) defines the treatment effect to be relative to the not-yet-treated counterfactual outcome: $Y_{i,t} - Y_{i,t}^e$ for $e > t$.

We can take the average of treatment effect 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 define 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 effects on the treated (CATT) ℓ periods from initial treatment is

$$CATT_{e,\ell} = E[Y_{i,e+\ell} - Y_{i,e+\ell}^\infty \mid E_i = e]. \quad (2)$$

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 period allows us to compare across cohorts while holding their exposure to the treatment constant.

2.2 Identifying assumptions

With the above definitions, we formalize three identifying assumptions for our event study design. The first assumption is a generalized form of 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 these assumptions.

Assumption 1. (Parallel trends in baseline outcomes.) $E[Y_{i,t}^\infty - Y_{i,s}^\infty \mid E_i = e]$ is the same for all $e \in \text{supp}(E_i)$ and for all s, t and is equal to $E[Y_{i,t}^\infty - Y_{i,s}^\infty]$.

If an application includes never-treated units so that $\infty \in \text{supp}(E_i)$, we need to especially consider whether these never-treated units satisfy 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 parallel trends assumption, then we should exclude them from the estimation to avoid violation of this assumption.

Our parallel trends assumption coincides with that of de Chaisemartin and D’Haultfœuille (2019). One could substitute this assumption with a stronger identifying assumption that baseline outcomes are mean independent of E_i i.e. at each t , $E[Y_{i,t}^\infty | E_i = e]$ is the same for all $e \in \text{supp}(E_i)$ and in particular is equal to $E[Y_{i,t}^\infty]$. This assumption is plausible when 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 parallel trends assumption. Callaway and Sant’Anna (2018) 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 $E[Y_{i,t}^\infty - Y_{i,s}^\infty | E_i = e]$ across cohorts to be zero.

Assumption 2. (No anticipatory behavior in $\ell \in \underline{g}$ periods prior to treatment.) *There is a (non-empty) set \underline{g} of pre-treatment periods such that $E[Y_{i,e+\ell}^e - Y_{i,e+\ell}^\infty | E_i = e] = 0$ for all $e \in \text{supp}(E_i)$ and all $\ell \in \underline{g}$.*

Assumption 2 requires potential outcomes $\ell \in \underline{g}$ periods before treatment to be equal to the baseline outcome on average as in Botosaru and Gutierrez (2018). If Assumption 2 holds for all pre-periods i.e. $\underline{g} = [-T, -1]$, there are no pre-trends since $CATT_{e,\ell} = 0$ for all $e \in \text{supp}(E_i)$ and all $\ell < 0$. 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. Depending on the application, it may still be plausible to assume no anticipation up till K periods before the treatment so that $\underline{g} = [-T, -K]$.

The no anticipation assumption proposed by Athey and Imbens (2018) is a deterministic condition which stipulates that $Y_{i,e+\ell}^e = Y_{i,e+\ell}^\infty$ 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 in the DGP. Callaway and Sant’Anna (2018) and Goodman-Bacon (2018) implicitly assume no anticipation by taking 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 $CATT_\ell$.*

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 just need to differ across cohorts in one relative period.

Our notion of treatment effect homogeneity does not preclude dynamic treatment effects. It just imposes that cohorts share the same path of treatment effect. The related literature sometimes formulates restriction 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 effect heterogeneity as “only across the time horizon”, which implies $CATT_{e,\ell}$ can vary over ℓ but not over e . Callaway and Sant’Anna (2018) allow for “arbitrary treatment effect heterogeneity” when $CATT_{e,\ell}$ varies across cohorts and over time. Similarly, de Chaisemartin and D’Haultfoeuille (2019) describe treatment effects may be “heterogeneous across groups and over time periods”. Goodman-Bacon (2018) allow heterogeneous 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 effect heterogeneity with time-varying treatment. Since researchers are naturally interested in dynamic treatment effects by using a “dynamic” specification, we do not restrict the path of treatment effect, and use “heterogeneity” to describe variation across cohorts only.

2.2.1 Sources of treatment effect heterogeneity

Assumption 3 is violated when different cohorts 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 Estimators from linear two-way FE regression

The regression under consideration is two-way (unit and time) fixed effects (FE) regression of the following form, estimated on a panel of $i = 1, \dots, N$ units for $t = 0, 1, \dots, T$ calendar time periods:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \mu_g \mathbf{1}\{t - E_i \in g\} + v_{i,t} \quad (3)$$

Here $Y_{i,t}$ is the outcome of interest for unit i at time t , E_i is the initial time of a binary absorbing treatment for unit i , and α_i and λ_t are the unit and time fixed effects.

Elements of $g \in \mathcal{G}$ are disjoint sets of *relative period* periods. By appropriately specifying \mathcal{G} , we can express a large number of specifications encountered in practice as illustrated in Section 3.1. While relative periods have range $\ell \in [-T, T]$, we allow some to be excluded from the specification and denote the excluded set with $g^{excl} = \{\ell : \ell \notin \cup_{g \in \mathcal{G}} g\}$. We denote by μ_g the relative period coefficients from regression (3), i.e. the population regression coefficients. Their corresponding OLS estimators are denoted by $\hat{\mu}_g$ respectively. We are interested in the properties of μ_g when there are variations in the initial treatment timing, and there may or may not be never-treated units. Ideally μ_g should be a convex average of $CATT_{e,\ell}$ for periods $\ell \in g$ from its corresponding set g .

3.1 Common specifications

To clarify how the relative period indicator $\mathbf{1}\{t - E_i \in g\}$ in regression (3) varies with time, we define $D_{i,t}^\ell := \mathbf{1}\{t - E_i = \ell\}$ 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 $D_{i,t}^\ell = 0$ for all ℓ and all t . We can represent the relative period indicator as

$$\mathbf{1}\{t - E_i \in g\} = \sum_{\ell \in g} \mathbf{1}\{t - E_i = \ell\} = \sum_{\ell \in g} D_{i,t}^\ell. \quad (4)$$

This representation also illustrates possible specifications that can be expressed as regression (3). The flexible specification using \mathcal{G} means our results are applicable for a large number of specifications encountered in practice.

One type of specification is a “static” specification where \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

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_g \sum_{\ell \geq 0} D_{i,t}^\ell + v_{i,t}. \quad (5)$$

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

Another type of specification is a “dynamic” specification where \mathcal{G} contains singletons $\mathcal{G} = \{-K, \dots, \{-2\}, \{0\}, \{1\}, \dots, \{L\}\}$. The indicators $\mathbf{1}\{t - E_i \in g\}$ correspond to treatment leads and lags, excluding distant ones that are more than K periods before treatment and more than L periods after treatment, as well as the period before treatment, or period -1. The “dynamic” specification thus takes the following form

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell=-K}^{-2} \mu_\ell D_{i,t}^\ell + \sum_{\ell=0}^L \mu_\ell D_{i,t}^\ell + v_{i,t}. \quad (6)$$

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

Excluding some relative period from the “dynamic” specification is necessary to avoid multicollinearity, either among the relative period indicators $D_{i,t}^\ell$, or with the unit and time fixed effects. For example, when there are no never-treated units i.e. $\infty \notin \text{supp}(E_i)$ 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 multicollinearity comes from the relative period indicators summing up to one for every unit $\sum_{\ell \in [-T, T]} D_{i,t}^\ell = 1$, and the other multicollinearity comes from the linear relationship between two-way fixed effects and the relative period indicators, namely $t - E_i = \ell$.

Excluding relative period 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 $D_{i,t}^{-1}$. The remaining two papers exclude $D_{i,t}^0$.

Excluding distant relative periods is however less common (only one of the eight papers we survey do so). Instead researchers “bin” or “trim” distant relative periods. For “binning”, researchers bin distant relative periods into $\underline{g} = [-T, -K)$ and $\bar{g} = (L, T]$ and estimate a “binned” specification

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_{\underline{g}} \sum_{\ell < -K} D_{i,t}^\ell + \sum_{\ell = -K}^{-2} \mu_\ell D_{i,t}^\ell + \sum_{\ell = 0}^L \mu_\ell D_{i,t}^\ell + \mu_{\bar{g}} \sum_{\ell > L} D_{i,t}^\ell + v_{i,t}. \quad (7)$$

without excluding any relative period so that $g^{\text{excl}} = \emptyset$. For “trimming”, researchers trim their panel to be balanced in relative periods. We discuss the implications of “binning” and “trimming” in Section 3.5.1 and 3.2.2 respectively.

3.2 Interpreting the coefficients under no assumptions

We now interpret μ_g when each of the three identifying assumptions fail. First we show that without any assumptions, we can write μ_ℓ 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 $\ell \in g'$ of other*

bins $g' \neq g$, and from relative periods excluded from the specification $\ell \in g^{excl}$:

$$\mu_g = \sum_{\ell \in g} \sum_e \omega_{e,\ell}^g \left(E[Y_{i,e+\ell} - Y_{i,0}^\infty | E_i = e] - E[Y_{i,e+\ell}^\infty - Y_{i,0}^\infty] \right) \quad (8)$$

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

$$+ \sum_{\ell \in g^{excl}} \sum_e \omega_{e,\ell}^g \left(E[Y_{i,e+\ell} - Y_{i,0}^\infty | E_i = e] - E[Y_{i,e+\ell}^\infty - Y_{i,0}^\infty] \right). \quad (10)$$

We use the superscript g to associate the weight $\omega_{e,\ell}^g$ with the coefficient μ_g . The weight $\omega_{e,\ell}^g$ associated with cohort e in relative period ℓ 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 regression (3), and two-way fixed effects.

The above proposition is a direct result of the double-demeaning properties of the two-way fixed effects. We defer the derivation to the Appendix, but mention the following four properties of the weights $\omega_{e,\ell}^g$.

- For relative periods of μ_g 's own i.e. $\ell \in g$, their associated weights as displayed in (8) sum to one $\sum_{\ell \in g} \sum_e \omega_{e,\ell}^g = 1$.
- For relative periods belonging to some other bin i.e. $\ell \in g'$ and $g' \neq g$, their associated weights as displayed in (9) sum to zero $\sum_{\ell \in g'} \sum_e \omega_{e,\ell}^g = 0$ for each bin g' .
- For relative periods not included in \mathcal{G} , the weights associated weights as displayed in (10) sum to negative one $\sum_{\ell \in g^{excl}} \sum_e \omega_{e,\ell}^g = -1$.
- If there are never-treated units i.e. $\infty \in \text{supp}(E_i)$, we have $\omega_{\infty,\ell}^g = 0$ for all g and ℓ .

We can easily estimate the weights using the cohort composition alone for any given specification of \mathcal{G} , namely regressing $D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}$ on all bin indicators included in regression (3), $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$, and two-way fixed effects. We illustrate the above decomposition results via two numerical examples.

3.2.1 The implication of never-treated units for weighting

Our decomposition results hold under the presence of never-treated units, which would apply to five out of the eight papers we survey. We construct a panel balanced in calendar time with $T = 3$. In Table 2, we estimate the weights underlying $\mu_{\bar{g}}$ and $\mu_{\underline{g}}$ from the following “binned” specification with $\mathcal{G} = \{\underline{g}, \bar{g}\}$, where $\underline{g} = [-T, -2]$ and $\bar{g} = [0, T]$:

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_{\underline{g}} \mathbf{1}\{t - E_i < -1\} + \mu_{\bar{g}} \mathbf{1}\{t - E_i \geq 0\} + v_{i,t} \quad (11)$$

Take $\mu_{\bar{g}}$ for example, to estimate the weight on $CATT_{e,\ell}$ in the linear combination introduced in Proposition 1, we regress $D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}$ on $\mathbf{1}\{t - E_i < -1\}$, $\mathbf{1}\{t - E_i \geq 0\}$ and two-way fixed effects. The coefficient on $\mathbf{1}\{t - E_i \geq 0\}$ is the weight $\omega_{e,\ell}^{\bar{g}}$. We illustrate the four properties listed below Proposition 1 using these weights, displayed in the second columns of Table 2 Panel (a) where we divide the units evenly into cohorts $E_i \in \{1, 2, 3\}$, and Panel (b) where we divide the units evenly into cohorts $E_i \in \{1, 2, 3, \infty\}$.

- For relative periods included in bin \bar{g} namely $\ell \in \{0, 1, 2\}$, the weights $\omega_{e,\ell}^{\bar{g}}$ sum to one;
- For relative periods included in bin \underline{g} namely $\ell \in \{-3, -2\}$, the weights $\omega_{e,\ell}^{\bar{g}}$ sum to zero;
- For relative periods excluded from \mathcal{G} namely $\ell = -1$, the weights $\omega_{e,\ell}^{\bar{g}}$ sum to negative one.
- The indicators for relative period are set to zero for never-treated units, which means $\omega_{\infty,\ell}^{\bar{g}} = 0$ for all ℓ .

In Panel (a), we consider cohort composition without never-treated units, and Panel (b) with never-treated units. Comparing the weights across the panels, we note that never-treated units change the cohort composition, affecting $\omega_{e,\ell}^{\bar{g}}$. While intuitively never-treated units should even out the weights, comparing weights displayed in Panel (b) with (a) makes it clear that is not always the case.

3.2.2 The implication of “trimming” for weighting

Among three out of the eight paper we survey, researchers trim their panel to be balanced in relative periods. In Table 3, we estimate the weights underlying the following “dynamic” specification where $\mathcal{G} = \{\{-2\}, \{0\}, \{1\}, \{2\}\}$

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell \in \{-2,0,1,2\}} \mu_\ell D_{i,t}^\ell + v_{i,t} \quad (12)$$

with indicators for relative periods $\ell \in \{-2,0,1,2\}$. For a panel to be balanced in these relative periods, at minimum we require $E \in [2,4]$ and $T \geq 6$. In Panel (a) of Table 3, we consider a panel balanced in calendar time with $T = 3$. We evenly divide units into cohorts $E_i \in \{1,2,3\}$. In Panel (b) of Table 3, we consider a panel balanced in relative period with $T = 6$. We evenly divide units into cohorts $E_i \in \{2,3,4\}$. While relative periods theoretically range from -4 to 4, we trim the panel to be balanced in relative period from -2 to 2. Thus the relative periods $\ell \notin [-2,2]$ are “trimmed”.

Take μ_{-2} for example, to estimate the weight on $CATT_{e,\ell'}$ in the linear combination introduced in Proposition 1, we regress $D_{i,t}^{\ell'} \cdot \mathbf{1}\{E_i = e\}$ on $D_{i,t}^\ell$ for $\ell \in \{-2,0,1,2\}$ and two-way fixed effects. The coefficient on $D_{i,t}^{-2}$ is the weight $\omega_{e,\ell'}^{-2}$. We illustrate the first three properties listed below Proposition 3 using these weights, displayed in the second first columns of Table 2 Panel (a) and (b).

- For $\ell = -2$, the weights $\omega_{e,\ell}^{-2}$ sum to 1;
- For other relative periods included in \mathcal{G} namely $\ell \in \{0,1,2\}$, the weights $\omega_{e,\ell}^{-2}$ sum to 0 for each of $\ell = 0,1,2$;
- For Panel (a), the relative periods excluded from \mathcal{G} are $\ell \in \{-3,-1\}$ and the weights $\omega_{e,\ell}^{-2}$ sum to -1. For Panel (b), the relative period time excluded from \mathcal{G} is only $\ell \in \{-1\}$ and the weights $\omega_{e,\ell}^{-2}$ sum to -1. The indicators for the imbalanced relative period $\ell \notin [-2,2]$ are set to zero due to trimming, so the weights $\omega_{e,\ell}^{-2}$ for trimmed relative periods $\ell \notin [-2,2]$ are set to zero mechanically.

We note that trimming also changes the cohort composition at relative periods ℓ compared to a panel balanced in calendar time, which explains the difference in weights $\omega_{e,\ell}^{-2}$ across Panel (a) and (b). However, while trimming mechanically sets the $\omega_{e,\ell}^{-2}$ to zero for trimmed relative periods $\ell \notin [-2, 2]$, it does not make the weights convex for its own relative period, namely $\ell = -2$. It also does not make the weights zero for other relative periods, namely $\ell \in \{-1, 0, 1, 2\}$.

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:*

$$\mu_g = \sum_{\ell \in g} \sum_e \omega_{e,\ell}^g CATT_{e,\ell} + \sum_{\ell' \notin g} \sum_e \omega_{e,\ell'}^g CATT_{e,\ell'}. \quad (13)$$

Under Assumption 1, we can provide a causal interpretation of the the difference in trends stated in Proposition 1, namely $E[Y_{i,e+\ell} - Y_{i,0}^\infty | E_i] - E[Y_{i,e+\ell}^\infty - Y_{i,0}^\infty] = CATT_{e,\ell} + \underbrace{E[Y_{i,t}^\infty - Y_{i,0}^\infty | E_i] - E[Y_{i,t}^\infty - Y_{i,0}^\infty]}_{=0}$ for $t = e + \ell$. The weights are still functions of cohort compositions, same as those in Proposition 1, suggesting μ_g can be written as non-convex averages of not only $CATT_{e,\ell}$ from own periods $\ell \in g$, but also $CATT_{e,\ell'}$ from other periods. This implies that μ_g could fall outside the convex hull of $CATT_{e,\ell \in g}$, which means it is possible for μ_g to be of the opposite sign to all $CATT_{e,\ell \in g}$. Calculating and examining the weights lets researchers gauge how treatment effect heterogeneity would interact with potential non-convex weighting in μ_g . When the weights have larger magnitude, treatment effect heterogeneity matters more as a particular $CATT_{e,\ell}$ can drive the overall estimates. When the weights are more uniform, treatment effect heterogeneity matter less.

3.4 Interpreting the coefficients under parallel trends and no anticipation assumptions

Proposition 3. *If Assumption 1 (parallel trends) holds and Assumption 2 holds for all $\ell < 0$ (no anticipatory behavior in all periods before the initial treatment), the population regression coefficient $\mu_{\underline{g}}$ for \underline{g} containing some leads prior to the treatment is a linear combination of post-treatment $CATT_{e,\ell'}$ for all $\ell' \geq 0$*

$$\mu_{\underline{g}} = \sum_{\ell' \geq 0} \sum_e \omega_{e,\ell'}^{\underline{g}} CATT_{e,\ell'}. \quad (14)$$

Invalidity of pretrend tests based on pre-period coefficients. Proposition 3 implies that when effects are not homogenous across cohorts, it is problematic to interpret non-zero estimates for $\mu_{\underline{g}}$ as evidence for pretrends, where \underline{g} contains some leads $\ell < 0$. 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, due to the influence of post-treatment $CATT_{e,\ell' \geq 0}$, this estimand $\mu_{\underline{g}}$ is not necessarily zero. This invalidates a common pretrend test used in practice, which is to test that none of the pre-period coefficients is individually statistically significant. 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.

Next we present through simulation two examples where FE estimates do not appropriately summarize the underlying pretrends. We consider specification (12) underlying Table 2 Panel (a) as discussed in Section 3.2.2, with a balanced panel where E_i is drawn uniformly from $\{1, 2, 3\}$ and $T = 3$. We defer simulation details to Online Appendix C.

While we set $CATT_{e,\ell < 0} = 0$ for all e , the FE estimates of $\hat{\mu}_{-2}$ from specification (12) are statistically distinguishable from zero, usually negative as shown in the left panel of Figure 1, which would suggest a spurious negative pretrend when none exists. We then set $CATT_{e,-1} = 3 \forall e$ to reflect anticipatory behavior in the period right before treatment. For each simulation, we re-estimate specification (12). The estimates of $\hat{\mu}_{-2}$ are now statistically indistinguishable from zero

as shown in the right panel, which would suggest a lack of pre-trend when one does exist. The reason for such failure is that the weights underlying μ_{-2} , as displayed in Column (1) of Table 2 Panel (a) and discussed in Section 3.2.2, are non-zero for post-periods, namely $\omega_{1,0}^{-2}$, $\omega_{2,0}^{-2}$, $\omega_{3,0}^{-2}$, $\omega_{1,1}^{-2}$ and $\omega_{2,1}^{-2}$. In our simulation we specify a large amount of treatment effect heterogeneity in these post-periods so that they affect μ_{-2} . See Callaway and Sant’Anna (2018) for alternative tests for pretrends that do not suffer from this drawback.

3.5 Interpreting the coefficients under all three assumptions

Proposition 4. *If Assumption 1 (parallel trends) holds, Assumption 2 holds for all $l < 0$ (no anticipatory behavior in all periods before the initial treatment), and Assumption 3 (treatment effect homogeneity) holds, then $CATT_{e,\ell} = CATT_\ell$ is constant across e for a given ℓ , and the population regression coefficient μ_g is equal to a linear combination of $CATT_{\ell \in g}$, minus a linear combination of $CATT_{\ell \notin g}$ from other relative periods*

$$\mu_g = \sum_{\ell \in g} \omega_\ell^g CATT_\ell + \sum_{g' \neq g} \sum_{\ell \in g'} \omega_\ell^g CATT_\ell + \sum_{\ell \in g^{excl}} \omega_\ell^g CATT_\ell \quad (15)$$

The weight $\omega_\ell^g = \sum_e \omega_{e,\ell}^g$ sums over the weights $\omega_{e,\ell}^g$ from Proposition 1. The weight ω_ℓ^g is also equal to the population regression coefficient on $\mathbf{1}\{t - E_i \in g\}$ from regressing $D_{i,t}^\ell$ on all the group indicators i.e. $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ included in (3) and two-way fixed effects.

3.5.1 Binning vs excluding distant relative periods

Binning is more common in practice to accommodate “unbalanced” relative periods i.e. relative periods for which $\ell < -\min(E_i)$ and $\ell > T - \max(E_i)$ because not all cohorts are observed for these relative periods. Six out of the eight applied papers we survey bin distant relative periods. Perhaps the popularity of binning is based on the belief that a “binned” specification is more interpretable.

Proposition 4 does not provide any theoretical advantage to one or the other. But examining the weights for a given specification cohort composition can guide our choice on whether we should

bin or exclude relative periods when choosing a specification. We provide a numerical example where excluding “unbalanced” relative periods can actually provide more reasonable weights.

We construct a panel balanced in calendar time with $N = 1000$ and $T = 6$. We divide the units evenly into cohorts $E_i \in \{2, 3, 4\}$. In Table 4 Panel (a), we estimate the weights underlying the following “binned” specification:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell \in \{-2, 0, 1, 2\}} \mu_\ell D_{i,t}^\ell + \mu_{\underline{g}} \mathbf{1}\{t - E_i < -2\} + \mu_{\bar{g}} \mathbf{1}\{t - E_i > 2\} + v_{i,t}. \quad (16)$$

In Table 4 Panel (b), we estimate the weights underlying the following specification that excludes distant relative periods $\ell \in \{-4, -3, -1, 3, 4\}$:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell \in \{-2, 0, 1, 2\}} \mu_\ell D_{i,t}^\ell + v_{i,t}. \quad (17)$$

In this example, binning actually adds $CATT_{-4}$ to several coefficients, as we have positive ω_{-4}^0 , ω_{-4}^1 and ω_{-4}^2 as shown in the first row of Panel (a). In contrast, excluding these “unbalanced” relative periods assigns more reasonable weights. It subtracts from the coefficients a weighted average of excluded periods effects, as the weights ω_ℓ^g for $\ell \in \{-4, -3, -1, 3, 4\}$ are all negative.

We provide some intuition for how weights can be less interpretable in the “binned” specification. For the “binned” specification the weights on distant relative periods $\ell' \in g'$ underlying μ_ℓ sum to zero i.e. $\sum_{\ell \in g'} \omega_\ell^g = 0$ for each bin g' . This suggests some of the weights have to be positive, implying we add the corresponding effects to the coefficient μ_ℓ . For the specification where we exclude distant relative periods $\ell' \in g'$ as well as $\ell = -1$, underlying μ_ℓ we subtract a linear combination of their effect as well as $CATT_{-1}$ as their weights sum to negative one i.e. $\sum_{\ell' \notin \mathcal{G}} \omega_{\ell'}^\ell = -1$. While some of the weights can still be positive, we are more likely to subtract $CATT_{\ell' \in g'}$ from μ_ℓ than binning.

4 Alternative estimation method

We propose a new estimation method that is robust to treatment effect heterogeneity. The goal of our method is to estimate an interpretable analog of the relative period coefficient μ_g from the two-way fixed effects regression, namely a weighted average of $CATT_{e,\ell}$ for $\ell \in g$ with reasonable weights (i.e. 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 :

$$\nu_g = \frac{1}{|g|} \sum_{\ell \in g} \sum_e CATT_{e,\ell} Pr\{E_i = e \mid E_i \in [-\ell, T - \ell]\} \quad (18)$$

One can aggregate $CATT_{e,\ell}$ to form other parameters of interest, such as those proposed by Callaway and Sant’Anna (2018). 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. (2018). 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 :

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \notin C} \sum_{\ell \neq -1} \delta_{e,\ell} (\mathbf{1}\{E_i = e\} \cdot D_{i,t}^\ell) + \epsilon_{i,t} \quad (19)$$

The exact specification depends on the cohort compositions for a given application. If there is a never-treated cohort i.e. $\infty \in \text{supp}\{E_i\}$, then we may set $C = \{\infty\}$ and estimate regression (19) on all observations. If there are no never-treated units i.e. $\infty \notin \text{supp}\{E_i\}$, then we may set $C = \{\max\{E_i\}\}$ i.e. the latest-treated cohort and estimate regression (19) on observations from $t = 0, \dots, \max\{E_i\} - 1$. Lastly, if there is a cohort that is always treated i.e. $0 \in \text{supp}\{E_i\}$, then we need to exclude this cohort from estimation.

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

$$\hat{v}_g = \frac{1}{|g|} \sum_{\ell \in g} \sum_e \hat{\delta}_{e,\ell} \widehat{\Pr}\{E_i = e \mid E_i \in [-\ell, T - \ell]\} \quad (20)$$

where $\hat{\delta}_{e,\ell}$ 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 $|g| = 1$.

The coefficient estimator $\hat{\delta}_{e,\ell}$ from regression (19) is a DID estimator for $CATT_{e,\ell}$ with particular choices of pre-periods and control cohorts. Since DID is likely a familiar estimator for applied researchers, we defer its definition and choices of pre-periods and control cohorts to Section 4.2.

Under parallel trends and no anticipation assumptions the coefficient estimator $\widehat{\delta}_{e,\ell}$ from Regression (19) 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 (19), 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 bootstrap as in Callaway and Sant’Anna (2018).

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

$$\widehat{\delta}_{e,\ell} = \frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s}) \cdot \mathbf{1}\{E_i = e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]} - \frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s}) \cdot \mathbf{1}\{E_i \in C\}]}{\mathbb{E}_N[\mathbf{1}\{E_i \in C\}]} \quad (21)$$

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 (19) in Step 1.

Using the above definition, in regression (19) 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. $\infty \in \text{supp}\{E_i\}$, then we set the control cohort to be never-treated units $C = \{\infty\}$. If there are no never-treated units i.e. $\infty \notin \text{supp}\{E_i\}$, then we set the control cohort $C = \{\max\{E_i\}\}$ i.e. the latest-treated cohort. However, without never-treated units we need to drop time periods

$t \geq \max\{E_i\}$ 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, then we 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 Assumption 1 holds and Assumption 2 holds for all pre-periods i.e. $\underline{g} = \{\ell : \ell < 0\}$, 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 (2018), the outcome regression approach by Heckman et al. (1997), and the doubly robust estimator recently proposed by Sant’Anna and Zhao (2018). The resulting estimates $\widehat{\delta}_{e,\ell}$ can then be plugged into step 3 to form our IW estimator.

Imposing no anticipation for all $\ell < 0$ means we can choose any pre-period $s < e$ and include in control cohorts C any cohorts not yet treated at $e + \ell$. Among all possible pre-periods, for regression (19) 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. Allowing some anticipation restricts the choices of s and C . 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 (for example, some $s < e - 2$) and a control group (for example, cohorts which will be treated only after at least 2 periods from now $C \subseteq \{c : \max\{e + \ell, e + \ell + 2\} < c \leq T\}$).

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 effect 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 likely to hold 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 in two-way fixed effects regressions would thus apply to their 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 re-

striction 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 5 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 individ-

uals with high out-of-pocket medical spending and low labor earnings, even when restricted to individuals who were ever hospitalized. Thus, it is not plausible that the baseline outcome $Y_{i,t}^\infty$ 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.

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)

$$Y_{i,t} = \alpha_i + \lambda_t + \mu_{-3}D_{i,t}^{-3} + \mu_{-2}D_{i,t}^{-2} + \mu_0D_{i,t}^0 + \mu_1D_{i,t}^1 + \mu_2D_{i,t}^2 + \mu_3D_{i,t}^3 + v_{i,t}. \quad (22)$$

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. (relative period in the original sample ranges from nine waves prior to the hospitalization to nine waves after the hospitalization.) To fully illustrate issues with this specification, we do not trim, but rather use a sample balanced in calendar time for $t = 0, \dots, 4$, 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 $\ell = -4$. But since we do not trim, our results are not directly comparable to Dobkin et al. (2018), even though they are quite similar.

We focus on a single coefficient μ_0 that is supposed to capture the contemporaneous effect of hospitalizations. As we find in Proposition 2 μ_0 can be decomposed as

$$\sum_{e=1}^4 \omega_{e,0}^0 CATT_{e,0} + \sum_{\ell=-3, \neq -1}^3 \sum_{e=1}^4 \omega_{e,\ell}^0 CATT_{e,\ell} + \sum_{\ell'=-4, -1} \sum_e \omega_{e,\ell'}^0 CATT_{e,\ell'}. \quad (23)$$

As discussed in Proposition 2 we can estimate the underlying weights $\omega_{e,\ell}^0$ by regressing $\mathbf{1}\{E_i = e\} \cdot D_{i,t}^\ell$ on the relative wave indicators included in specification (22) i.e. $\{D_{i,t}^\ell\}_{\ell=-3, \neq -1}^3$ and two-way fixed effects. The coefficient estimator of $D_{i,t}^0$ in such regression, $\widehat{\omega}_{e,\ell}^0$, consistently estimates $\omega_{e,\ell}^0$.

Figure 3 plots these estimated weights. To match with our decomposition results and to provide more intuition behind the weights, we note that (a) the four weights from relative wave $\ell = 0$ sum to one; (b) the weights for other included relative waves $\ell \in \{-3, -2, 1, 2, 3\}$ 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 large for leads of treatments (negative relative waves), which suggest that the FE estimate $\widehat{\mu}_0$ is particularly sensitive to estimates of pre-trends, and does not isolate the contemporaneous effect of hospitalizations.

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 (19) as

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \in \{1,2,3\}} \sum_{\ell=-3, \neq -1}^2 \delta_{e,\ell} \mathbf{1}\{E_i = e\} \cdot D_{i,t}^\ell + \epsilon_{i,t} \quad (24)$$

for $t = 0, \dots, 3$. 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 periods. 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{v}_ℓ 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 6, we report the FE estimates $\widehat{\mu}_\ell$ and the IW estimates \widehat{v}_ℓ , 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 small in comparison and transitory.

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, $\widehat{\mu}_0$ still 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{v}_ℓ 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 from two-way fixed effects regressions in settings with variation in treatment timing and treatment effect heterogeneity. We show that under some assumptions, the coefficients μ_ℓ on the indicator for being ℓ periods away from the treatment can be decomposed as a linear combination of cohort-specific average treatment effect on the treated $CATT_{e,\ell}$. However, the weights can be non-convex, and non-zero on relative periods $\ell' \neq \ell$. These weights are unlikely what researchers have in mind with two-way fixed effects regression, as under heterogeneous treatment effects, μ_ℓ 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 .

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 heterogeneous 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. IW estimates, on the other hand, are weighted averages of the underlying $CATT$ estimates with weights representative of cohort share.

References

- Abadie, Alberto**, “Semiparametric Difference-in-Differences Estimators,” *The Review of Economic Studies*, 2005, 72 (1), 1–19.
- 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 2018.
- Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey**, “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 2013, 81 (2), 535–580.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-way fixed effects estimators with heterogeneous treatment effects,” Working Paper arXiv:1803.08807 [econ.EM], ArXiv October 2019.

- 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*, 2018, *forthcoming*.
- 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.
- Robins, James**, “A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect,” *Mathematical Modelling*, 1986, 7 (9), 1393 – 1512.
- Robins, James M.**, “Causal Inference from Complex Longitudinal Data,” in Maia Berkane, ed., *Latent Variable Modeling and Applications to Causality*, Springer New York New York, NY 1997, pp. 69–117.

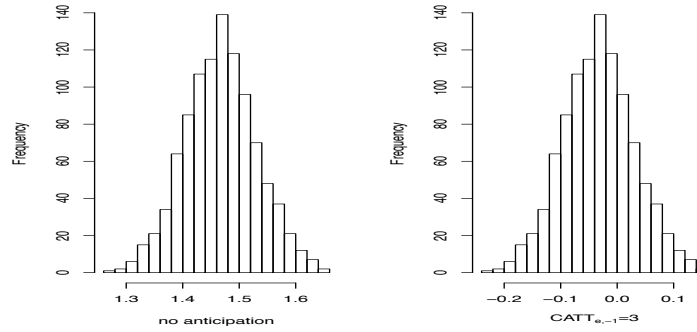
Roth, Jonathan, “Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends,” Working Paper, Harvard University July 2019.

Sant’Anna, Pedro H. C. and Jun B. Zhao, “Doubly Robust Difference-in-Differences Estimators,” Technical Report 2018.

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.

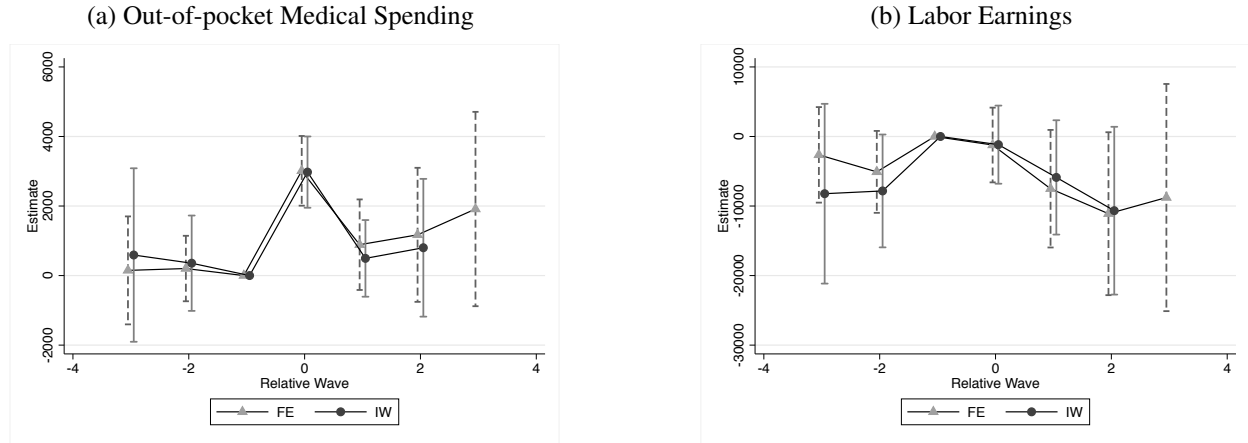
Tewari, Ishani, “The Distributive Impacts of Financial Development: Evidence from Mortgage Markets during US Bank Branch Deregulation,” *American Economic Journal: Applied Economics*, October 2014, 6 (4), 175–96.

Figure 1: Simulated Distribution of Lead Estimates without and with Anticipatory Behavior



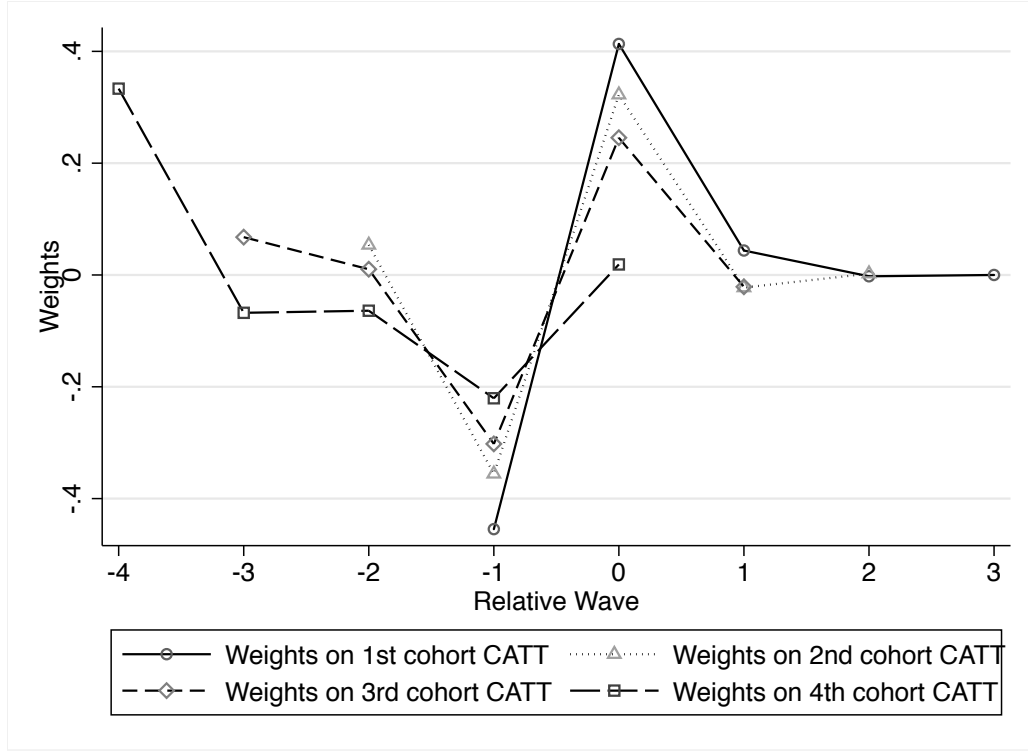
Notes: This figure plots the histogram of $\hat{\mu}_{-2}$, the FE and IW estimates for the treatment effect one period before treatment, across 1000 simulated samples. The true underlying treatment effects one period before treatment are set to zero for all cohorts in these simulations in the left panel, whereas the true underlying treatment effects one period before treatment are set to be three for all cohorts in these simulations in the right panel.

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



Notes: Each figure plots FE estimates $\hat{\mu}_\ell$ in triangles and IW estimates \hat{v}_ℓ 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: Weights $\hat{\omega}_{e,\ell}^0$ on cohort specific effect estimates $\hat{\delta}_{e,\ell}$ in forming point estimate $\hat{\mu}_0$



Notes: The FE estimate for the instantaneous effect of hospitalization $\hat{\mu}_0$ is a linear combination of $\hat{\delta}_{e,\ell}$'s, estimates for cohort-specific effects $CATT_{e,\ell}$'s from all cohorts e and relative waves ℓ . This figure plots the weight $\hat{\omega}_{e,\ell}^0$ associated with each $\hat{\delta}_{e,\ell}$ in forming the FE estimate $\hat{\mu}_0$.

Table 1: 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: 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.

Table 2: Numerical illustration of weights: the implication of never-treated units

(a) $E_i \in \{1, 2, 3\}$			(b) $E_i \in \{1, 2, 3, \infty\}$		
$\omega_{e,\ell}^g$	$[-T, -2]$	$[0, T]$	$\omega_{e,\ell}^g$	$[-T, -2]$	$[0, T]$
$\omega_{3,-3}^g$	0.041	0.292	$\omega_{3,-3}^g$	0.190	0.119
$\omega_{2,-2}^g$	0.417	-0.084	$\omega_{2,-2}^g$	0.405	-0.060
$\omega_{3,-2}^g$	0.542	-0.208	$\omega_{3,-2}^g$	0.405	-0.060
$\omega_{1,-1}^g$	-0.458	-0.208	$\omega_{1,-1}^g$	-0.333	-0.333
$\omega_{2,-1}^g$	-0.334	-0.334	$\omega_{2,-1}^g$	-0.333	-0.333
$\omega_{3,-1}^g$	-0.208	-0.458	$\omega_{3,-1}^g$	-0.333	-0.333
$\omega_{1,0}^g$	-0.208	0.542	$\omega_{1,0}^g$	-0.024	0.298
$\omega_{2,0}^g$	-0.084	0.417	$\omega_{2,0}^g$	-0.024	0.298
$\omega_{3,0}^g$	-0.375	0.375	$\omega_{3,0}^g$	-0.262	0.274
$\omega_{1,1}^g$	0.292	0.041	$\omega_{1,1}^g$	0.190	0.119
$\omega_{2,1}^g$	0.000	0.000	$\omega_{2,1}^g$	-0.048	0.095
$\omega_{1,2}^g$	0.375	-0.375	$\omega_{1,2}^g$	0.167	-0.083

Notes: We construct a panel balanced in calendar time with $N = 1000$ and $T = 3$. We divide the units evenly into cohorts. For Panel (a) with $E_i \in \{1, 2, 3\}$, we set $E_i = 1$ for one third of the units and so on. For panel (b) with $E_i \in \{1, 2, 3, \infty\}$, we set $E_i = 1$ for one fourth of the units and so on.

Table 3: Numerical illustration of weights: the role of trimming

(a) Balanced in calendar time: $E_i \in \{1, 2, 3\}$ and $T = 3$

$\omega_{e,\ell}^g$	μ_{-2}	μ_0	μ_1	μ_2
$\omega_{3,-3}^g$	-0.500	0.500	1.000	1.500
$\omega_{2,-2}^g$	0.532	0.031	-0.125	-0.094
$\omega_{3,-2}^g$	0.468	-0.031	0.125	0.094
$\omega_{1,-1}^g$	-0.032	-0.531	-0.875	-1.406
$\omega_{2,-1}^g$	-0.406	-0.407	-0.375	-0.281
$\omega_{3,-1}^g$	-0.062	-0.563	-0.750	-0.813
$\omega_{1,0}^g$	-0.062	0.437	0.250	0.187
$\omega_{2,0}^g$	-0.032	0.469	0.125	0.594
$\omega_{3,0}^g$	0.094	0.093	-0.375	-0.781
$\omega_{1,1}^g$	0.094	0.093	0.625	0.219
$\omega_{2,1}^g$	-0.094	-0.093	0.375	-0.219
$\omega_{1,2}^g$	0.000	0.000	0.000	1.000

(b) Balanced in relative time: $E_i \in \{2, 3, 4\}$ and $T = 6$

$\omega_{e,\ell}^g$	μ_{-2}	μ_0	μ_1	μ_2
$\omega_{4,-4}^g$	0.000	0.000	0.000	0.000
$\omega_{3,-3}^g$	0.000	0.000	0.000	0.000
$\omega_{4,-3}^g$	0.000	0.000	0.000	0.000
$\omega_{2,-2}^g$	0.167	-0.167	-0.333	-0.500
$\omega_{3,-2}^g$	0.381	0.177	0.212	0.260
$\omega_{4,-2}^g$	0.452	-0.011	0.121	0.240
$\omega_{2,-1}^g$	-0.381	-0.177	-0.212	-0.260
$\omega_{3,-1}^g$	-0.417	-0.402	-0.303	-0.387
$\omega_{4,-1}^g$	-0.202	-0.421	-0.485	-0.354
$\omega_{2,0}^g$	-0.035	0.412	0.182	0.146
$\omega_{3,0}^g$	0.071	0.357	0.000	0.071
$\omega_{4,0}^g$	-0.036	0.230	-0.182	-0.218
$\omega_{2,1}^g$	0.131	0.064	0.485	0.282
$\omega_{3,1}^g$	-0.083	-0.098	0.303	-0.113
$\omega_{4,1}^g$	-0.048	0.035	0.212	-0.169
$\omega_{2,2}^g$	0.119	-0.132	-0.121	0.331
$\omega_{3,2}^g$	0.048	-0.035	-0.212	0.169
$\omega_{4,2}^g$	-0.167	0.167	0.333	0.500
$\omega_{2,3}^g$	0.000	0.000	0.000	0.000
$\omega_{3,3}^g$	0.000	0.000	0.000	0.000
$\omega_{2,4}^g$	0.000	0.000	0.000	0.000

Notes: For Panel (a) with $E_i \in \{1, 2, 3\}$ and $T = 3$, we construct a panel balanced in calendar time with $N = 1000$ and divide the units evenly into cohorts. That is, we set $E_i = 1$ for one third of the units and so on. We keep all observations $\{i, t\}$ for $i = 1, \dots, N$ and $t = 0, \dots, T$. For panel (b) with $E_i \in \{2, 3, 4\}$ and $T = 6$, we construct a panel balanced in relative time with $N = 1000$ and divide the units evenly into cohorts. That is, we set $E_i = 2$ for one third of the units and so on. We keep observations $\{i, t\}$ for which $E_i - t \in [-2, 2]$.

Table 4: Numerical illustration of weights: binning vs excluding distant relative times

(a) Binning “unbalanced” relative periods				
ω_{ℓ}^g	$\{-2\}$	$\{0\}$	$\{1\}$	$\{2\}$
ω_{-4}^g	-0.536	0.488	1.000	1.464
ω_{-3}^g	0.536	-0.488	-1.000	-1.464
ω_{-2}^g	1.000	0.000	0.000	0.000
ω_{-1}^g	-1.000	-1.000	-1.000	-1.000
ω_0^g	0.000	1.000	0.000	0.000
ω_1^g	0.000	0.000	1.000	0.000
ω_2^g	0.000	0.000	0.000	1.000
ω_3^g	-0.464	0.512	1.000	1.536
ω_4^g	0.464	-0.512	-1.000	-1.536

(b) Excluding “unbalanced” relative periods				
ω_{ℓ}^g	$\{-2\}$	$\{0\}$	$\{1\}$	$\{2\}$
ω_{-4}^g	-0.172	-0.036	-0.016	-0.011
ω_{-3}^g	-0.346	-0.072	-0.031	-0.022
ω_{-2}^g	1.000	0.000	0.000	0.000
ω_{-1}^g	-0.433	-0.597	-0.448	-0.249
ω_0^g	0.000	1.000	0.000	0.000
ω_1^g	0.000	0.000	1.000	0.000
ω_2^g	0.000	0.000	0.000	1.000
ω_3^g	-0.038	-0.229	-0.413	-0.509
ω_4^g	-0.011	-0.067	-0.092	-0.208

Notes: We construct a panel balanced in calendar time with $N = 1000$ and $T = 6$. We divide the units evenly into cohorts $E_i \in \{2, 3, 4\}$. That is, we set $E_i = 2$ for one third of the units and so on. Panel (a) reports weights underlying the “binned” specification, where distant relative times are collected in $\underline{g} = \{-4, -3\}$ and $\bar{g} = \{3, 4\}$. Panel (b) reports weights underlying the specification that excludes distant relative times $\{-4, -3, 3, 4\}$.

Table 5: Summary Statistics of the HRS sample

	<i>N</i>	Mean	Std. Dev
<i>Panel A. Demographics</i>			
Age at admission	656	56	2.29
Male	656	0.456	0.498
Year of admission	656	2,007	2.11
<i>Panel B. Race/ethnicity</i>			
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
<i>Panel C. Insurance status</i>			
Medicaid	656	0.05	0.219
Private	656	0.715	0.452
Medicare	656	0.072	0.259
<i>Panel D. Pre-hospitalization outcome</i>			
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 6: Estimates for the Effect of Hospitalization on Outcomes

(a) Out-of-pocket Medical Spending					
	FE	IW	Estimates for $CATT_{e,\ell}$		
ℓ wave relative to hospitalization	estimates	estimates	$\widehat{\delta}_{1,\ell}$	$\widehat{\delta}_{2,\ell}$	$\widehat{\delta}_{3,\ell}$
-3	149 (792)	591 (1273)	-	-	591 (1273)
-2	203 (480)	356 (700)	-	299 (967)	411 (1030)
-1	0	0	0	0	0
0	3,013 (511)	2,975 (523)	2,826 (1038)	3,031 (704)	3,092 (998)
1	888 (664)	494 (562)	825 (912)	107 (653)	-
2	1,172 (983)	800 (1011)	800 (1011)	-	-
3	1,914 (1426)	-	-	-	-
(b) Labor Earnings					
	FE	IW	Estimates for $CATT_{e,\ell}$		
ℓ wave relative to hospitalization	estimates	estimates	$\widehat{\delta}_{1,\ell}$	$\widehat{\delta}_{2,\ell}$	$\widehat{\delta}_{3,\ell}$
-3	-2,642 (3504)	-8,228 (6594)	-	-	-8,228 (6594)
-2	-5,089 (3005)	-7,830 (4139)	-	-7,691 (6349)	-7,964 (4766)
-1	0	0	0	0	0
0	-1,225 (2742)	-1,175 (2866)	455 (5593)	3,032 (4153)	-7,107 (5883)
1	-7,508 (4312)	-5,886 (2866)	-1,670 (6500)	-10,826 (4863)	-
2	-11,102 (5976)	-10,670 (6155)	-10,670 (6155)	-	-
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 (Panel A) and labor earnings (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.

A Notation glossary and properties to double-demeaning

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

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

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:

$$\ddot{X}_{i,t} = X_{i,t} - E[X_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T X_{i,s} + \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] \quad (25)$$

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

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 , $\frac{1}{T+1} \sum_{s=0}^T X_{i,s}$. 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:

$$E[\ddot{X}_{i,t}] = E \left[X_{i,t} - E[X_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T X_{i,s} + \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] \right] \quad (26)$$

$$= E[X_{i,t}] - E[X_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] + \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] = 0 \quad (27)$$

and

$$\sum_{t=0}^T \ddot{X}_{i,t} = \sum_{t=0}^T X_{i,t} - \sum_{t=0}^T E[X_{i,t}] - \sum_{t=0}^T \frac{1}{T+1} \sum_{s=0}^T X_{i,s} + \sum_{t=0}^T \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] \quad (28)$$

$$= \sum_{t=0}^T X_{i,t} - \sum_{t=0}^T E[X_{i,t}] - \sum_{s=0}^T X_{i,s} + \sum_{s=0}^T E[X_{i,s}] = 0 \quad (29)$$

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

$$= E \left[\left(X_{i,t} - \frac{1}{T+1} \sum_{s=0}^T X_{i,s} \right) \ddot{Z}_{i,t} \right] - E \left[\left(E[X_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] \right) \ddot{Z}_{i,t} \right] \quad (30)$$

$$= E \left[\left(X_{i,t} - \frac{1}{T+1} \sum_{s=0}^T X_{i,s} \right) \ddot{Z}_{i,t} \right] - \left(E[X_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T E[X_{i,s}] \right) \cdot E[\ddot{Z}_{i,t}] \quad (31)$$

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

$$\sum_{t=0}^T E[X_{i,t} \ddot{Z}_{i,t}] - \sum_{t=0}^T E \left[\left(\frac{1}{T+1} \sum_{s=0}^T X_{i,s} \right) \ddot{Z}_{i,t} \right] = \sum_{t=0}^T E[X_{i,t} \ddot{Z}_{i,t}] - E \left[\left(\frac{1}{T+1} \sum_{s=0}^T X_{i,s} \right) \sum_{t=0}^T \ddot{Z}_{i,t} \right] = \sum_{t=0}^T E[X_{i,t} \ddot{Z}_{i,t}] \quad (32)$$

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

$$\ddot{Z}_{i,t} = Z_{i,t} - E[Z_{i,t}] - \frac{1}{T+1} \sum_{s=0}^T Z_{i,s} + \frac{1}{T+1} \sum_{s=0}^T E[Z_{i,s}] \quad (33)$$

$$= Z_i - E[Z_i] - Z_i + E[Z_i] = 0 \quad (34)$$

□

B Proofs

Proof of Proposition 1

Proof. Collect the relative time indicators in a column vector $\mathbf{D}_{i,t} = (\mathbf{1}\{t - E_i \in g\})_{g \in \mathcal{G}}^\top$. Collect their corresponding coefficients in a column vector $\mu = (\mu_g)_{g \in \mathcal{G}}^\top$. Partialling out the unit and time fixed effects, Regression (3) is

$$\ddot{Y}_{i,t} = \mu^\top \ddot{\mathbf{D}}_{i,t} + v_{i,t} \quad (35)$$

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

$$\mu_g = \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \ddot{\mathbf{D}}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} Y_{i,t}] \quad (36)$$

where \mathbf{e}_g is a column vector with 1 in the entry corresponding to the entry of $\mathbf{1}\{t - E_i \in g\}$ in $\mathbf{D}_{i,t}$, and 0 otherwise.

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

$$\begin{aligned} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} Y_{i,t}] &= \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} Y_{i,t}] - \underbrace{\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} Y_{i,0}^\infty]}_{=0} = \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} (Y_{i,t} - Y_{i,0}^\infty)] = \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} E[Y_{i,t} - Y_{i,0}^\infty | E_i]] \end{aligned} \quad (37)$$

$$= \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} E[Y_{i,t} - Y_{i,0}^\infty | E_i]] - \sum_{t=0}^T \underbrace{E[\ddot{\mathbf{D}}_{i,t}]}_{=0} E[Y_{i,t}^\infty - Y_{i,0}^\infty] \quad (38)$$

$$= \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} (E[Y_{i,t} - Y_{i,0}^\infty | E_i] - E[Y_{i,t}^\infty - Y_{i,0}^\infty])]. \quad (39)$$

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

$$E[Y_{i,t} - Y_{i,0}^\infty | E_i] - E[Y_{i,t}^\infty - Y_{i,0}^\infty] =: f(E_i, t) = \sum_{e=0}^{\infty} f(e, t) \cdot \mathbf{1}\{E_i = e\} \quad (40)$$

$$= \sum_{e=0}^{\infty} \sum_{\ell=-T}^T f(e, e + \ell) \cdot \mathbf{1}\{t - e = \ell\} \cdot \mathbf{1}\{E_i = e\} \quad (41)$$

$$= \sum_{\ell=-T}^T \sum_{e=0}^{\infty} D_{i,t}^\ell \mathbf{1}\{E_i = e\} \cdot f(e, e + \ell). \quad (42)$$

Even though we sum over $e = 0, \dots, \infty$, 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 $E_i = e'$). So the range of the summation is still sensible. We can replace $\mathbf{1}\{t - e = \ell\} \cdot \mathbf{1}\{E_i = e\}$ with the more compact notation $D_{i,t}^\ell \mathbf{1}\{E_i = e\}$ because only under the joint event $E_i = e$ and $E_i = t + \ell$, the product is one:

$$D_{i,t}^\ell \mathbf{1}\{E_i = e\} = \mathbf{1}\{t - e = \ell\} \cdot \mathbf{1}\{E_i = e\}. \quad (43)$$

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

$$\mu_g = \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T \sum_{\ell=-T}^T \sum_{e=0}^{\infty} E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}] f(e, e + \ell) \quad (44)$$

$$= \sum_{t=0}^T \sum_{l \in g} \sum_{e=0}^{\infty} \underbrace{\mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}]}_{=\omega_{e,\ell}^g} f(e, e + \ell) \quad (45)$$

$$+ \sum_{t=0}^T \sum_{g' \neq g} \sum_{\ell \in g'} \sum_{e=0}^{\infty} \underbrace{\mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}]}_{=\omega_{e,\ell}^g} f(e, e + \ell) \quad (46)$$

$$+ \sum_{t=0}^T \sum_{\ell \in g^{excl}} \sum_{e=0}^{\infty} \underbrace{\mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}]}_{=\omega_{e,\ell}^g} f(e, e + \ell) \quad (47)$$

The superscript g in $\omega_{e,\ell}^g$ indexes μ_g . The subscript ℓ in $\omega_{e,\ell}^g$ indexes $f(e, e + \ell)$. The expression above the braces makes it clear that 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 the bin indicators i.e. $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ and two-way fixed effects included in (3). 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 i.e. $\ell \in g$, the weights sum up to one $\sum_{\ell \in g} \sum_e \omega_{e,\ell}^g = 1$.
2. For relative times belong to some other bin i.e. $\ell \in g'$ and $g' \neq g$, the weights sum up to zero $\sum_{\ell \in g'} \sum_e \omega_{e,\ell}^g = 0$ for each bin g' .
3. For relative times not contained in \mathcal{G} i.e. $\ell \in g^{excl}$, the weights sum up to minus one $\sum_{\ell \in g^{excl}} \sum_e \omega_{e,\ell}^g = -1$.
4. If there are never-treated units i.e. $\infty \in \text{supp}(E_i)$, we have $\omega_{\infty,\ell}^g = 0$ for all g and ℓ .

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

$$\sum_{t=0}^T \sum_{\ell \in g} \sum_e \omega_{e,\ell}^g = \sum_{t=0}^T \sum_{\ell \in g} \sum_e \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell \cdot \mathbf{1}\{E_i = e\}] \quad (48)$$

$$= \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \sum_{\ell \in g} D_{i,t}^\ell \cdot \sum_e \mathbf{1}\{E_i = e\}] \quad (49)$$

$$= \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{1}\{t - E_i \in g\}] \quad (50)$$

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

$$\sum_{\ell \in g^{excl}} \sum_e \omega_{e,\ell}^g = \mathbf{e}_g^\top \sum_{\ell \in g^{excl}} \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} D_{i,t}^\ell] \quad (51)$$

$$= \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E \left[\ddot{\mathbf{D}}_{i,t} \sum_{\ell \in g^{excl}} D_{i,t}^\ell \right] \quad (52)$$

$$= \mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} (1 - \sum_{g \in \mathcal{G}} \sum_{\ell \in g} D_{i,t}^\ell)] \quad (53)$$

$$= -\mathbf{e}_g^\top \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \sum_{g \in \mathcal{G}} \sum_{\ell \in g} D_{i,t}^\ell] \quad (54)$$

$$= -\mathbf{e}_g^\top \sum_{g \in \mathcal{G}} \left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{1}\{t - E_i \in g\}] = -1 \quad (55)$$

where the third equality follows from $\sum_{\ell \in g^{excl}} D_{i,t}^\ell + \sum_{g \in \mathcal{G}} \sum_{\ell \in g} D_{i,t}^\ell = \sum_{-T \leq \ell \leq T} D_{i,t}^\ell = 1$. 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

$$\left(\sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^T E[\ddot{\mathbf{D}}_{i,t} \mathbf{1}\{t - E_i \in g\}] \quad (56)$$

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 $\omega_{\infty,\ell}^g = 0$. This is because $D_{i,t}^\ell = 0$ for all ℓ when $E_i = \infty$. \square

Proof of Proposition 2

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

$$E[Y_{i,e+\ell} - Y_{i,0}^\infty | E_i] - E[Y_{i,e+\ell}^\infty - Y_{i,0}^\infty] = CATT_{e,\ell} + \underbrace{E[Y_{i,t}^\infty - Y_{i,0}^\infty | E_i] - E[Y_{i,t}^\infty - Y_{i,0}^\infty]}_{=0} \quad (57)$$

for $t = e + \ell$. \square

Proof of Proposition 3

Proof. Under no anticipation for all leads $\ell < 0$, we have $CATT_{e,\ell} = 0$ for all $\ell < 0$, setting the corresponding terms to zero in the linear combination underlying $\mu_{\underline{g}}$ displayed in Proposition 2.

□

Proof of Proposition 4

Proof. Under homogeneous treatment effects, we have $CATT_{e,\ell} = CATT_{\ell}$ for all e . This simplifies Proposition 2 to

$$\sum_{\ell \in \mathcal{G}} \left(\sum_e \omega_{e,\ell}^g \right) CATT_{\ell} + \sum_{g' \neq g} \sum_{\ell \in \mathcal{G}'} \left(\sum_e \omega_{e,\ell}^g \right) CATT_{\ell} + \sum_{\ell \in \mathcal{G}^{excl}} \left(\sum_e \omega_{e,\ell}^g \right) CATT_{\ell}. \quad (58)$$

Recall 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 the bin indicators i.e. $\{\mathbf{1}\{t - E_i \in g\}\}_{g \in \mathcal{G}}$ and two-way fixed effects included in (3). Summing over e would imply the weight $\omega_{\ell}^g := \sum_e \omega_{e,\ell}^g$ is equal to the population regression coefficient on $\mathbf{1}\{t - E_i \in g\}$ from regressing $D_{i,t}^{\ell}$ 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 (3).

□

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

$$E\left[\frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s}) \cdot \mathbf{1}\{E_i = e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]}\right] = E\left[E\left[\frac{\mathbb{E}_N[(Y_{i,e+\ell} - Y_{i,s}) \cdot \mathbf{1}\{E_i = e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]} \mid E_i\right]\right] \quad (59)$$

$$= E\left[\frac{\mathbb{E}_N\left[E[(Y_{i,e+\ell} - Y_{i,s}) \cdot \mathbf{1}\{E_i = e\} \mid E_i]\right]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]}\right] \quad (60)$$

$$= E\left[\frac{\mathbb{E}_N\left[E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e] \cdot \mathbf{1}\{E_i = e\}\right]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]}\right] \quad (61)$$

$$= E\left[\frac{E[(Y_{i,e+\ell} - Y_{i,s}) \mid E_i = e] \cdot \mathbb{E}_N[\mathbf{1}\{E_i = e\}]}{\mathbb{E}_N[\mathbf{1}\{E_i = e\}]}\right] \quad (62)$$

$$= E[(Y_{i,e+\ell} - Y_{i,s}) \mid E_i = e]. \quad (63)$$

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

$$E[Y_{i,e+\ell} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+\ell} - Y_{i,s} \mid E_i \in C] \quad (64)$$

$$= E[Y_{i,e+\ell}^e - Y_{i,s}^e \mid E_i = e] - \sum_{c \in C} Pr\{E_i = c \mid E_i \in C\} E[Y_{i,e+\ell}^c - Y_{i,s}^c \mid E_i = c] \quad (65)$$

$$= E[Y_{i,e+\ell}^e - Y_{i,s}^\infty \mid E_i = e] - \sum_{c \in C} Pr\{E_i = c \mid E_i \in C\} E[Y_{i,e+\ell}^\infty - Y_{i,s}^\infty \mid E_i = c] \quad (66)$$

$$= E[Y_{i,e+\ell}^e - Y_{i,e+\ell}^\infty \mid E_i = e] + E[Y_{i,e+\ell}^\infty - Y_{i,s}^\infty \mid E_i = e] \quad (67)$$

$$- \sum_{c \in C} Pr\{E_i = c \mid E_i \in C\} E[Y_{i,e+\ell}^\infty - Y_{i,s}^\infty \mid E_i = c] \quad (68)$$

$$= E[Y_{i,e+\ell}^e - Y_{i,e+\ell}^\infty \mid E_i = e] + E[Y_{i,e+\ell}^\infty - Y_{i,s}^\infty] - E[Y_{i,e+\ell}^\infty - Y_{i,s}^\infty] \quad (69)$$

$$= E[Y_{i,e+\ell}^e - Y_{i,e+\ell}^\infty \mid E_i = e] \quad (70)$$

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

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.

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \neq \max\{E_i\}} \sum_{\ell \neq -1} \delta_{e,\ell} (\mathbf{1}\{E_i = e\} \cdot D_{i,t}^\ell) + \epsilon_{i,t} \quad (71)$$

$$= \alpha_i + \lambda_t + \mathbf{B}_{i,t}^\top \boldsymbol{\delta} + \epsilon_{i,t} \quad (72)$$

on observations from $t = 0, \dots, \max\{E_i\} - 1$ 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 $N_e := \sum_{i=1}^N \mathbf{1}\{E_i = e\}$ 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 $\text{vec}(A)$ 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}^ℓ 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

$$\widehat{\mathbf{f}}^0 = \begin{pmatrix} \frac{N_1}{N_1+N_2} & 0 \\ 0 & \frac{N_2}{N_1+N_2} \end{pmatrix} \quad (73)$$

and its probability limit is

$$\mathbf{f}^0 = \begin{pmatrix} Pr\{E_i = 1 | 1 \leq E_i \leq 2\} & 0 \\ 0 & Pr\{E_i = 2 | 1 \leq E_i \leq 2\} \end{pmatrix}. \quad (74)$$

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

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 (3), we can use

$$\widehat{v}_g := \frac{1}{|g|} \sum_{\ell \in g} \sum_{e \in h^\ell} \frac{N_e}{\sum_{e \in h^\ell} N_e} \widehat{\delta}_{e,\ell} = \frac{1}{|g|} \sum_{\ell \in g} \text{vec}(\widehat{\mathbf{f}}^\ell)^\top \widehat{\boldsymbol{\delta}}. \quad (75)$$

With a few standard assumptions (which we present together below as Assumption 4) on Regression (19), 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: $\{(E_i, \mathbf{Y}_i) : i = 1, 2, \dots, N\}$ are i.i.d. draws from their joint distribution where $\mathbf{Y}_i = (Y_{i,0}, \dots, Y_{i,T})^\top$ 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 $\ddot{\mathbf{B}}_{i,t}^\top$, double-demeaned version of $\mathbf{B}_{i,t}^\top$. 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 v_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*

$$\widehat{\nu}_\ell \rightarrow_p \sum_{e \in h^\ell} \Pr \{E_i = e \mid E_i \in h^\ell\} CATT_{e,\ell} = \text{vec}(\mathbf{f}^\ell)^\top \boldsymbol{\delta}. \quad (76)$$

The asymptotic distribution of this estimator is

$$\sqrt{N} \left(\widehat{\nu}_\ell - \text{vec}(\mathbf{f}^\ell)^\top \boldsymbol{\delta} \right) \rightarrow_d N \left(0, \boldsymbol{\delta}^\top \Sigma_{f^\ell} \boldsymbol{\delta} + \Sigma_\ell \right) \quad (77)$$

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

$$\mathbf{V}_{\mathbf{B}} = \sum_{t=0}^{\max\{E_i\}-1} E[\mathbf{B}_{i,t} \mathbf{B}_{i,t}^\top] \quad \Sigma_\ell = \text{vec}(\mathbf{f}^\ell)^\top \mathbf{V}_{\mathbf{B}}^{-1} \text{Var} \left(\sum_{t=0}^{\max\{E_i\}-1} \mathbf{B}_{i,t} \ddot{\epsilon}_{i,t}^2 \mathbf{B}_{i,t}^\top \right) \mathbf{V}_{\mathbf{B}}^{-1} \text{vec}(\mathbf{f}^\ell). \quad (78)$$

Proof of Proposition 6

Proof. We first show the asymptotic normality of the weights. The weights $\widehat{\mathbf{f}}^\ell$ are consistent since

$$\frac{N_e}{\sum_{e \in h^\ell} N_e} \rightarrow_p \frac{\Pr\{E_i=e\}}{\Pr\{E_i \in h^\ell\}} = \Pr\{E_i = e \mid E_i \in h^\ell\} \text{ by the Law of Large Numbers and Slutsky's theorem.}$$

Note that $\frac{N_e}{\sum_{e \in h^\ell} N_e}$ is also the regression coefficient estimator from the following cross-sectional regression

$$\mathbf{1}\{E_i = e\} = \beta \mathbf{1}\{E_i \in h^\ell\} + \eta_i(e) \quad (79)$$

with population regression coefficient equal to $\beta = \Pr\{E_i = e \mid E_i \in h^\ell\}$. Then by OLS asymptotics

which holds as E_i are iid by assumption and $\eta_i(e)$ is bounded, we have

$$\sqrt{N} \left(\frac{N_e}{\sum_{e \in h^\ell} N_e} - \Pr\{E_i = e \mid E_i \in h^\ell\} \right) \rightarrow_d N \left(0, \frac{E[\mathbf{1}\{E_i \in h^\ell\}^2 \eta_i^2(e)]}{E[\mathbf{1}\{E_i \in h^\ell\}^2]^2} \right). \quad (80)$$

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

$$\frac{E[\eta_i^2(e) \mid E_i \in h^\ell] \Pr\{E_i \in h^\ell\}}{\Pr\{E_i \in h^\ell\}^2} = \frac{E[\eta_i^2(e) \mid E_i \in h^\ell]}{\Pr\{E_i \in h^\ell\}}. \quad (81)$$

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 $E[\eta_i(e) \eta_i(e') \mid E_i \in h^\ell] / \Pr\{E_i \in h^\ell\}$. Thus, $\text{vec}(\widehat{\mathbf{f}}^\ell)$ has asymptotic distribution

$\sqrt{N} \left(\text{vec}(\widehat{\mathbf{f}}^\ell) - \text{vec}(\mathbf{f}^\ell) \right) \rightarrow_d N \left(0, \Sigma_{f^\ell} \right)$. Σ_{f^ℓ} is a matrix with diagonal entries equal to $\frac{E[\eta_i^2(e)|E_i \in h^\ell]}{\Pr\{E_i \in h^\ell\}}$, and off-diagonal entries equal to $\frac{E[\eta_i(e)\eta_i(e')|E_i \in h^\ell]}{\Pr\{E_i \in h^\ell\}}$.

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

$$\sqrt{N}(\widehat{\delta} - \delta) \rightarrow_d N \left(0, \mathbf{V}_{\ddot{\mathbf{B}}}^{-1} \text{Var} \left(\sum_{t=0}^{T-1} \ddot{\mathbf{B}}_{i,t} \ddot{\epsilon}_{i,t}^2 \ddot{\mathbf{B}}_{i,t}^\top \right) \mathbf{V}_{\ddot{\mathbf{B}}}^{-1} \right) \quad (82)$$

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

Lastly, by the delta method, we have

$$\sqrt{N} \left(\text{vec}(\widehat{\mathbf{f}}^\ell)^\top \widehat{\delta} - \text{vec}(\mathbf{f}^\ell)^\top \delta \right) \rightarrow_d N \left(0, \delta^\top \Sigma_{f^\ell} \delta + \Sigma_\ell \right) \quad (83)$$

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

$$\frac{\mathbf{V}_{\ddot{\mathbf{B}}}^{-1} \text{Cov}(\mathbf{1}\{E_i \in h^\ell\} \eta_i(e), \ddot{\mathbf{B}}_{i,t} \ddot{\epsilon}_{i,t})}{E[\mathbf{1}\{E_i \in h^\ell\}^2]^2} = \frac{\mathbf{V}_{\ddot{\mathbf{B}}}^{-1} E[\ddot{\mathbf{B}} \eta_i(e) \ddot{\epsilon}_{i,t} | E_i \in h^\ell]}{E[\mathbf{1}\{E_i \in h^\ell\}^2]^2}. \quad (84)$$

Since $\ddot{\mathbf{B}}$ and $\eta_i(e)$ are functions of E_i , we have $E[\ddot{\mathbf{B}} \eta_i(e) \ddot{\epsilon}_{i,t} | E_i \in h^\ell] = E[\ddot{\mathbf{B}} \eta_i(e) E[\ddot{\epsilon}_{i,t} | E_i, E_i \in h^\ell]]$.

Furthermore, specification (19) is saturated in E_i and relative time so $E[\ddot{\epsilon}_{i,t} | E_i] = 0$. This proves Proposition 6. \square

C Simulation design

We generate 1000 simulated datasets with $N = 1000$ and $T = 3$ according to the following DGP

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e=1}^3 \sum_{\ell=1-e}^{T-e} \delta_{e,\ell} \mathbf{1}\{E_i = e\} \cdot D_{i,t}^\ell + \epsilon_{i,t} \quad (85)$$

where the individual fixed effects and time fixed effects are set to the id numbers $\alpha_i = i \in \{1, \dots, N\}$ and $\lambda_t = t \in \{0, \dots, T\}$. For each simulation, we draw E_i uniformly from $\{1, 2, 3\}$. We analyze the case where the DGP is a model of dynamic and non-stationary treatment effects. In particular, we set $\delta_{1,0} = 2$, $\delta_{1,1} = 18$, $\delta_{1,2} = 19$, $\delta_{2,0} = 3$, $\delta_{2,1} = 4$, $\delta_{3,0} = 4$ and $\epsilon_{i,t} \sim N(0, 1)$. First, we set $\delta_{e,\ell < 0} =$

0 $\forall e$ for the case of no anticipatory behavior. For each simulation, we estimate the following FE specification

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\ell=-2,0,1,2} \mu_\ell D_{i,t}^\ell + \epsilon_{i,t} \quad (86)$$

for $t = 0, \dots, 3$.