Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects

Sarah Abraham*
Department of Economics, MIT
and
Liyang Sun
Department of Economics, MIT (lsun20@mit.edu)

September 6, 2018

Abstract

In event studies where the timing of treatment varies across units, researchers often estimate dynamic treatment effects using fixed effects regressions. In this paper we cast event studies in a potential outcomes framework, and formulate assumptions that give the estimands of these regressions a direct causal interpretation as the population average treatment effect on the treated. However when the homogeneity assumptions do not hold and treatment effects vary across cohorts and over time, estimands of these regressions may not be causally interpretable. In fixed effect models where the sole regressor is treatment status, the OLS coefficient is a non-convex average of the cohort-specific treatment effects. When regressors containing lags and leads of treatment are added, the OLS coefficient corresponding to a given lead or lag picks up spurious terms consisting of treatment effects from other periods. In particular, this makes Granger causality test in which coefficients on leads provide evidence for lack of or existence of pre-trends invalid. We propose alternative estimators that are guaranteed to identify convex averages of the cohort-specific treatment effects under heterogeneity.

Keywords: Differences-in-Differences, Fixed Effects, Panel Data Model

*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, Kevin Kainan Li, Whitney Newey, and James Poterba for helpful 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/.
1 Introduction

Rich panel data has fueled a growing literature estimating dynamic treatment effects. In this context, “event studies” are an increasingly common extension of differences-in-differences (DID) designs for estimating the dynamic effects of discrete shocks and non-transient treatments such as the effect of a hospitalization on individuals over time (Dobkin et al., 2018). Event studies exploit variation in the timing of a unit’s treatment, which allows for identification of the path of treatment effects even when pure control units (i.e. never treated units) are not observed. To estimate the treatment effects in an event study framework, researchers often use linear fixed effects (FE) regressions (Chetty et al. 2014; Finkelstein et al. 2017; Kline et al. 2017; Kleven et al. 2018; Zwick 2018).

In this paper, we inform the estimation of event studies through three contributions. First, we cast event studies in a potential outcomes framework and discipline the possible forms of treatment effect heterogeneity. Then, we use the potential outcomes framework to describe the behavior of linear two-way FE regressions including unit and time fixed effects, when used to estimate treatment effects in an event study framework.¹ We show that when strict homogeneity assumptions do not hold, the FE estimands can be non-convex averages of heterogeneous treatment effects, and thus not causally interpretable. Lastly, motivated by these negative results, we propose alternative estimators that are guaranteed to estimate convex averages of treatment effects even under the most general form of treatment effect heterogeneity. We close with an empirical example that illustrates both how and why FE estimators can be misleading in comparison to our suggested alternative estimators.

We begin by providing a potential outcomes framework for event studies to clarify three things: the causal parameters of interest, how they are identified, and the relevant concepts of treatment effect homogeneity. Our focus is the set of event studies where units select into a binary treatment over multiple periods with unknown underlying time trends and treatment is an absorbing state. We can subsequently categorize treated units into “cohorts” based on when they first receive the treatment. For each cohort, we can define relative time as the time relative to the initial treatment.

¹We find qualitatively similar results for another set of common regressions for event studies with only time fixed effects. For details, see Online Appendix C.
The defining feature of event studies is the presence of multiple cohorts, which serves as the source of identification for estimation and therefore also governs the set of causal parameters that can be recovered. Within a cohort, calendar time and relative time are collinear, so we cannot separately identify dynamic treatment effects in the presence of time trends. However, with multiple cohorts we can separate these two sets of effects by comparing the trends in average outcomes between treated cohorts and cohorts yet to be treated if we are willing to assume parallel trends and no anticipatory behavior. For any given cohort, the differences in trends identify the average treatment effects for this cohort at different relative times. We call these cohort-specific average treatment effects on the treated (CATTs). By extension, any convex average of CATTs is also a causally interpretable estimate of the average treatment effects on the treated.

Cohorts are the level of variation used to identify calendar time trends separately from dynamic treatment effects, therefore they are also the key level of concern for confounding heterogeneity. Using the potential outcomes framework, we define the two relevant dimensions of heterogeneity in CATTs: variation within cohorts across relative time (non-stationarity) and within relative time across cohorts (heterogeneity across cohorts). We use these dimensions of heterogeneity to clarify what assumptions are required for the FE estimands to be causally interpretable.

There are two common specifications of linear two-way FE regressions: static and dynamic. In the static specification, researchers estimate a single discrete effect by including a single indicator for a unit being treated at time \( t \). We find that this FE estimator estimates a linear combination of CATTs across all cohorts and relative times where the weights add up to one but can be negative. When treatment effects are heterogeneous or non-stationary, this estimand need not correspond to an average treatment effect for any group of cohorts and may even be outside the convex hull of the set of CATTs.

In the dynamic specification, researchers include leads and lags of the treatment indicator as regressors. Researchers may use a dynamic specification because they believe the treatment effects change over time or want to test the validity of the simpler static specification.\(^2\) When treatment effects are heterogeneous, we find the dynamic specification is similar to the Granger causality model popular in DID designs (Autor, 2003). See Deshpande and Li (2017) for an example of interpreting coefficients on the leads as pre-trends in event

\(^2\)This specification is similar to the Granger causality model popular in DID designs (Autor, 2003). See Deshpande and Li (2017) for an example of interpreting coefficients on the leads as pre-trends in event
tion also does not return causally interpretable estimands. The estimand associated with a particular lead (or lag) $l$ can be a non-convex average of $CATT$s from all periods, not just $l$. Thus the interpretation of this estimand as an average treatment effect $l$ periods since initial treatment is confounded by two factors: the inclusion of spurious terms reflecting treatment effects from periods other than $l$ and the possibility of negative weights on some $CATT$s. This cautions against interpreting the lead coefficient estimates as indicative of underlying pre-trends, and the lag coefficient estimates as the dynamic treatment effects. In some cases these two factors are sufficient to produce estimates of the opposite sign to the true underlying $CATT$s, or to indicate pre-trends when in fact none exists. To our knowledge, these negative results regarding the dynamic specification have not been documented.

We propose alternative estimators that are guaranteed to estimate a causally interpretable convex average of $CATT$s even under treatment effect heterogeneity. Under the parallel trends and no anticipatory behavior assumptions we can estimate the full set of $CATT$s from an interacted specification that is saturated in relative time and cohort indicators. We can then aggregate these using any set of well behaved weights to obtain a causally interpretable average treatment effect. In particular, as an alternative to the dynamic FE specification, we can form a weighted average of the $CATT$ estimates associated with each relative time with weights equal to the share of each cohort. These alternative estimates are thus causally interpretable as estimates for the average treatment effect in a given relative time representative of the underlying cohort shares.

We demonstrate the empirical relevance of our negative findings regarding the dynamic FE specification by re-estimating the economic consequences of hospitalization seen in Dobkin et al. (2018) for an elderly sample. We estimate the effect of hospitalization on the out-of-pocket medical spending and labor earnings on an elderly sample of the Health and Retirement Survey (HRS). This provides a context where we expect heterogeneous treatment effects over time because individuals gradually age into Medicare and retirement. Using our alternative estimator with weights equal to the share of each cohort, we find that hospitalization increases out-of-pocket medical spending and decreases labor earnings. In
contrast, we show that the FE estimator gives results of the opposite sign, suggesting hospitalization decreases medical spending and increases labor earnings. This example illustrates how FE estimates can yield misleading results in the presence of heterogeneous treatment effects.

This paper complements several contemporaneous papers that also analyze event studies (Athey and Imbens (2018); Borusyak and Jaravel, 2017; Callaway and Sant’Anna, 2018). Our analysis also contributes to a recent literature analyzing linear two-way fixed effects regressions applied to a setting with units selecting into binary treatment over time (de Chaisemartin and D’Haultfœuille, 2018; Imai and Kim, 2017). While these papers focus on the implications of treatment effect heterogeneity for the static FE specification, our analysis also covers the popular dynamic FE specification, where we find additional reasons for concern in causally interpreting FE estimates. We compare our results with these papers in more detail after presenting the static FE specification in Section 3.2.1.

In the next section, we introduce the potential outcomes framework for event studies. Section 3 derives the estimands of linear two-way FE regressions and presents the negative results on them. Section 4 develops our alternative estimators. Section 5 discusses applications of our results and Section 6 concludes. All proofs are contained in the Online Appendix, along with the extension to alternative identifying assumptions and variations of linear two-way FE regressions; in Online Appendix F we allow for covariates. In addition to the inverse propensity score reweighted estimator that is also independently proposed by Callaway and Sant’Anna (2018), we propose a doubly robust score that can be used to estimate dynamic treatment effects as an extension to Chernozhukov et al. (2017).

## 2 Event studies in a potential outcomes framework

We assume a random sample of $N$ independently drawn units from an infinitely large population. For each unit $i = 1, \ldots, N$ at time $t = 0, \ldots, T$ where $T$ is fixed, we observe an outcome variable $Y_{i,t}$ and a binary treatment status variable $D_{i,t} \in \{0,1\}$: $D_{i,t} = 0$ if $i$ has not been treated by period $t$ and $D_{i,t} = 1$ if $i$ has been treated by period $t$.

In event studies, treatment is an absorbing state and the treatment path $\{D_{i,t}\}_{t=0}^{T}$ is a non-decreasing sequence of zeros and then ones, i.e. $D_{i,s} \leq D_{i,t}$ for $s < t$. Thus, the
treatment path of a unit can be uniquely characterized by the time period of the initial
treatment, denoted by the scalar random variable \( E_i = \min \{ t : D_{i,t} = 1 \} \). We call this
random variable \( E_i \) “event time” as it refers to the onset of the treatment. In our notation,
cohort \( e \) is the set of units for which \( E_i = e \) and the treatment path is uniquely defined. We
denote \( E_i = \infty \) for units never treated. This set up is similar to that in Athey and Imbens
(2018) where their “adoption date” corresponds to our “event time”, but they focus on a
randomization-based perspective that parallels our sampling-based perspective.

For our main results, we consider the general case where \( E_i \) is supported on \( \{0, 1, 2, 3, \ldots, T\} \),
so that there can be units selecting into treatment in every period and all units are treated
by the last period \( (t = T) \). This case corresponds to the scenario when researchers only
observe or choose to focus on units that are ever treated in a fixed time frame. Furthermore,
for units that receive the treatment first \( (E_i = 0) \), researchers do not observe their
previous outcomes when not treated. A common special case used by researchers normal-
izes calendar time so that no units are treated in the first period \( (t = 0) \). This case can
be easily extrapolated from our more general case by excluding the first cohort so that no
units are treated at \( t = 0 \). Another extension of our case is one where we observe units
never treated in the panel (possibly treated after the panel \( E_i > T \), or a pure control group
that is never treated \( E_i = \infty \) ), i.e. \( \sum_{e=0}^{T} Pr \{ E_i = e \} < 1 \). We do not focus on this case for
our main results on FE regressions, as many event studies restrict their analysis to units
that ever receive the treatment (often because there are concerns that the never-treated
group is not a representative control for treated groups). However, our negative results on
FE regressions are not affected by the inclusion of a never-treated group and the proof can
be easily modified to accommodate this case, as we show in Online Appendix A.

Potential outcomes describe realizations of the outcome variable \( Y_{i,t} \) that arise in re-
response to a hypothetical treatment path \( e \). We denote the potential outcome in period
t under treatment path \( e \) by \( Y^e_{i,t} \). We denote the potential outcome if unit \( i \) never re-
ceives treatment as \( Y^\infty_{i,t} \), which we call the “baseline outcome”. A “treatment effect” is
therefore defined as the difference between the baseline outcome and the potential out-
come: \( Y^e_{i,t} - Y^\infty_{i,t} \). We observe each unit under only a single treatment path \( E_i = e \), so
the observed outcome for unit \( i \) is simply \( Y_{i,t} = Y^e_{i,t} \). To clarify the link between observed
outcomes, potential outcomes, and treatment effects, we can write this realized outcome as a sum of the baseline outcome and the treatment path specific treatment effects:

\[ Y_{i,t} = Y_{i,t}^{\infty} + \sum_{0 \leq e \leq T} (Y_{i,t}^{e} - Y_{i,t}^{\infty}) \cdot 1 \{ E_{i} = e \}. \]  

(1)

For a given cohort \( e \) the treatment effects of interest are the cohort-specific average treatment effects on the treated in time period \( t \), or \( E[Y_{i,t}^{e} - Y_{i,t}^{\infty} \mid E_{i} = e] \). For comparison across cohorts in an event study framework, we shift from calendar time index \( t \) to relative time index \( l \) which denotes the periods since treatment; for cohort \( e \), \( l \) ranges from \(-e\) to \( T - e \) because at most we observe \( e \) periods before initial treatment and \( T - e \) periods after initial treatment. Thus, the central set of causal parameters for our paper are the quantities \( CATT_{e,l} \) defined below.

**Definition 1.** We define the cohort-specific average treatment effects on the treated \( l \) periods from initial treatment as \( CATT_{e,l} = E[Y_{i,e+l}^{e} - Y_{i,e+l}^{\infty} \mid E_{i} = e] \).

Each \( CATT_{e,l} \) represents the average treatment effect \( l \) periods from initial treatment for the cohort of units first treated at time \( e \). Next we turn to the two assumptions necessary to identify \( CATT_{e,l} \) (see Proposition 4).

### 2.1 Identifying assumptions

Two assumptions are sufficient for identification of \( CATT_{e,l} \): a parallel trends assumption and a no anticipatory behavior assumption. For our parallel trends assumption, we assume that growth in baseline outcomes is mean independent of the event time \( E_{i} \).

**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 \( E[Y_{i,t}^{\infty} - Y_{i,s}^{\infty}] \).

One could substitute this assumption with a stronger identifying assumption that baseline outcomes are mean independent of the event time i.e. at each \( t \), \( E[Y_{i,t}^{\infty} \mid E_{i} = e] \) 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. While this stronger assumption does not affect our results on two-way (unit and time) fixed effects regressions, it allows
identification in regressions with only time fixed effects. See Online Appendix C for more
details on regressions with only time fixed effects under the assumption of random timing.

The no anticipatory behavior assumption precludes treatment from influencing out-
comes before treatment takes place. If this assumption is violated and there is anticipatory
behavior, then trends in observed outcomes of future treated cohorts do not identify the
trends in baseline outcomes but rather a combination of trends in baseline outcomes and
leading treatment effects (i.e. anticipatory behavior).

**Assumption 2.** (No anticipatory behavior.) $Y_{i,t}^e = Y_{i,t}^\infty$ for all $t < e$. Equivalently, $Y_{i,e+l}^e = Y_{i,e+l}^\infty$ for all $l < 0$. This implies no pre-trends since $CATT_{e,l} = 0$ for all $e \in \text{supp}(E_i)$ and all $l < 0$.

This assumption requires potential outcomes before treatment to be equal to the base-
line outcome. This is most plausible if the full treatment paths are not known to units
because 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 repre-
sent baseline outcomes. For example, Hendren (2017) shows that knowledge of future job
loss leads to decreases in consumption. Formally, let $Y_{i,t}$ be consumption and $D_{i,t}$ be job
loss. Future job loss $D_{i,1} = 1$ may lower current level of consumption $Y_{i,0}^1$ and we cannot
assume $Y_{i,0}^1$ equals $Y_{i,0}^\infty$.

We maintain both Assumption 1 and Assumption 2 throughout our analysis unless
otherwise noted.

### 2.2 Possible forms of treatment effect heterogeneity

We now return to the set of causal parameters, $CATT_{e,l}$, to define and discuss possible
forms of treatment effect heterogeneity. As the notation shows, $CATT_{e,l}$ are cohort- and
lag-specific average treatment effects. Therefore in an event study (which includes multiple
cohorts and lags) heterogeneity in $CATT_{e,l}$ can arise among each of these two dimensions:
within a given cohort across lags and within a given lag across cohorts. (By starting
from $CATT_{e,l}$, an average treatment effect, we are allowing for heterogeneity across units.
Heterogeneity across units is only relevant in so far as it generates variation in $CATT_{e,l}$
across cohorts or lags, as we discuss below.)
2.2.1 Definitions of heterogeneity

First, we define stationary and non-stationary treatment effects, which describe heterogeneity within a cohort in the relative time dimension $l$.

**Definition 2.** (Stationarity and non-stationarity). If $CATT_{e,l} = CATT_{e,l'}$ for all lags of treatment $l, l' \geq 0$, then we say treatments effects for cohort $e$ are **stationary**; in contrast, if $CATT_{e,l} \neq CATT_{e,l'}$ for any lag of treatment $l, l' \geq 0$, then we say treatment effects for cohort $e$ are **non-stationary**.

Stationary effects impact units immediately upon treatment and on average persist at the same level for all treated periods. These cases are likely from permanent changes (e.g. the effect of a tax rate change on labor supply) but could also be generated by a one-time shock with constant and persistent effects. Treatment effects are likely to be non-stationary when there is learning from and adaptation to the treatment over time, such as workers switching to a new firm or recovery from a transient shocks like an adverse health shock, as we will explore in our application.

Second, we define cross-cohort homogeneity and heterogeneity in treatment effects to describe the heterogeneity within a relative time $l$ across cohorts.

**Definition 3.** (Cross-cohort homogeneity and heterogeneity). For each lag of treatment $l \geq 0$, if $CATT_{e,l}$ does not depend on cohort $e$, then we say treatment effects are **homogenous across cohorts**; in contrast, if for any $l \geq 0$, $CATT_{e,l}$ varies by $e$, then we say treatment effects are **heterogeneous across cohorts**.

Cross-cohort homogeneity means that each cohort experiences the same profile of treatment effects on average and $CATT_{e,l}$ at any given lag $l$ is the same across cohorts. Treatment effects need to be the same across cohorts in every lag for homogeneity to hold, whereas for heterogeneity to occur, treatment effects just need to differ across cohorts in one lag.

In event studies, heterogeneity across cohorts could arise for many reasons. Cohorts may differ in their covariates, which affect how they respond to treatment. For a concrete example which we will explore further in our application, if treatment effects differ with age then when age is correlated with cohort we will have heterogeneous effects (see Section
5 for details). After controlling for covariates, cohorts may still vary in their response to the treatment if units select treatment timing based on treatment effects. This case corresponds to essential heterogeneity as discussed in Heckman et al. (2006). 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).

2.2.2 Possible homogeneity assumptions

We now translate the two definitions governing heterogeneity into two corresponding assumptions on homogeneity in $CATT_{e,l}$ that we might impose for estimation. As we show later, these additional assumptions are sufficient for the FE estimators to yield causally interpretable estimates.

**Assumption 3.** (Stationary treatment effects for all cohorts). For any given cohort $e$, treatment effects are stationary, i.e. $CATT_{e,l} = CATT_{e,l'}$ for all different lags of treatment $l, l' \geq 0$.

**Assumption 4.** (Treatment effects homogeneity). For each lag of treatment $l \geq 0$, $CATT_{e,l}$ does not depend on cohort $e$.

Note that neither of Assumptions 3 and 4 is implied by the other. If both assumptions hold, then all cohorts experience the same constant treatment effect on average. If neither assumption holds then there is at least one lag where not all cohorts experience the same average treatment effect. If Assumption 3 does not hold but Assumption 4 does (treatment effects are non-stationary but not homogenous across cohorts) then all cohorts share the same non-constant treatment path on average. The final and perhaps least intuitive case is one where treatment effects are heterogeneous across cohorts but still stationary for all cohorts (Assumption 3 holds but Assumption 4 does not). This may occur when the treatment intensity varies across cohorts but the effect is constant within each cohort. For example, a job training program might bring a one-time permanent wage increase to participants, but the magnitude of the increase experienced by an individual depends on when he participates.
3 Estimators from linear two-way FE regression

3.1 Two specifications: static and dynamic

Treatment effects in event studies are often estimated by a linear regression with two-way (unit and time) fixed effects. There are two common specifications: static and dynamic. The static specification is

\[ Y_{i,t} = \alpha_i + \lambda_t + \gamma D_{i,t} + \xi_{i,t} \quad (2) \]

where \( D_{i,t} \) is the treatment indicator in time period \( t \). The other common specification, which we refer to as the dynamic specification, adds leads and lags of treatment to the above specification:

\[ Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-T}^{T} \mu_l D_{i,t}^l + \nu_{i,t} \quad (3) \]

where \( D_{i,t}^l \) is an indicator for being \( l \) time periods relative to \( i \)'s initial treatment (\( l = 0 \) is the year of initial treatment). Even though we sum relative time \( l \) over its possible range, we actually need to exclude at least two relative time indicators due to multicollinearities. One multicollinearity arises from the linear dependence between unit (or time) fixed effects and relative time indicators. Another multicollinearity arises from the linear relationship between cohort, calendar time and relative time. See Borusyak and Jaravel (2017) for more details. In practice, researchers often exclude more than two relative times from (3). For example, they may exclude all leads greater than \( K \) and lags greater than \( L \).\(^3\) In Online Appendix A.1, we show that as long as \( CATT_{e,l} \) associated with the excluded relative times are all zero, our main results are unaffected. Throughout this section we will maintain this assumption and leave the case where \( CATT_{e,l} \neq 0 \) for excluded relative times to the Online Appendix.

To estimate the effect of the treatment on outcome growth, researchers have also considered a first-differenced version of specifications (2) and (3).\(^4\) We defer a discussion of the first-differenced variation to Online Appendix B since the results are qualitatively similar.

\(^3\)Excluding leads can be justified by the no anticipatory behavior assumption (Assumption 2); excluding lags can be justified if treatment effects are zero for the excluded lags.

\(^4\)For empirical examples of these specifications, see Parker et al. (2013) and Hendren (2017).
3.2 Estimators

Throughout this section we assume a balanced panel of $N$ independently drawn units and $T + 1$ time periods where $T$ is fixed. We denote by $\gamma$ and $\mu_l$ the estimands of specifications (2) and (3), i.e. the population regression coefficients. Their corresponding estimators are denoted by $\hat{\gamma}$ and $\hat{\mu}_l$ respectively. Below we show that while these FE estimators recover linear combinations of causal parameters $CATT_{e,l}$, the weights are in general unreasonable and need not be positive without additional assumptions that restrict treatment effects to be stationary or homogeneous.

3.2.1 Static specification estimators

**Proposition 1.** (Static specification estimand). Denote $\bar{D}_{e,\cdot} = Pr\{E_i \leq e\} = (T - e)/(T + 1)$ and $\bar{D} = (T - E[E_i])/(T + 1)$. Under Assumption 1 and 2, the population regression coefficient $\gamma$ from specification (2) is

$$\gamma = \sum_{e=0}^{T} \sum_{l=0}^{T-e} \omega_{e,l} CATT_{e,l}$$

where the weights are

$$\omega_{e,l} := \frac{Pr\{E_i = e\} \cdot (1 - \bar{D}_{e+1} - \bar{D}_{e} + \bar{D})}{\sum_{t=0}^{T} E[\bar{D}_{i,t}^2]}$$

and these weights sum to one.

Proposition 1 reveals that, while the static FE estimand is a linear combination of $CATT_{e,l}$, the weights generally do not coincide with sample frequencies and need not be non-negative. The numerator of the weight, $1 - \bar{D}_{e+1} - \bar{D}_{e} + \bar{D}$, is the residual from predicting treatment status $D_{i,t}$ with unit and time fixed effects. It thus downweights (to the point of potentially negatively weighting) the long-run treatment effects (large $l$) for cohorts with an early onset of treatment (small $e$) because the fixed effects would overpredict their treatment probability.

As a concrete example, if all units are treated at some time within the panel, some weights will necessarily be negative. Since all units have been treated by the final period, we have $\bar{D}_{\cdot,T} = 1$. Furthermore, $\bar{D}_{e,\cdot} > \bar{D}$ for cohorts treated early in the panel ($e < E[E_i]$).
The weights associated with $CATT_{e,T-e}$ for these cohorts in the last period are all negative, therefore $\gamma$ is a non-convex average of $CATT_{e,l}$. It would be further from a convex average when the $CATT_{e,l}$ associated with negative weights are of large magnitude. In some cases $\gamma$ may be outside the convex hull of all $CATT_{e,l}$ and $\gamma$ can even have the opposite sign from all $CATT_{e,l}$, as we will demonstrate in our empirical application.

**Special Case: Stationary and Homogeneous Treatment Effects.** From Proposition 1, we can also see that when treatment effects are stationary for all cohorts (Assumption 3) and homogeneous (Assumption 4), the static FE estimand $\gamma$ will be the causal effect of interest. With these strict assumptions, $CATT_{e,l}$ does not depend on $e$ or $l$ so we can move the constant $CATT$ outside of the summation and the weights sum to one; thus the population regression coefficient $\gamma$ recovers the constant $CATT$.

**Comparison to the literature.** The above analysis contributes to a recent literature analyzing linear two-way fixed effects regressions (de Chaisemartin and D’Haultfœuille, 2018; Imai and Kim, 2017). We show the similar result that the FE estimand is a linear combination of average treatment effects in each cohort, however due to the structure of event studies, we are able to relax their assumptions on treatment effects. In particular, in contrast to de Chaisemartin and D’Haultfœuille (2018), we allow for non-stationary treatment effects, meaning treatment effects can vary with time since treatment for each cohort. Relatedly, we also allow for lasting treatment effects, unlike Imai and Kim (2017) where by assumption the treatment has no lasting effect beyond the period of initial treatment.

We also extend the analysis of Borusyak and Jaravel (2017) for static specification event studies to allow for treatment effect heterogeneity in addition to non-stationary treatment effects. Finally, while Callaway and Sant’Anna (2018) allude to possible issues with equation (2), their focus is to develop an inverse propensity score reweighted estimator for $CATT_{e,l}$. We compare their estimator to ours in Online Appendix F when we discuss the extension of our results to allow for covariates.

### 3.2.2 Dynamic specification estimators

Next we analyze what dynamic FE estimators estimate without restrictions on treatment effects. To our knowledge, no paper has yet studied the dynamic specification when treat-
ment effects are heterogenous. We find that the FE estimands for $l$ lags to treatment can be non-convex averages of $CATT_{e,l}$ as well as $CATT_{e,l'}$ from other periods. Incorrectly assuming homogenous effects can yield an estimate with the opposite sign to all $CATT$ from exclusively $l$ periods after initial treatment or yield non-zero pre-trend estimates even when there are no pre-trends. We state these results more formally in two corresponding propositions below.

**Proposition 2.** Under Assumption 1 (parallel trends) only, the population regression coefficient on the indicator for $l$ periods from treatment is a linear combination of $CATT_{e,l}$ as well as $CATT_{e,l'}$ for $l' \neq l$,

\[
\mu_l = \sum_{0 \leq e \leq T-l} \omega^l_{e,l} CATT_{e,l} + \sum_{l' \neq l} \sum_{0 \leq e \leq T-l'} \omega^l_{e,l'} CATT_{e,l'}.
\] (4)

The weights on $CATT_{e,l}$ sum up to one $\sum_{-l \leq e \leq T-l} \omega^l_{e,l} = 1$. The weights on $CATT_{e,l'}$ from each other relative times $l' \neq l$ sum up to zero $\sum_{-l' \leq e \leq T-l'} \omega^l_{e,l'} = 0$. The weight $\omega^l_{e,l'}$ is the population regression coefficient of $D_{i,t}^l$ from regressing $1\{E_i = e\} \cdot D_{i,t}^{l'}$ on the relative time indicators included in (3) i.e. $\{D_{i,t}^{l'}\}_{T_t = -T}$ and two-way fixed effects.

Proposition 2 shows that the FE estimand for the lagged treatment effect, $\mu_l$ where $l \geq 0$, averages over treatment effects across cohorts from multiple time periods: $CATT_{e,l}$ as well as $CATT_{e,l'}$ for $l' \neq l$. Therefore, this estimand $\mu_l$ is not isolating the average treatment effects $l$ periods after initial treatment, but rather estimating a linear combination of effects from multiple periods; additionally, although these weights sum to one, the weights need not be non-negative.

**Proposition 3.** If Assumptions 1 (parallel trends) and 2 (no anticipatory behavior) both hold, the population regression coefficient on the indicator for $l$ leads prior to the treatment is a linear combination of post-treatment $CATT_{e,l'}$ for all $l' \geq 0$

\[
\mu_l = \sum_{l' \geq 0} \sum_{0 \leq e \leq T-l'} \omega^l_{e,l'} CATT_{e,l'},
\] (5)

where for each $l'$, the weights sum up to zero $\sum_{-l' \leq e \leq T-l'} \omega^l_{e,l'} = 0$.

Proposition 3 shows that the FE estimand for pre-trends, $\mu_l$ where $l < 0$, is not zero even when there are no pre-trends by assumption. Under the no anticipatory behavior
assumption, cohort-specific treatment effects prior to treatment are all zero: $CATT_{e,l} = 0$ for all $l < 0$. Any linear combination of these $CATT_{e,l}$ is zero. However, due to the influence of post-treatment $CATT_{e,l'} \geq 0$, this estimand $\mu_l$ is not necessarily zero. Similarly, when the no anticipatory behavior assumption does not hold, the FE estimand $\mu_l$ averages over pre-treatment effects and post-treatment effects; thus it is possible for them to cancel each other out and we would not be able to detect pre-trends with the lead FE estimates $\hat{\mu}_l$. When effects are not homogenous across cohorts, it is problematic to interpret a non-zero estimates on $D_{i,t}^l$ for $l < 0$ as evidence for pre-trends because non-zero post-treatment $CATT_{e,l}$ can affect the lead FE estimators. In Online Appendix D we present through simulation two examples where FE estimates do not appropriately summarize underlying pre-trends.

**Special Case: Homogeneous Treatment Effects.** In the special case where effects are homogenous across cohorts, i.e. $CATT_{e,l}$ is constant across $e$ for a given $l$. In such a case, the lagged FE estimand $\mu_l$ for $l \geq 0$ is equal to $CATT_{e,l}$ and the FE dynamic estimators do recover causally interpretable estimates. If we further assume no anticipation, the lead FE estimand $\mu_l$ for $l < 0$ is equal to zero.

4 Alternative estimation method

We propose a new specification for estimating treatment effects in event studies. Unlike the FE estimators, the estimands of this alternative method are weighted averages of $CATT_{e,l}$ with reasonable weights (i.e. weights that sum to one and are non-negative) even under heterogeneous effects.

Recall that without homogeneity across cohorts $CATT_{e,l}$ vary by $e$ so the error term $\nu_{i,t}$ involves the difference between $CATT_{e,l}$ and $CATT_{e',l}$ for different cohorts $e \neq e'$. The error term is thus correlated with treatment timing $E_i$, which is captured in the unit fixed effects. This observation hints at using an interacted specification saturated in relative time indicators $D_{i,t}^l$ and cohort indicators $1\{E_i = e\}$ to estimate each of the $CATT_{e,l}$; the resulting estimates for $CATT_{e,l}$ are averaged to provide the final estimate. Following Gibbons et al. (2018), we call these estimators “interaction-weighted” (IW) estimators. In what follows we begin by defining a general DID estimator for $CATT_{e,l}$ as a building block.
for the interaction-weighted estimators and then proceed to describe the construction and properties of our alternative estimator.

4.1 Differences-in-differences for \( CATT_{e,l} \)

We use the notion \( E_N \) to abbreviate the symbol \( \frac{1}{N} \sum_{i=1}^{N} \). We define the DID estimator for a \( CATT_{e,l} \) as

\[
\frac{E_N[(Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i = e\}]}{E_N[1\{E_i = e\}]} - \frac{E_N[(Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i \in C\}]}{E_N[1\{E_i \in C\}]} \tag{6}
\]

for some \( s < e \) and \( C \subseteq \{c : e + l < c \leq T\} \) where \( C \) is the set of cohorts used as control cohorts. We assume \( \sum_{i=1}^{N} 1\{E_i = e\} \) and \( \sum_{i=1}^{N} 1\{E_i \in C\} \) are non-zero so that this estimator is well-defined. If we do not observe any unit in cohort \( e \) or \( C \) is empty, we cannot estimate the corresponding \( CATT_{e,l} \) using a DID estimator.

The DID is the difference between the average change in outcomes for cohort \( e \), which is exactly \( l \) periods relative to treatment, and average change in outcomes for cohorts that have not been treated by \( t = e + l \). When well-defined, it is an unbiased and consistent estimator for \( CATT_{e,l} \), a fact that we build on in deriving the probability limit of the IW estimator. We state this in the following proposition.

**Proposition 4.** Assume cohort \( e \) and the control cohorts are non-empty so that the DID estimator for \( CATT_{e,l} \) is well-defined. Under Assumption 1 and 2, the DID estimator is an unbiased and consistent estimator for \( CATT_{e,l} \).

Note that under Assumption 2 (no anticipatory behavior), Assumption 1 (parallel trends) has many testable implications. There are multiple DID estimators for \( CATT_{e,l} \) since we can set different \( s < e \) and take subsets of \( \{c : e + l < c \leq T\} \) as the control cohorts. One can thus form an omnibus test for the validity of these two assumptions. See Callaway and Sant’Anna (2018) for one such test.

It is also possible to relax the parallel trends assumption to allow the timing of treatment to depend on covariates. We state the conditional parallel trends assumption in Online Appendix F, and present the doubly robust scores for estimating \( CATT_{e,l} \) consistently and efficiently as an extension to the inverse propensity score reweighted estimator proposed by Abadie (2005) and Callaway and Sant’Anna (2018).
4.2 Interaction-weighted estimators

Interaction-weighted (IW) estimators average estimates for $CATT_{e,t}$ with reasonable weights to form causally interpretable average treatment effect estimators. Below we describe the estimation procedure in three steps (with more detailed definitions stated in Definition 4 of Online Appendix A for the static and dynamic estimators). First, we estimate $CATT_{e,t}$ using a linear two-way fixed effects specification that interacts relative time indicators with cohort indicators:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e=1}^{T-1} \sum_{l=1}^{T-e} \delta_{e,l}(1\{E_i = e\} \cdot D_{i,t}^l) + \epsilon_{i,t}$$  \hspace{1cm} (7)

on $t = 0, \ldots, T - 1$ and $e = 1, \ldots, T$. (Among regressors we exclude interactions with $1\{E_i = T\}$ and for each $e = 1, \ldots, T - 1$, we exclude interactions with $D_{i,t}^e$. We need to drop time period $T$ because everyone is treated in the last period. The DID estimators for $CATT_{e,T}$ are thus not well-defined as $C = \emptyset$. We need to exclude cohort 0 from the sample because for cohort 0 we do not observe its units when not treated. The DID estimators for $CATT_{0,t}$ are thus not well-defined.) Second, we estimate a set of appropriate weights that is the sample share of each cohort in the relevant period(s). Third and finally, we take weighted averages of $CATT_{e,t}$ estimates from step 1 to form average treatment effect estimates with weight estimates from step 2.

The coefficient estimator $\hat{\delta}_{e,t}$ from Regression (7) is a consistent DID estimator for $CATT_{e,t}$ in the form of (6) with $s = 0$ and $C = \{T\}$. (Among all possible DID estimators for $CATT_{e,t}$, we propose Regression (7) because it is a natural extension to the dynamic FE specification (3) by saturating in relative time indicators and cohort indicators.) Thus, the IW estimator is consistent for a weighted average of $CATT_{e,t}$ with weights equal to the share of each cohort in the relevant period(s). These weights are guaranteed to be convex and have an interpretation as the representative shares corresponding to each $CATT_{e,t}$.

With a few standard assumptions (which we present as Assumption 5 in Online Appendix A) on Regression (7), we can show that the IW estimators are asymptotically normal and derive its asymptotic variance.
4.3 Comparing FE and IW estimators

Assuming homogeneous treatment effects both the dynamic FE and IW estimators are consistent for $CATT_{e,t}$, which are constant across $e$ for a given $l$. If further treatment effects are stationary and there is no anticipatory behavior, both the static FE and IW estimators are consistent for $CATT_{e,l}$, which is constant across $e$ and $l > 0$. The FE estimators may have smaller variance when they are consistent because specifications (2) and (3) estimate fewer coefficients. In particular, the FE estimators are more efficient when regressors in specifications (2) and (3) are linear transformations of regressors in specification (7) and errors are homoskedastic and serially uncorrelated. In the presence of heterogeneous treatment effects, only the saturated specification (7) guarantees causally interpretable estimates. It may reduce the standard deviation of residuals due to better fit. The IW estimators may thus be more precise than FE estimators, which we find to be the case in our empirical application.

5 Applications

5.1 Setting and data

We illustrate the empirical importance of our findings by extending the Dobkin et al. (2018) analysis on the dynamic effects of hospitalization. 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 Survey (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” 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). Unlike Dobkin et al. (2018), we restrict our analysis to a subsample of these individuals who appear throughout waves 7-11 (roughly 2004-2012) to maintain a balanced panel with a reasonable sample size. Our sample of analysis therefore includes HRS respondents with index hospitalization during waves 8-11. We also include adults at older ages, spanning 50-97 where as for their main results, Dobkin et al.
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. In our terminology, we categorize individuals into cohorts based on $E_i$, which is defined as the survey wave of their index hospitalization. Since we restrict the sample to individuals who were ever hospitalized between wave 8-11, there are four cohorts $E_i \in \{1, 2, 3, 4\}$.

The treatment status $D_{i,t}$ is thus an indicator for whether a respondent has had a hospitalization by wave $t$. Although hospitalization itself may not be an absorbing state, we are trying to model the impact of having had any hospitalization. Thus, an event study is an appropriate research design as it allows us to trace out the path of treatment effects following a negative health shock (even though the shock itself may be transient). Our outcomes of interest $Y_{i,t}$ include 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 1 presents basic summary statistics for our analysis sample before hospitalization. By design, our sample is older than that of Dobkin et al. (2018) because they restrict to non-elderly adults with ages 50-59 at the time of index hospitalization; the age difference also contributes to other observed demographic differences. We have a slightly lower fraction of men in our sample, as well as a higher fraction white and lower fraction black. Many individuals have also aged into Medicare coverage (for which the qualifying age is 65).
Table 1: Summary Statistics of the HRS sample

<table>
<thead>
<tr>
<th>Panel A. Demographics</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at admission</td>
<td>2,813</td>
<td>68</td>
<td>9</td>
</tr>
<tr>
<td>Male</td>
<td>2,813</td>
<td>0.425</td>
<td>0.494</td>
</tr>
<tr>
<td>Year of admission</td>
<td>2,813</td>
<td>2,007</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Race/ethnicity</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>2,813</td>
<td>0.097</td>
<td>0.296</td>
</tr>
<tr>
<td>Black</td>
<td>2,813</td>
<td>0.126</td>
<td>0.332</td>
</tr>
<tr>
<td>White</td>
<td>2,813</td>
<td>0.817</td>
<td>0.387</td>
</tr>
<tr>
<td>Other race</td>
<td>2,813</td>
<td>0.057</td>
<td>0.232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Insurance status</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicaid</td>
<td>2,813</td>
<td>0.057</td>
<td>0.232</td>
</tr>
<tr>
<td>Private</td>
<td>2,813</td>
<td>0.508</td>
<td>0.5</td>
</tr>
<tr>
<td>Medicare</td>
<td>2,813</td>
<td>0.57</td>
<td>0.495</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D. Pre-hospitalization outcome</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out-of-pocket medical spending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 7 (E_i \geq 1)</td>
<td>2,813</td>
<td>3,143</td>
<td>8,797</td>
</tr>
<tr>
<td>Wave 8 (E_i \geq 2)</td>
<td>1,980</td>
<td>2,339</td>
<td>5,767</td>
</tr>
<tr>
<td>Wave 9 (E_i \geq 3)</td>
<td>1,269</td>
<td>1,858</td>
<td>2,746</td>
</tr>
<tr>
<td>Wave 10 (E_i = 4)</td>
<td>536</td>
<td>2,412</td>
<td>4,105</td>
</tr>
<tr>
<td>Earnings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 7 (E_i \geq 1)</td>
<td>2,813</td>
<td>23,656</td>
<td>49,163</td>
</tr>
<tr>
<td>Wave 8 (E_i \geq 2)</td>
<td>1,980</td>
<td>20,539</td>
<td>43,201</td>
</tr>
<tr>
<td>Wave 9 (E_i \geq 3)</td>
<td>1,269</td>
<td>17,723</td>
<td>40,486</td>
</tr>
<tr>
<td>Wave 10 (E_i = 4)</td>
<td>536</td>
<td>14,773</td>
<td>39,246</td>
</tr>
</tbody>
</table>

In Panel D, we compare the mean outcomes for individuals who have not yet been hospitalized by each subsequent wave. The number of respondents hospitalized in each cohort is \(N_1 = 833\), \(N_2 = 711\), \(N_3 = 733\), and \(N_4 = 536\). There are apparent time trends in our outcomes of interest prior to hospitalization as we observe distributional changes across waves. Out-of-pocket medical spending fluctuates and earnings decrease with each wave on average as more individuals are retired in each subsequent wave.
5.2 Assumptions and estimation

Parallel trends (Assumption 1). We think that assuming the baseline outcome $Y_{i,t}^{\infty}$ is mean independent of the timing of the shock is too strong, because even among individuals who are ever hospitalized the hospitalization is likely to be earlier among sicker individuals with high out-of-pocket medical spending and low labor earnings. However, the parallel trends assumption is more plausible as it allows the timing to depend on unobserved time-invariant characteristics such as chronic disease. Note that for their main specification Dobkin et al. (2018) use a different identifying assumption that, conditional on ever being hospitalized, the timing of hospitalization is uncorrelated with deviations of the baseline outcome from a linear trend in relative time. Thus, their main specification deviates from the dynamic FE specification (3), which prevents a direct comparison between their results and our results, even though their results are qualitatively similar to our IW estimates.5

No anticipatory behavior (Assumption 2). Because we have restricted to conditions that are likely unexpected hospitalizations, it is plausible that there is no anticipatory behavior. This assumption may be violated if individuals have private information about the probability of these hospitalizations over time and respond proactively prior to hospitalization.

Treatment effect heterogeneity (Assumptions 3 and 4). For the two outcomes we study, we think the assumptions of stationarity and homogenous treatment effects are likely to be violated. 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 affected by the labor market: 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. We suspect these sources of heterogeneity may be exacerbated in this sample because the elderly population has experienced an especially rapid change in healthcare spending and labor participation over time.

Estimation. We use both FE and IW estimators to estimate the impact of hospitaliza-

5In their robustness analysis, Dobkin et al. (2018) considers a specification with two-way fixed effects, which is similar to the dynamic FE specification (3). In Online Appendix E, we replicate their results and explain why their specification alleviates the issue we identify.
tion on out-of-pocket medical spending and labor earnings. We first estimate the following two specifications without survey weights. For the linear two-way dynamic FE specification (3), we estimate

$$ Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-2}^{3} \mu_l D_{i,t}^l + \nu_{i,t} $$

(8)

for $t = 0, \ldots, 4$. For the saturated specification (7), we estimate

$$ Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \in \{1, 2, 3\}} \sum_{l=1-e}^{T-e} \delta_{e,l} 1\{E_i = e\} \cdot D_{i,t}^l + \epsilon_{i,t} $$

(9)

for $t = 0, \ldots, 3$. We drop $t = 4$ from (9) because everyone has been hospitalized by $t = 4$, and $CATT_{e,l}$ in $t = 4$ are not identified. Then for the IW estimators $\hat{\nu}_l$, our proposed alternative estimators to the FE estimators $\hat{\mu}_l$, we take weighted averages of $\hat{\delta}_{e,l}$ with weights that are sample share of each cohort $e$ across cohorts that are observed $l$ periods after hospitalization. We take $Y_{i,t}$ equal to out-of-pocket medical spending as well as labor earnings.

5.3 Results and discussion

In Table 2, we report the FE estimates $\hat{\mu}_l$ and the IW estimates $\hat{\nu}_l$, as well as the underlying $CATT_{e,l}$ estimates $\hat{\delta}_{e,l}$. While not reported, we fail to reject the joint hypothesis of parallel trends and no anticipation based on a Wald test for $H_0 : \mu_{-2} = \mu_{-1} = 0$ or $H_0 : \delta_{3,-2} = \delta_{3,-1} = \delta_{2,-1} = 0$.

Recall that the IW estimands $\nu_l$ are weighted averages of estimates for $CATT_{e,l}$, with weights equal to the share of cohort $e$ across cohorts that experience at least $l$ periods of treatment. Therefore by construction, the IW estimates fall within the convex hull of the $CATT_{e,l}$ estimates and have an interpretation as an average effect of the treatment $l$ periods after initial treatment. In contrast, the FE estimates are not within the convex hull of the $CATT_{e,l}$ estimates and thus do not have a causal interpretation.

Figure 1 provides a visual comparison between the dynamic FE estimates for the impact of hospitalization and the IW estimates for two outcomes: out-of-pocket medical spending and labor earnings. The IW estimates show qualitatively similar results to Dobkin et al. (2018). The effect on out-of-pocket spending is significant and positive for $l = 0$ (roughly
Table 2: Estimates for the Effect of Hospitalization on Outcomes

(a) Out-of-pocket Medical Spending

<table>
<thead>
<tr>
<th>$l$ wave relative to hospitalization</th>
<th>FE estimates</th>
<th>IW estimates</th>
<th>Estimates for $CATT_{e,l}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\mu}_l$</td>
<td>$\hat{\nu}_l$</td>
<td>$\hat{\delta}_{1,l}$</td>
</tr>
<tr>
<td>-2</td>
<td>-630</td>
<td>-301</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(571)</td>
<td>(291)</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>-1,122</td>
<td>-386</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(907)</td>
<td>(248)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>160</td>
<td>1,297</td>
<td>676</td>
</tr>
<tr>
<td></td>
<td>(1255)</td>
<td>(366)</td>
<td>(506)</td>
</tr>
<tr>
<td>1</td>
<td>-1,396</td>
<td>224</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>(1644)</td>
<td>(379)</td>
<td>(404)</td>
</tr>
<tr>
<td>2</td>
<td>-1,706</td>
<td>251</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td>(2010)</td>
<td>(468)</td>
<td>(468)</td>
</tr>
<tr>
<td>3</td>
<td>-1,775</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(2394)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Labor Earnings

<table>
<thead>
<tr>
<th>$l$ wave relative to hospitalization</th>
<th>FE estimates</th>
<th>IW estimates</th>
<th>Estimates for $CATT_{e,l}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\mu}_l$</td>
<td>$\hat{\nu}_l$</td>
<td>$\hat{\delta}_{1,l}$</td>
</tr>
<tr>
<td>-2</td>
<td>1,600</td>
<td>1,099</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(1743)</td>
<td>(897)</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>2,475</td>
<td>387</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(2953)</td>
<td>(956)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2,221</td>
<td>-527</td>
<td>-557</td>
</tr>
<tr>
<td></td>
<td>(4268)</td>
<td>(1449)</td>
<td>(1866)</td>
</tr>
<tr>
<td>1</td>
<td>1,352</td>
<td>-3,337</td>
<td>-3,518</td>
</tr>
<tr>
<td></td>
<td>(5609)</td>
<td>(1664)</td>
<td>(2015)</td>
</tr>
<tr>
<td>2</td>
<td>773</td>
<td>-5,375</td>
<td>-5,375</td>
</tr>
<tr>
<td></td>
<td>(6965)</td>
<td>(2230)</td>
<td>(2229)</td>
</tr>
<tr>
<td>3</td>
<td>2,434</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(8396)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 $\hat{\mu}_l$. The sample includes observations from wave $t = 0, 1, 2, 3, 4$. Columns 3-5 report the estimates for $CATT_{e,l}$ from $\hat{\delta}_{e,l}$. 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 $\hat{\delta}_{e,l}$’s across cohorts $e$ who are $l$ periods from hospitalization. Standard errors (clustered on the individual) are shown in parentheses.
Figure 1: FE vs IW Estimates of the Effects of Hospitalization on Outcomes

(a) Out-of-pocket Medical Spending

(b) Labor Earnings

Notes: Each figure plots FE estimates \( \hat{\mu}_l \) in triangles and IW estimates \( \hat{\nu}_l \) in circles against relative wave \( l \), with their respective 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.

1 year after hospitalization). The effect on earnings becomes significantly negative starting in \( l = 1 \) (roughly 3 years after hospitalization) and remains large for \( l = 2 \). In contrast, the dynamic FE estimates are all statistically insignificant, even though the underlying \( CATT_{e,l} \) are mostly precisely estimated as shown in Table 2. Furthermore, while they are not statistically different, the FE point estimates are of the opposite signs of the IW point estimates, and are misleading as they suggest that out-of-pocket medical spending decrease while earnings increase after hospitalization.

The dynamic specification produces wrong-signed and imprecise estimates because the \( \hat{\mu}_l \) are sensitive to effects from other time periods. Recall by Proposition 2, the FE estimand \( \mu_l \) is

\[
\sum_{1 \leq e \leq T-l} \omega_{e,l}^l CATT_{e,l} + \sum_{l' \neq l} \sum_{-l' \leq e \leq T-l'} \omega_{e,l'}^l CATT_{e,l'}. \tag{10}
\]

The point estimate \( \hat{\mu}_l \) is the sample analog of expression (10).

We illustrate the issues with the FE estimator by focusing on the estimation of a single coefficient, \( \mu_0 \). For \( l = 0 \), we can estimate the underlying weights \( \omega_{e,0}^0 \) by regressing \( 1 \{ E_i = e \} \cdot D_{i,t}^l \) on the relative time indicators in specification (8) i.e. \( \{D_{i,t}^l\}_{l=-2}^3 \) and two-
way fixed effects; the coefficient estimator of $D_{i,t}^0$ consistently estimates $\omega_{e,t}^0$. The point estimate $\hat{\mu}_0$ is the product of these estimated weights $\hat{\omega}_{e,t}^0$ and $\hat{\delta}_{e,t}$. Figure 2 plots these estimated weights. The weights are large for leads of treatments (negative relative waves), which suggest that the FE estimate $\hat{\mu}_0$ is particularly sensitive to estimates of pre-trends, and does not isolate the contemporaneous effect of hospitalizations.

Figure 2: Weights $\hat{\omega}_{e,t}^0$ on cohort specific effect estimates $\hat{\delta}_{e,t}$ 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 $\hat{\delta}_{e,t}$’s, estimates for cohort-specific effects $CATT_{e,t}$’s from all cohorts $e$ and relative waves $l$. This figure plots the weight $\hat{\omega}_{e,t}^0$ associated with each $\hat{\delta}_{e,t}$ in forming the FE estimate $\hat{\mu}_0$.

We found no evidence of anticipatory behavior, so in the limit, the FE estimator $\hat{\mu}_0$ should not be affected by pre-trends since $CATT_{e,t} = 0$ for all $l < 0$; however in finite samples, the estimates $\hat{\delta}_{e,t}$ where $l < 0$ are not necessarily zero and can influence $\hat{\mu}_0$ (and its standard error) if their weights are non-negligible as shown in Figure 2.
6 Conclusions

This paper analyzes the behavior of commonly used linear two-way fixed effects (FE) estimators under heterogeneous treatment effects. We cast event studies in a potential outcomes framework and show that cohort-specific average treatment effects on the treated \( (CATT_{e,l}) \) are identified under parallel trends and no anticipation assumptions. We then clarify the notions of heterogeneity in event studies: \( CATT_{e,l} \) can vary within cohorts over time (non-stationarity) and within a relative time period across cohorts (heterogeneity across cohorts).

We derive the FE estimands for static and dynamic specifications and show that they are linear combinations of \( CATT_{e,l} \) which may include negative weights. This means that when treatment effects are heterogenous, the static FE estimate may not correspond to a causal effect as the estimand may fall outside the convex hull of \( CATT_{e,l} \). In the dynamic specification estimating relative time specific effects, in addition to non-convex weights, the FE estimator associated with \( l \) periods relative to initial treatment may pick up spurious terms consisting of treatment effects from periods other than \( l \).

Given the negative results on FE estimators, we propose “interaction-weighted” (IW) estimators for estimating dynamic treatment effects in event studies. These estimators are formed by first estimating \( CATT_{e,l} \) with a regression saturated in cohort and relative time indicators, and then averaging estimates of \( CATT_{e,l} \) across \( e \) at a given \( l \). These estimators are easy to implement and robust to heterogenous treatment effects; the IW estimator associated with relative time \( l \) is guaranteed to estimate a convex average of \( CATT_{e,l} \) using weights that are sample share of each cohort \( e \).

Finally, we illustrate the empirical relevance of our results by comparing FE and IW estimates for the dynamic effects of hospitalization on the economic outcomes of an elderly sample. We find that IW estimates are more precise and recover economically plausible estimates of the opposite sign from misleading FE estimates. We show that the implicit weights for the FE estimates are non-convex, and that the FE estimates of contemporaneous effects are sensitive to pre-trends estimates and negatively influenced by long-run effects, which makes them causally uninterpretable.
References


**URL:** http://www.jstor.org/stable/3700681


**URL:** http://www.aeaweb.org/articles?id=10.1257/aer.p20171038


**URL:** https://ideas.repec.org/a/oup/qjecon/v129y2014i3p1141-1219.html


**URL:** http://www.nber.org/papers/w23472

**URL:** [http://www.aeaweb.org/articles?id=10.1257/aer.20161038](http://www.aeaweb.org/articles?id=10.1257/aer.20161038)


**URL:** [https://ideas.repec.org/a/tpr/restat/v88y2006i3p389-432.html](https://ideas.repec.org/a/tpr/restat/v88y2006i3p389-432.html)


**URL:** [http://www.aeaweb.org/articles?id=10.1257/aer.20151655](http://www.aeaweb.org/articles?id=10.1257/aer.20151655)


**URL:** [http://www.nber.org/papers/w24219](http://www.nber.org/papers/w24219)


**URL:** [http://www.aeaweb.org/articles?id=10.1257/aer.103.6.2530](http://www.aeaweb.org/articles?id=10.1257/aer.103.6.2530)


**URL:** [http://www.nber.org/papers/w24382](http://www.nber.org/papers/w24382)
Online Appendix

A Proofs

Proof of Proposition 1

Proof. Partiailling out the unit and time fixed effects, Regression (2) is

\[ \tilde{Y}_{i,t} = \gamma \tilde{D}_{i,t} + \epsilon_{i,t} \]

where \( \tilde{X}_{i,t} \) is time- and cross-sectional demeaned version of \( X_{i,t} \) i.e. \( \tilde{X}_{i,t} = X_{i,t} - \bar{X}_t - \bar{X}_i + \bar{X} \), and \( \bar{X}_t = E[X_{i,t}], \bar{X}_i = \frac{1}{T+1} \sum_{t=0}^{T} X_{i,t}, \bar{X} = \frac{1}{T+1} \sum_{t=0}^{T} E[X_{i,t}] \). It is easy to see that for \( \tilde{Y}_{i,t} \) and any \( \tilde{Z}_{i,t} \), demeaned versions of \( X_{i,t} \) and \( Z_{i,t} \) respectively, we have \( E[\tilde{X}_{i,t} \tilde{Z}_{i,t}] = E[X_{i,t} \tilde{Z}_{i,t}] = E[\tilde{X}_{i,t} Z_{i,t}] \). The population regression coefficient on \( D_{i,t} \) from specification (2), i.e. the probability limit of the coefficient estimator \( \hat{\gamma} \), is then

\[ \gamma = \frac{\frac{1}{T+1} \sum_{t=0}^{T} E[\tilde{Y}_{i,t} \tilde{D}_{i,t}]}{\frac{1}{T+1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}]} = \frac{\sum_{t=0}^{T} E[Y_{i,t} \tilde{D}_{i,t}]}{\sum_{t=0}^{T} E[D_{i,t} \tilde{D}_{i,t}]} = \frac{\sum_{t=0}^{T} E[(Y_{i,t} - Y_{i,0}) \tilde{D}_{i,t} \mid E_i]}{\sum_{t=0}^{T} E[D_{i,t} \tilde{D}_{i,t}]} \]

Note that \( \tilde{D}_{i,t} \) is a function of \( E_i \). By definition \( e \) is the time at first treatment so \( D_{i,t} = 1 \) in all times for all \( t \geq e \) and \( D_{i,t} = 0 \) in all times \( t < e \). The time fixed effects simplify to \( \tilde{D}_{i,t} = Pr \{ D_{i,t} = 1 \} = Pr \{ E_i \leq t \} \), the share of treated units at time \( t \). The individual effects become the share of treated periods, \( \tilde{D}_i \) := \( \bar{D}_{e_i} := \frac{T-e_i}{T+1} \). The mean \( \bar{D} = \frac{T-E[E_i]}{T+1} \) is fixed for all \( i \). For a given cohort, \( E_i = e \) and \( \tilde{D}_{i,t} = D_{i,t} - D_{i,e} + D \) is thus non-random. Using this expression, we can write

\[ \gamma = \frac{\sum_{t=0}^{T} E[\tilde{D}_{i,t} E[Y_{i,t}^{E_i} - Y_{i,0}^{\infty} \mid E_i]]}{\sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}]} = \frac{E[Y_{i,t}^{E_i} - Y_{i,0}^{\infty} \mid E_i = e]}{E[Y_{i,t}^{\infty} - Y_{i,0}^{\infty} \mid E_i = e, D_{i,t} = 1]} \]

\[ + \sum_{e>t} Pr\{E_i = e\}(0 - \tilde{D}_{e_i} + D) \frac{E[Y_{i,t}^{E_i} - Y_{i,0}^{\infty} \mid E_i = e, D_{i,t} = 1]}{E[Y_{i,t}^{\infty} - Y_{i,0}^{\infty} \mid E_i = e, D_{i,t} = 0]} \]

The equality above the first brace follows from no anticipation (Assumption 2). The equality above the second brace follows from parallel trends and no anticipation (Assumption 1 and 2).
Furthermore, the weights on $E[Y^e_{i,t} - Y^\infty_{i,0} | E_i = e]$ for all cohorts $e$ and $E[Y^\infty_{i,t} - Y^\infty_{i,0}]$ sum up to zero.

$$
\sum_{e \leq t} Pr\{E_i = e\}(1 - \bar{D}_{i,t} - \bar{D}_{e, \cdot} + \bar{D}) + \sum_{e > t} Pr\{E_i = e\}(0 - \bar{D}_{i,t} - \bar{D}_{e, \cdot} + \bar{D})
= Pr\{D_{i,t} = 1\} - \bar{D}_{\cdot,t} - \sum_e \bar{D}_{e, \cdot} + \bar{D} = \bar{D}_{\cdot,t} - \bar{D}_{\cdot,t} - \bar{D} + \bar{D} = 0.
$$

With this, we can distribute $E[Y^\infty_{i,t} - Y^\infty_{i,0}]$ across the treated units and obtain the following expression:

$$
\frac{\sum_{e=0}^{T-e} \sum_{i=0}^{T} Pr\{E_i = e\}(1 - \bar{D}_{t,e+l} - \bar{D}_{e, \cdot} + \bar{D}) \sum_{t=0}^{T} E[D_{i,t} \bar{D}_{i,t}] E[Y^e_{i,e+l} - Y^\infty_{i,e+l} | E_i = e]}{\sum_{e=0}^{T} \sum_{i=0}^{T} Pr\{E_i = e\}(1 - \bar{D}_{t,e+l} - \bar{D}_{e, \cdot} + \bar{D})}
$$

To see that the weights sum to one, note that the denominator of the weights can be written as

$$
\sum_{t=0}^{T} E[D_{i,t} \bar{D}_{i,t}] = \sum_{t=0}^{T} \sum_{e \leq t} Pr\{E_i = e\}(1 - \bar{D}_{i,t} - \bar{D}_{e, \cdot} + \bar{D})
= \sum_{e=0}^{T-e} \sum_{i=0}^{T} Pr\{E_i = e\}(1 - \bar{D}_{t,e+l} - \bar{D}_{e, \cdot} + \bar{D})
$$

which is equal to the sum of the numerator of all weights.

Proof of Propositions 2 and 3

Proof. Following a similar argument in the proof for Proposition 1, the population regression coefficients $\mu_l$ from regression (3) are:

$$
(\mu_l)_{-T \leq l \leq T} = \left( \sum_{t=0}^{T} E[\bar{D}_{i,t} D_{i,t}^l] \right)^{-1} \sum_{t=0}^{T} E[\bar{D}_{i,t} Y_{i,t}]
$$

31
where \( \bar{D}_{i,t} = \begin{pmatrix} \bar{D}_{i,t}^{-T} \\ \vdots \\ \bar{D}_{i,t}^T \end{pmatrix} \) is a column vector collecting \( \bar{D}_{i,t}^l \). Again, \( \bar{X}_{i,t} \) is time- and cross-sectional demeaned version of \( X_{i,t} \).

To further develop the expression for the population regression coefficients \( \mu_l \), we note that

\[
\sum_{t=0}^{T} E[\bar{D}_{i,t} Y_{i,t}] = \sum_{t=0}^{T} E[\bar{D}_{i,t} (Y_{i,t} - Y_{i,0}^\infty)] = \sum_{t=0}^{T} E[\bar{D}_{i,t} E[Y_{i,t} - Y_{i,0}^\infty | E_i]]
\]

\[
= \sum_{t=0}^{T} E[\bar{D}_{i,t} E[Y_{i,t} - Y_{i,0}^\infty | E_i]] - \sum_{t=0}^{T} E[\bar{D}_{i,t}] E[Y_{i,t}^\infty - Y_{i,0}^\infty] = 0
\]

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

Under the parallel trends, we have \( E[Y_{i,t}^\infty - Y_{i,0}^\infty | E_i] = E[Y_{i,t}^\infty - Y_{i,0}^\infty] \). Then the difference in trends \( E[Y_{i,t} - Y_{i,0}^\infty | E_i] - E[Y_{i,t}^\infty - Y_{i,0}^\infty] \) is equal to average treatment effects \( E[Y_{i,t} - Y_{i,0}^\infty | E_i] \).

The expression for population regression coefficients \( \mu_l \) simplifies to

\[
\left( \sum_{t=0}^{T} E[\bar{D}_{i,t} \bar{D}_{i,t}^T] \right)^{-1} \sum_{t=0}^{T} E[\bar{D}_{i,t} E[Y_{i,t} - Y_{i,0}^\infty | E_i]]
\]

Furthermore, we can write the average treatment effect for a given cohort as a product of relative time indicators \( D_{i,t}^l \) and cohort-specific average treatment effects \( CATT_{E_i,l} \):

\[
E[Y_{i,t} - Y_{i,0}^\infty | E_i] = \begin{pmatrix} D_{i,t}^{-T} \\ \vdots \\ D_{i,t}^T \end{pmatrix}^T \begin{pmatrix} E[Y_{i,t} - Y_{i,0}^\infty | E_i] \end{pmatrix} = \begin{pmatrix} D_{i,t}^{-T} \\ \vdots \\ D_{i,t}^T \end{pmatrix}^T \begin{pmatrix} CATT_{E_i,-T} \\ \vdots \\ CATT_{E_i,T} \end{pmatrix}
\]

\[\text{Note that } \sum_{t=-T}^{T} \bar{D}_{i,t}^l = 0 \text{ and we need to drop one } \bar{D}_{i,t}^l \text{ due to multicollinearity. Another multicollinearity is } \sum_{t=-T}^{T} (l - l') \bar{D}_{i,t}^l = 0 \text{ for } l' \text{ the excluded relative time indicator variable, which is caused by a linear relationship between relative time indicators and two-way fixed effects. These collinearities are discussed by Borusyak and Jaravel (2017).} \]
For a given cohort \( e \), \( l \) ranges from \(-e\) to \( T - e\) because at most we observe \( e \) periods before initial treatment and \( T - e \) periods after initial treatment. For indices that are out of range (\( l < -e \) or \( l > T - e \)), the above expression is well-defined even though \( \text{CATT}_{e,l} \) is undefined because the corresponding \( D^l_{i,t} = 0 \) for all \( t \). Similarly, if there are never-treated units with \( E_i > T \), \( D_{i,t} \) is a zero vector for these units so the above expression is well-defined regardless of their \( \text{CATT}_{e,l} \).

We can make two more simplifications to \( f(E_i) \). If the no anticipation assumption holds, we can set \( \text{CATT}_{E_i,-T} \) through \( \text{CATT}_{E_i,-1} \) to zero. For relative times \( \bar{l} \) excluded from specification (3), we assume \( \text{CATT}_{E_i,\bar{l}} \) are zero. See Section A.1 for results when \( \text{CATT}_{E_i,\bar{l}} \) are not zero.

Using this matrix notation, the expression for population regression coefficients \( \mu_l \) simplifies to

\[
\left( \sum_{t=0}^{T} E[\hat{D}_{i,t} D^l_{i,t}] \right)^{-1} \sum_{t=0}^{T} E[\hat{D}_{i,t} D^l_{i,t} f(E_i)].
\]

While \( D_{i,t} D^l_{i,t} \) is a diagonal matrix with diagonal entries equal to one for \( t = E_i + l \) and zero otherwise, \( D_{i,t} D^l_{i,t} \) is not a diagonal matrix. This means that treatment effects at lags not equal to \( l \) might affect the coefficient on \( D^l_{i,t} \). This proves that the population regression coefficient on \( D^l_{i,t} \) is a weighted average of \( \text{CATT}_{e,l} \) for \( 0 \leq e \leq T \) and all \( l' \neq l \).

Specifically, we can write \( \mu_l \) as a weighted average

\[
\sum_{0 \leq e \leq T-l} \omega^l_{e,l} \text{CATT}_{e,l} + \sum_{l' \neq l} \sum_{0 \leq e \leq T-l'} \omega^l_{e,l'} \text{CATT}_{e,l'},
\]

where \( \sum_{-l \leq e \leq T-l} \omega^l_{e,l} = 1 \) and for each \( l' \neq l \), \( \sum_{-l' \leq e \leq T-l'} \omega^l_{e,l'} = 0 \). The weight \( \omega^l_{u,l'} \) is equal to the population regression coefficient on \( D^l_{i,t} \) from regressing \( D^l_{i,t} \cdot 1 \{ E_i = e \} \) on relative time indicators and two-way fixed effects. To see that \( \sum_{-l \leq e \leq T-l} \omega^l_{e,l} = 1 \), note that the sum of weights for \( \text{CATT}_{e,l} \) across \( 0 \leq e \leq T - l \) in \( \mu_l \) is equal to

\[
\sum_{0 \leq e \leq T-l} \left( \sum_{t=0}^{T} E[\hat{D}_{i,t} D^l_{i,t}] \right)^{-1} \left( \sum_{t=0}^{T} E[\hat{D}_{i,t} D^l_{i,t} \cdot 1 \{ E_i = e \}] \right) e^l_t
\]

where \( e_t \) is a column vector with 1 in the entry corresponding to the entry of \( D^l_{i,t} \) in \( D_{i,t} \), and 0 otherwise. Multiplying with \( e^l_t \) returns the population regression coefficient on \( D^l_{i,t} \).
The above expression simplifies to
\[
\left( \sum_{t=0}^{T} E[D_{i,t}D_{i,t}'] \right)^{-1} \left( \sum_{0 \leq e \leq T-l} E[D_{i,t}D_{i,e+l}'] \right) e_{i}^{\top} = \left( \sum_{t=0}^{T} E[D_{i,t}D_{i,t}'] \right)^{-1} \left( \sum_{t=0}^{T} E[D_{i,t}D_{i,t}'] \right) e_{i}^{\top}
\]
It is thus the population regression coefficient on \( D_{i,t} \) from regressing \( D_{i,t} \) on \( D_{i,t} \) and the unit and time fixed effects, which is just one. Similarly, for each \( l' \neq l \), the sum of weights for \( CATT_{e,l'} \) across \( 1 \leq e \leq T-l' \) is population regression coefficient on \( D_{i,l'}^{e} \) from regressing \( D_{i,l'}^{e} \) on \( D_{i,t} \) and the unit and time fixed effects, which is zero. □

**Proof of Proposition 4**

*Proof.* Provided that the DID estimator is well-defined, the DID estimator is an unbiased and consistent estimator for \( E[Y_{i,e+l} - Y_{i,s} | E_{i} = e] - E[Y_{i,e+l} - Y_{i,s} | E_{i} \in C] \). 7 To show that it is an unbiased and consistent estimator for \( CATT_{e,l} \), we show \( E[Y_{i,e+l} - Y_{i,s} | E_{i} = e] - E[Y_{i,e+l} - Y_{i,s} | E_{i} \in C] = CATT_{e,l} \).

Since \( s < e \) and \( c > e + l \), we have
\[
E[Y_{i,e+l} - Y_{i,s} | E_{i} = e] - E[Y_{i,e+l} - Y_{i,s} | E_{i} \in C] = E[Y_{i,e+l}^{e} - Y_{i,s}^{e} | E_{i} = e] - \sum_{c \in C} Pr \{ E_{i} = c | E_{i} \in C \} E[Y_{i,e+l}^{c} - Y_{i,s}^{c} | E_{i} = c]
\]
\[
= E[Y_{i,e+l}^{e} - Y_{i,s}^{e} | E_{i} = e] - \sum_{c \in C} Pr \{ E_{i} = c | E_{i} \in C \} E[Y_{i,e+l}^{\infty} - Y_{i,s}^{\infty} | E_{i} = c]
\]
\[
= E[Y_{i,e+l}^{e} - Y_{i,s}^{\infty} | E_{i} = e] + E[Y_{i,e+l}^{\infty} - Y_{i,s}^{\infty} | E_{i} = e] - \sum_{c \in C} Pr \{ E_{i} = c | E_{i} \in C \} E[Y_{i,e+l}^{\infty} - Y_{i,s}^{\infty} | E_{i} = c]
\]
\[
= E[Y_{i,e+l}^{e} - Y_{i,s}^{\infty} | E_{i} = e] + E[Y_{i,e+l}^{\infty} - Y_{i,s}^{\infty} | E_{i} = e] - E[Y_{i,e+l}^{\infty} - Y_{i,s}^{\infty}]
\]
\[
= E[Y_{i,e+l}^{e} - Y_{i,s}^{\infty} | E_{i} = e]
\]

7Take the first term for an example. For unbiasedness, note that by the Law of Iterated Expectations
\[
E \left[ \frac{E_{N}[Y_{i,e+l} - Y_{i,s}] \cdot 1 \{ E_{i} = e \}}{E_{N}[1 \{ E_{i} = e \}]} \right] = E \left[ E \left[ \frac{E_{N}[Y_{i,e+l} - Y_{i,s}] \cdot 1 \{ E_{i} = e \}}{E_{N}[1 \{ E_{i} = e \}]} \right] \mid E_{i} \right]
\]
\[
= E \left[ \frac{E[E_{N}[Y_{i,e+l} - Y_{i,s}] \cdot 1 \{ E_{i} = e \}]}{E_{N}[1 \{ E_{i} = e \}]} \right] = E[(Y_{i,e+l} - Y_{i,s}) \mid E_{i} = e].
\]
For consistency, by the Law of Large Numbers the numerator and the denominator converge in probability to \( E[(Y_{i,e+l} - Y_{i,s}) \cdot 1 \{ E_{i} = e \}] \) and \( Pr \{ E_{i} = e \} \) respectively. By the Law of Iterated Expectations and Slutsky’s theorem, \( g^{l} \) converges in probability to \( E[Y_{i,e+l} - Y_{i,s} | E_{i} = e] \).

34
where the second equality follows from Assumption 2 and the fourth equality follows from Assumption 1.

**Definition 4.** The IW estimator for a weighted average of $CATT_{e,t}$ is constructed via the following three steps:

**Step 1** Estimate the $CATT_{e,t}$.

\[
y_{i,t} = \alpha_i + \lambda_t + \sum_{e=1}^{T-1} \sum_{l=1-e}^{T-1} \delta_{e,l} (1\{E_i = e\} \cdot D_{l,t}) + \epsilon_{i,t}
\]

\[
y_{i,t} = \alpha_i + \lambda_t + \beta_{i,t}^t \delta + \epsilon_{i,t} \tag{11}
\]

on $t = 0, \ldots, T - 1$ and $e = 1, \ldots, T$. Note that among regressors we exclude interactions with $1\{E_i = T\}$ and for each $e = 1, \ldots, T - 1$, we exclude interactions with $D_{e,t}^\infty$.

Here $B_{i,t}$ is a column vector collecting $1\{E_i = e\} \cdot D_{l,t}$ for each cohort $1 \leq e \leq T - 1$ with relative time spanning $1 - e \leq l \leq T - 1 - e$.

Similarly, $\delta$ is a column vector collecting the coefficients $\delta_{e,l}$ on $1\{E_i = e\} \cdot D_{l,t}^t$. The matrix notation is used later to derive the asymptotic variance of IW estimators.

We estimate the interacted specification.

**Step 2** Estimate the weights for the static or dynamic estimator.

Denote by $N_e := \sum_{i=1}^{N} 1\{E_i = e\}$ the number of units in cohort $e$. Below vec$(A)$ vectorizes matrix $A$ by stacking its columns.

1. **Static.** This set of weights is used in Step 3.1 to form a static estimator. Define $\hat{s}$ to be a matrix with its $(t,e)^{th}$ entry equal to $1\{t - e \geq 0\} \cdot (N_e/\sum_{e=1}^{T-1} N_e)/(T - e)$. Here $1\{t - e \geq 0\}$ indicates post-treatment, $N_e/\sum_{e=1}^{T-1} N_e$ is equal to the sample share of cohort $e$, and $T - e$ is the number of periods of treatment for cohort $e$. Denote by $s$ the probability limit of $\hat{s}$, which is a matrix with its $(t,e)^{th}$ entry equal to $1\{t - e \geq 0\} \cdot Pr\{E_i = e \mid 1 \leq E_i \leq T - 1\} / (T - e)$. In the Online Appendix, we show that the weight matrix estimator $\hat{s}$ is asymptotically normal $\sqrt{N} (\text{vec}(\hat{s}) - \text{vec}(s)) \to_d N(0, \Sigma_s)$.

---

8We need to drop time period $T$ because $CATT_{e,T}$ are not identified. We need to exclude cohort 0 from the sample because $CATT_{0,l}$ are not identified. See footnote ?? for more details.
2. **Dynamic.** This set of weights is used in Step 3.2 to form a dynamic estimator. Define \( \hat{f}^l \) to be a matrix with its \((t, e)^{th}\) entry equal to \(1\{t - e = l\} \cdot N_e/\sum_{e=1}^{T-1-l} N_e\). Here \(1\{t - e = l\}\) indicates when cohort \(e\) experiences exactly \(l\) periods of treatment and \(N_e/\sum_{e=1}^{T-1-l} N_e\) is equal to the sample share of units in cohort \(e\) among units that experience at least \(l\) periods of treatment. Denote by \(f^l\) the probability limit of \(\hat{f}^l\), which is a matrix with its \((t, e)^{th}\) entry equal to \(1\{t - e = l\} \cdot Pr\{E_i = t - l \mid 1 - l \leq E_i \leq T - 1 - l\}\). For example, for \(T = 3\) and \(l = 0\),

\[
\hat{f}^0 = \left( \begin{array}{cc} \frac{N_1}{N_1+N_2} & 0 \\ 0 & \frac{N_2}{N_1+N_2} \end{array} \right)
\]

and its probability limit is

\[
f^0 = \left( \begin{array}{cc} Pr\{E_i = 1 \mid 1 \leq E_i \leq 2\} & 0 \\ 0 & Pr\{E_i = 2 \mid 1 \leq E_i \leq 2\} \end{array} \right).
\]

In the Online Appendix, we show that the weight matrix estimator \(\hat{f}^l\) is asymptotically normal \(\sqrt{N}(vec(\hat{f}^l) - vec(f^l)) \rightarrow_d N(0, \Sigma_f^l)\).

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

1. **Static.** To form an estimator alternative to the static FE estimator \(\hat{\gamma}\) from Regression (2), we can use

\[
\hat{\kappa} := \sum_{e=1}^{T-1} \frac{N_e}{\sum_{e=1}^{T-1} N_e} \sum_{0 \leq t \leq T-1-e} \frac{1}{T-e} \hat{\delta}_{e,l} = vec(\hat{\mathbf{s}})^T \hat{\delta}.
\]

2. **Dynamic.** To form an estimator alternative to the dynamic FE estimator \(\hat{\mu}_l\) from Regression (3), we can use

\[
\hat{\nu}_l := \sum_{e=1}^{T-1-l} \frac{N_e}{\sum_{e=1}^{T-1-l} N_e} \hat{\delta}_{e,l} = vec(\hat{f}^l)^T \hat{\delta}.
\]

Note that \(\hat{\delta}_{e,l}\) from Regression (7) is a DID estimator for \(CATT_{e,l}\) in the form of (6) with \(s = 0\) and \(C = \{T\}\), which is consistent.\(^9\) Thus, as we state more formally below, \(\hat{\kappa}\)

\(^9\)Among all possible DID estimators for \(CATT_{e,l}\), we propose Regression (7) because it is a natural extension to the dynamic FE specification (3) by saturating in relative time indicators and cohort indicators.
is consistent for a weighted average of $\text{CATT}_{e,t}$ with weights equal to the share of cohort $e$ divided by $T - e$, the number of periods of treatment received by $e$; $\tilde{\nu}_t$ is consistent for a weighted average of $\text{CATT}_{e,t}$ with weights equal to the share of cohort $e$ across cohorts that have experienced at least $l$ periods of treatment. These weights are guaranteed to be convex and have an interpretation as representative shares corresponding to each $\text{CATT}_{e,t}$.

One can also form other weighted averages of $\hat{\delta}_{e,t}$ for other target parameters of interest such as the average treatment effects across a subset of lags or a subset of cohorts.

With a few standard assumptions (which we present together below as Assumption 5) on Regression (7), we can show that the IW estimators are asymptotically normal. We use the notation $\ddot{X}_{i,t}$ to denote double-demeaning $X_{i,t} - \overline{X}_{i,\cdot} + \overline{X}$, where $\overline{X}_{i,\cdot} = \sum_{t=0}^{T-1} X_{i,t}/T$, $\overline{X}_{\cdot,t} = E[X_{i,t}]$ and $\overline{X} = \sum_{t=0}^{T-1} E[X_{i,t}]/T$.

**Assumption 5.** (The saturated regression assumptions).

1. There are observations from at least two cohorts, one for $e = 1, \ldots, T - 1$ and the other for $e = T$.

2. Independent, identically distributed cross-sectional observations: $\{(E_i, Y_i) : i = 1, 2, \ldots, N\}$ are i.i.d. draws from their joint distribution where $Y_i = (Y_{i,0}, \ldots, Y_{i,T-1})^\top$ is a $T \times 1$ vector.

3. Large outliers are unlikely: $(B_{i,t}, \epsilon_{i,t})$ have nonzero finite fourth moments.

4. Denote by $\hat{B}$ the data matrix, whose rows consist of $\hat{B}_{i,t}^\top$, double-demeaned version of $B_{i,t}^\top$. Assume $\hat{B}$ has full rank. If $\hat{B}$ is reduced-rank because cohort $e$ is empty, then discard regressors involving $1\{E_i = e\}$.

Denote by $\delta$ the probability limit of $\hat{\delta}$, which is a vector of $\text{CATT}_{e,t}$. We next state the asymptotic distribution of the IW estimators. 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 5.** (Consistency and asymptotic normality of the IW estimator for overall treatment effects). Under Assumptions 1, 2 and 5, the IW estimator converges in probability
The asymptotic distribution of this estimator is

$$\sqrt{N} (\hat{\kappa} - vec(s)^T \delta) \rightarrow_d N(0, \delta^T \Sigma_s \delta + \Sigma)$$

for $\Sigma_s$ the asymptotic variance of $\sqrt{N} (vec(\hat{s}) - vec(s))$ where $\hat{s}$ is the weight matrix estimator and

$$V_B = \sum_{t=0}^{T-1} E[\hat{B}_{i,t} \hat{B}_{i,t}^T] \quad \Sigma = vec(s)^T V_B^{-1} Var \left( \sum_{t=0}^{T-1} \hat{B}_{i,t} \epsilon_{i,t} \hat{B}_{i,t}^T \right) V_B^{-1} vec(s).$$

**Proposition 6.** (Consistency and asymptotic normality of the IW estimators for dynamic treatment effects). Under Assumptions 1, 2 and 5, the IW estimator converges in probability to

$$\hat{\nu}_t \rightarrow_p \sum_{e=1}^{T-1} Pr \{ E_i = e \mid 1 - l \leq E_i \leq T - 1 - l \} \text{ATT}_{e,t} = vec(f^t)^T \delta.$$

The asymptotic distribution of this estimator is

$$\sqrt{N} (\hat{\nu}_t - vec(f^t)^T \delta) \rightarrow_d N(0, \delta^T \Sigma_{f1} \delta + \Sigma)$$

for $\Sigma_{f1}$ the asymptotic variance of $\sqrt{N}(vec(\hat{f}^t) - vec(f^t)) \rightarrow_d$ where $\hat{f}^t$ is the weight matrix estimator and

$$V_B = \sum_{t=0}^{T-1} E[\hat{B}_{i,t} \hat{B}_{i,t}^T] \quad \Sigma_{f1} = vec(f^t)^T V_B^{-1} Var \left( \sum_{t=0}^{T-1} \hat{B}_{i,t} \epsilon_{i,t} \hat{B}_{i,t}^T \right) V_B^{-1} vec(f^t).$$

**Proof of Proposition 5 and 6**

**Proof.** We first show the asymptotic normality of the weights. Recall that $\hat{s}$ is a matrix with its $(t,e)^{th}$ entry equal to $1 \{ t - e \geq 0 \} \frac{N_e}{\sum_{e=1}^{T} N_e} \frac{1}{T-e}$. Here $1 \{ t - e \geq 0 \}$ indicates post-treatment, $\frac{N_e}{\sum_{e=1}^{T} N_e}$ is equal to the sample share of cohort $e$, and $T - e$ is the number of periods of treatment for cohort $e$. Denote by $s$ the probability limit of $\hat{s}$, which is a matrix with its $(t,e)^{th}$ entry equal to $1 \{ t - e \geq 0 \} \frac{Pr\{E_{i=e}\leq E_{i}\leq T-1\}}{T-e}$ since $\frac{N_e}{\sum_{e=1}^{T} N_e} \rightarrow_p \frac{Pr\{E_{i=e}\}}{Pr\{1\leq E_{i}\leq T-1\}} = Pr \{ E_i = e \mid 1 \leq E_i \leq T - 1 \}$ by the Law of Large Numbers and Slutsky’s
Note that $\sqrt{E}$ OLS asymptotics which holds as $\text{vec}$ thus and has nonzero finite fourth moments. The asymptotic distribution of this estimator is $E$ correlated with covariance $\text{cross-sectional regression}$ $N$ theorem. Note that applies because by assumption after double demeaning, the data ($\ddot{e}$ with diagonal entries equal to $\beta$ $\sum_{t=1}^{T-1} \sum_{e=1}^{N_e} e_{t,e} \{1 \leq E_{i,t} \leq T - 1\} \rightarrow_d N \left( 0, \frac{E[1 \{1 \leq E_i \leq T - 1\}^2 \eta(e)_i^2]}{E[1 \{1 \leq E_i \leq T - 1\}^2]^2} \right)$.

Note that $1 \{1 \leq E_i \leq T - 1\}^2 = 1 \{1 \leq E_i \leq T - 1\}$ so

$$\frac{E[1 \{1 \leq E_i \leq T - 1\}^2 \eta(e)_i^2]}{E[1 \{1 \leq E_i \leq T - 1\}^2]^2} = \frac{E[\eta(e)_i^2 \mid 1 \leq E_i \leq T - 1]Pr \{1 \leq E_i \leq T - 1\}}{Pr \{1 \leq E_i \leq T - 1\}^2} = \frac{E[\eta(e)_i^2 \mid 1 \leq E_i \leq T - 1]}{Pr \{1 \leq E_i \leq T - 1\}}.$$ 

Similarly, for a pair of cohorts with $e \neq e'$, $\sum_{t=1}^{N_e} \sum_{e=1}^{N_e} \text{are asymptotically correlated with covariance } E[\eta(e)_i \eta(e')_i \mid 1 \leq E_i \leq T - 1] / Pr \{1 \leq E_i \leq T - 1\}$. Thus, $\text{vec}(\hat{s})$ has asymptotic distribution $\sqrt{N}(\text{vec}(\hat{s}) - \text{vec}(s)) \rightarrow_d N(0, \Sigma_a)$. $\Sigma_a$ is a matrix with diagonal entries equal to

$$1 \{t - e \geq 0\} \frac{E[\eta(e)_i^2 \mid 1 \leq E_i \leq T - 1]}{Pr \{1 \leq E_i \leq T - 1\} (T - e)^2},$$

and off-diagonal entries equal to

$$1 \{t - e \geq 0\} 1 \{t' - e' \geq 0\} \frac{E[\eta(e)_i \eta(e')_i \mid 1 \leq E_i \leq T - 1]}{Pr \{1 \leq E_i \leq T - 1\} (T - e) (T - e')}.$$ 

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

$$\sqrt{N} \left( \hat{\delta} - \delta \right) \rightarrow_d N \left( 0, V^{-1}_B \text{Var} \left( \sum_{t=0}^{T-1} \tilde{B}_{i,t} \tilde{e}_{i,t}^2 \tilde{B}_{i,t}^T \right) V^{-1}_B \right)$$

where $V_B = \sum_{t=0}^{T-1} E[\tilde{B}_{i,t}^T \tilde{B}_{i,t}]$. 39
Lastly, by the delta method, we have
\[
\sqrt{N} \left( \text{vec}(\tilde{s})^T \tilde{\delta} - \text{vec}(s)^T \delta \right) \rightarrow_d N \left( 0, \delta^T \Sigma \delta + \Sigma \right)
\]
where \( \Sigma = \text{vec}(s)^T V_B^{-1} \text{Var} \left( \sum_{t=0}^{T-1} \tilde{B}_{i,t} \tilde{e}_{i,t} V_B^{-1} \text{vec}(s) \right) \). This follows because \( \text{vec}(\tilde{s}) \) and \( \tilde{\delta} \) are uncorrelated: the asymptotic covariance between \( N \sum_{t=1}^{T} \eta(e_i) \tilde{e}_{i,t} \) is equal to
\[
V_B^{-1} \text{Cov} \left( \tilde{B}_{i,t} \tilde{e}_{i,t} \mid 1 \leq E_i \leq T - 1 \right)
\]
where \( \tilde{B} \) and \( \eta(e_i) \) are functions of \( E_i \), we have \( E[\tilde{B}_{i,t} \tilde{e}_{i,t} \mid 1 \leq E_i \leq T - 1] = E[\tilde{B}_{i,t} \tilde{e}_{i,t} \mid E_i, 1 \leq E_i \leq T - 1] \). Furthermore, specification (7) is saturated in \( E_i \) and relative time so \( E[\tilde{e}_{i,t} \mid E_i] = 0 \). This proves Proposition 5.

The proof of Proposition 6 is identical to the proof of Proposition 5. \( \Box \)

A.1 What relative times to exclude from dynamic FE specifications

In practice, researchers often exclude more than two relative times from specification (3). For example, all leads greater than \( K \) and lags greater than \( L \) are excluded, as well as relative time \( -1 \). For example, Dobkin et al. (2018) sets \( K = -3 \) and \( L = 3 \) in their “nonparametric event study”. This results in the following specification
\[
Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=0}^{L} \mu_l D_{i,t}^l + \sum_{l=-K}^{-2} \mu_l D_{i,t}^l + \upsilon_{i,t}. \tag{12}
\]
The population regression coefficients \( \mu_l \) are then a weighted average of all \( CATT_{e,l'} \) for \( l' \) included in (12), less a weighted average of all \( CATT_{e,\bar{l}} \) for \( \bar{l} \) excluded from (12). When \( CATT_{e,\bar{l}} = 0 \) for all \( \bar{l} < -K \) and \( \bar{l} > L \), the normalization term collapses to a weighted average of all \( CATT_{e,-1} \). If additionally \( CATT_{e,-1} \) is constant across all \( e \), coefficient estimates \( \hat{\mu}_l \) for included relative times are normalized relative to any effect the period prior to treatment. This finding cautions against excluding \( l = -1 \) from (3) when treatment effects are heterogenous and there is anticipation so that \( CATT_{e,-1} \) is likely nonzero, since it is more than a level normalization relative to the period prior to treatment, and could further complicate the interpretation of FE estimates \( \hat{\mu}_l \).
We formalize this result in the following proposition.

**Proposition 7.** Under Assumption 1, the population regression coefficient on the indicator for \( l \) periods relative to treatment is

\[
\mu_l = \sum_{0 \leq e \leq T-l} \omega_{l,e}^l \text{CATT}_{e,l} + \sum_{K \leq l' \leq L, l' \neq -1} \sum_{0 \leq e \leq T-l'} \omega_{e,l'} \text{CATT}_{e,l'} + \sum_{i < -K, =-1, > L} \sum_{0 \leq e \leq T-l'} \omega_{e,l} \text{CATT}_{e,l},
\]

where \( \sum_{-l \leq e \leq T-l} \omega_{l,e}^l = 1 \). For each \( l' \neq l \) included in (12), the weights on \( \text{CATT}_{e,l'} \) sum up to 0. The weights on \( \text{CATT}_{e,l} \) for excluded \( l \) sum up to -1.

**Proof.** Following a similar argument in the proof for Proposition 9, the population regression coefficients \( \mu_l \) from regression (12) are

\[
\left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top f(E_i)] = \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top f(E_i)] + \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top \tilde{f}(E_i)]
\]

where the vectors are

\[
D_{i,t} = \begin{pmatrix} D_{i,t}^{-T} \\ \vdots \\ D_{i,t}^0 \\ \vdots \\ D_{i,t}^T \end{pmatrix}, \quad \bar{D}_{i,t} = \begin{pmatrix} D_{i,t}^{-K} \\ \vdots \\ D_{i,t}^{-2} \\ \vdots \\ D_{i,t}^L \end{pmatrix}, \quad \tilde{D}_{i,t} = \begin{pmatrix} D_{i,t}^{-T} \\ \vdots \\ D_{i,t}^{-K} \\ \vdots \\ D_{i,t}^{L+1} \end{pmatrix}
\]

\[
f(E_i) = \begin{pmatrix} \text{CATT}_{E_i,-T} \\ \vdots \\ \text{CATT}_{E_i,0} \\ \vdots \\ \text{CATT}_{E_i,T} \end{pmatrix}, \quad \bar{f}(E_i) = \begin{pmatrix} \text{CATT}_{E_i,-K} \\ \vdots \\ \text{CATT}_{E_i,-2} \\ \vdots \\ \text{CATT}_{E_i,L} \end{pmatrix}, \quad \tilde{f}(E_i) = \begin{pmatrix} \text{CATT}_{E_i,-T} \\ \vdots \\ \text{CATT}_{E_i,-K} \\ \vdots \\ \text{CATT}_{E_i,L+1} \end{pmatrix}
\]

The bar denotes the included relative times and the tilde denotes the excluded relative times. Note that compared to the proof for Proposition (9), the extra term is

\[
\left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top \tilde{f}(E_i)] \right).
\]
This explains the extra term in the expression for $\mu_t$:

$$
\sum_{\tilde{l} < -K, \tilde{l} > -1, \tilde{l} > L} \sum_{0 \leq \tilde{e} \leq T - \tilde{l}} \omega_{e, \tilde{l}} CATT_{e, \tilde{l}}
$$

where $\omega_{e, \tilde{l}}$ is the population regression coefficient on $D_{i,t}^l$ from regressing $1 \{E_i = e\} \cdot D_{i,t}^l$ on the included relative time indicators and two-way fixed effects. To see that the weights on $CATT_{e, \tilde{l}}$ from all excluded relative time $\tilde{l}$ and cohorts $e$ add up to $-1$, note that the sum of these weights is the entry corresponding to $D_{i,t}^l$ in the following column vector

$$
\sum_{\tilde{l} < -K, \tilde{l} > -1, \tilde{l} > L} \left( \sum_{t=0}^{T} E[\bar{D}_{i,t} \bar{D}_{i,t}^T] \right)^{-1} \sum_{t=0}^{T} E[\bar{D}_{i,t} D_{i,t}^l] = -\sum_{K \leq l \leq L, l \neq -1} \left( \sum_{t=0}^{T} E[\bar{D}_{i,t} \bar{D}_{i,t}^T] \right)^{-1} \sum_{t=0}^{T} E[\bar{D}_{i,t} D_{i,t}^l]
$$

where the third equality follows from $\sum_{-T \leq t \leq T} D_{i,t}^l = 1$. The last line simplifies to a column vector of $-1$ because 1) regression coefficients on $\bar{D}_{i,t}$ from regressing 1 on them are all zeros; 2) each term in the summation

$$
\left( \sum_{t=0}^{T} E[\bar{D}_{i,t} \bar{D}_{i,t}^T] \right)^{-1} \left( \sum_{t=0}^{T} E[\bar{D}_{i,t} D_{i,t}^l] \right)
$$

equals the population regression coefficient on $\bar{D}_{i,t}$ from regressing $D_{i,t}^l$ on $\bar{D}_{i,t}$ and the unit and time fixed effects, which is a column vector with one in the entry corresponding to $D_{i,t}^l$. \qed
B Results on first-differenced version of FE models

Consider the following variation to FE models:

\[ Y_{i,t} - Y_{i,t-1} = \lambda_t + \gamma D_{i,t} + \xi_{i,t} \]  \hspace{1cm} (13)

\[ Y_{i,t} - Y_{i,t-1} = \lambda_t + \sum_{t=-T+1}^{T} \mu_l D_{i,t} + \epsilon_{i,t} \]  \hspace{1cm} (14)

on \( t = 1, \ldots, T \). The only change compared to the FE models is that outcome is replaced with its first-difference. This variation is used when researchers want the coefficients on \( D_{i,t} \) and relative time indicators to capture effects on growth in \( Y_{i,t} \); a common outcome for this model is consumption. Below we show similar negative results on these versions of FE models: specifications (13) and (14) estimate some weighted averages of \( CATT_{e,l} \).

The weights in specifications (13) are non-negative but these weights do not correspond to any intuitive weights as they are proportional to treatment variance. The weights in specifications (14) can be negative.

**Proposition 8.** Denote \( \bar{D}_{\cdot,e} \) and \( \dot{D}_{i,t} = D_{i,t} - \bar{D}_{\cdot,t} \). Under Assumption 1 and 2, the population regression coefficient \( \gamma \) from (13) is

\[
\gamma = \frac{T}{\sum_{t=1}^{T} E[D_{i,t}\dot{D}_{i,t}]} \left( \sum_{t=1}^{T} E[(Y_{i,t} - Y_{i,t-1}) \dot{D}_{i,t}] \right)
\]

where the weights are non-negative \( \omega_{e,t} := \text{Pr}\{E_i = e | E_i \leq e + l\} \frac{\text{Pr}\{E_i \leq e + l\}}{\sum_{e=0}^{T} \text{Pr}\{E_i \leq e + l\}} \text{ and these weights sum up to one.}

**Proof.** Following a similar argument in the proof for Proposition 1, the population regression coefficient on \( D_{i,t} \) from (13) is

\[
\gamma = \frac{\sum_{t=1}^{T} E[(Y_{i,t} - Y_{i,t-1}) \dot{D}_{i,t}]}{\sum_{t=1}^{T} E[D_{i,t}\dot{D}_{i,t}]} = \frac{\sum_{t=1}^{T} E[E[(Y_{i,t} - Y_{i,t-1}) \dot{D}_{i,t} | E_i]]}{\sum_{t=1}^{T} E[D_{i,t}\dot{D}_{i,t}]}
\]

\[
= \frac{1}{\sum_{t=0}^{T} E[D_{i,t}\dot{D}_{i,t}]} \left( \sum_{t=1}^{T} \left\{ \sum_{e \leq t} \text{Pr}\{E_i = e\} \left( 1 - \bar{D}_{\cdot,t} \right) E[Y_{i,t}^e - Y_{i,t-1}^e | E_i = e, D_{i,t} = 1] \right. \right. \\
+ \left. \left. \sum_{e > t} \text{Pr}\{E_i = e\} \left( 0 - \bar{D}_{\cdot,t} \right) E[Y_{i,t}^e - Y_{i,t-1}^e | E_i = e, D_{i,t} = 0] \right\} \right).
\]
The equality above the brace follows from parallel trends and no anticipation (Assumption 1 and 2).

Furthermore, the weights on $E[Y_{i,t}^e - Y_{i,t-1}^e \mid E_i = e, D_{i,t} = 1]$ for all cohorts $e$ and $E[Y_{i,t}^\infty - Y_{i,t-1}^\infty]$ sum up to zero:

$$\sum_{e \leq t} Pr \{ E_i = e \} \left( 1 - \bar{D}_{.,t} \right) + \sum_{e > t} Pr \{ E_i = e \} \left( 0 - \bar{D}_{.,t} \right) = Pr \{ E_i \leq t \} - \bar{D}_{.,t} = \bar{D}_{.,t} - \bar{D}_{.,t} = 0$$

With this, we can distribute $E[Y_{i,t}^\infty - Y_{i,t-1}^\infty]$ across the treated units and obtain the following expression:

$$\sum_{t=0}^T \frac{1}{\sum_{t=0}^T E[D_{i,t} \hat{D}_{i,t}]} \sum_{t=1}^T \left\{ \sum_{e \leq t} Pr \{ E_i = e \} \left( 1 - \bar{D}_{.,t} \right) \left( E[Y_{i,t}^e - Y_{i,t-1}^e \mid E_i = e] - E[Y_{i,t}^\infty - Y_{i,t-1}^\infty] \right) \right\} \cdot E[Y_{i,e+1}^e - Y_{i,e+1}^\infty - (Y_{i,e+1}^e - Y_{i,e+1}^\infty) \mid E_i = e] \quad (17)$$

Note that $\bar{D}_{.,t} \left( 1 - \bar{D}_{.,t} \right)$ is the variance of treatment at time $t$ as $\bar{D}_{.,t} = Pr\{D_{i,t} = 1\}$.

To see that the weights sum to one, note that the denominator of the weights can be written as

$$\sum_{t=1}^T E[D_{i,t} \hat{D}_{i,t}] = \sum_{t=1}^T \sum_{e \leq t} Pr \{ E_i = e \} \left( 1 - \bar{D}_{.,t} \right)$$

$$= \sum_{e=0}^T \sum_{l=\max\{0,1-e\}}^{T-e} \bar{D}_{.,e+l} \left( 1 - \bar{D}_{.,e+l} \right) \frac{Pr \{ E_i = e \} \left( 1 - \bar{D}_{.,e+l} \right)}{\sum_{t=1}^T E[D_{i,t} \hat{D}_{i,t}]}$$
While these weights are non-negative and proportional to the share of treated cohorts
\( Pr \{ E_i = e \mid E_i \leq e + l \} \), they produce a treatment variance weighted average, which might not be intuitive.

Special Case: Stationary Treatment Effects. When treatment effects are stationary for all cohorts (Assumption 3), we have
\[
CATT_{e,l} - CATT_{e,l-1} = 0
\]
for all \( e \) and \( l > 0 \). This means there is no effect on the growth in outcomes beyond the first period of treatment \( (l > 0) \). Under this assumption, the population regression coefficient becomes a linear combination of cohort-specific effects on growth during the first period of treatment:

\[
\gamma = \sum_{e=1}^{T} \omega_{e,0} \cdot CATT_{e,0}
\]

where the weight is non-negative \( \omega_{e,0} := Pr \{ E_i = e \mid E_i \leq e \} \cdot \bar{D}_{e} \cdot e \cdot (1 - \bar{D}_{e} \cdot e) \) and these weights sum to less than one.

Next we analyze what the dynamic first-differencing specification (14) estimates without restrictions on treatment effects.

Proposition 9. Under Assumption 1 only, the population regression coefficient on the indicator for \( l \) periods from treatment is a weighted average of \( CATT_{e,l} - CATT_{e,l-1} \) for all \( e \) as well as \( CATT_{e,l'} - CATT_{e,l'-1} \) for all \( e \) and \( l' \neq l \),

\[
\mu_l = \sum_{0 \leq e \leq T-l} \omega_{e,l} \cdot (CATT_{e,l} - CATT_{e,l-1}) + \sum_{l' \neq l} \sum_{0 \leq e \leq T-l'} \omega_{e,l'} \cdot (CATT_{e,l'} - CATT_{e,l'-1}), \quad (18)
\]

where \( \sum_{-l \leq e \leq T-l} \omega_{e,l} = 1 \) and for each \( l' \neq l \), \( \sum_{-l \leq e \leq T-l'} \omega_{e,l'} = 0 \).

The weight \( \omega_{e,l'} \) is the projection coefficient of \( D_{e,t} \) from regressing \( 1 \{ E_i = e \} \cdot D_{i,t} \) on time fixed effects and the rest of the relative time indicators in (14) \( \{ D_{i,t} \}_{t \neq l} \).

Proof. Following a similar argument in the proof for Proposition 8, the population regression coefficients \( \mu_l \) from regression (14) are

\[
(\mu_l)_{-T+1 \leq t \leq T} = \left( \sum_{t=1}^{T} E[\hat{D}_{i,t} D_{i,t}^\prime] \right)^{-1} \left( \sum_{t=1}^{T} E[\hat{D}_{i,t} (Y_{i,t} - Y_{i,t-1})] \right)
\]
where $\hat{D}_{i,t} = \begin{pmatrix} \hat{D}_{i,t}^{T+1} \\ \vdots \\ \hat{D}_{i,t}^T \end{pmatrix}$ and again, $\hat{X}_{i,t}$ is time-demeaned version of $X_{i,t}$.

To further develop the expression for the population regression coefficients $\mu$, we note that

$$
\sum_{t=1}^{T} E[\hat{D}_{i,t} (Y_{i,t} - Y_{i,t-1})] = \sum_{t=1}^{T} E[\hat{D}_{i,t} E[Y_{i,t} - Y_{i,t-1}|E_i]]
$$

$$
= \sum_{t=1}^{T} E[\hat{D}_{i,t} E[Y_{i,t} - Y_{i,t-1}|E_i]] - \sum_{t=1}^{T} \left( E[\hat{D}_{i,t} E[Y_{i,t}^{\infty} - Y_{i,t-1}^{\infty}]] = 0 \right)
$$

$$
= \sum_{t=1}^{T} E[\hat{D}_{i,t} (E[Y_{i,t} - Y_{i,t-1}|E_i] - E[Y_{i,t}^{\infty} - Y_{i,t-1}^{\infty}]])]
$$

Under the parallel trends, we have $E[Y_{i,t}^{\infty} - Y_{i,t-1}^{\infty}|E_i] = E[Y_{i,t}^{\infty} - Y_{i,t-1}^{\infty}]$. Thus the expression for population regression coefficients $\mu$ now simplifies to

$$
\left( \sum_{t=1}^{T} E[\hat{D}_{i,t} D_{i,t}^T] \right)^{-1} \left( \sum_{t=1}^{T} E[\hat{D}_{i,t} E[Y_{i,t} - Y_{i,t}^{\infty} - (Y_{i,t-1} - Y_{i,t-1}^{\infty})|E_i]] \right).
$$

Furthermore, we can write

$$
E[Y_{i,t} - Y_{i,t}^{\infty} - (Y_{i,t-1} - Y_{i,t-1}^{\infty})|E_i] = \begin{pmatrix} D_{i,t}^{T+1} \\ \vdots \\ D_{i,t}^0 \\ \vdots \\ D_{i,t}^T \end{pmatrix}^T \begin{pmatrix} E[Y_{i,E_i-T+1} - Y_{i,E_i-T+1}^{\infty} - (Y_{i,E_i-T} - Y_{i,E_i-T}^{\infty})|E_i] \\ \vdots \\ E[Y_{i,E_i-E_i} - Y_{i,E_i-E_i}^{\infty} - (Y_{i,E_i-1} - Y_{i,E_i-1}^{\infty})|E_i] \\ \vdots \\ E[Y_{i,E_i} - Y_{i,E_i}^{\infty} - (Y_{i,E_i+T-1} - Y_{i,E_i+T-1}^{\infty})|E_i] \end{pmatrix}
$$

$$
= \begin{pmatrix} D_{i,t}^{T+1} \\ \vdots \\ D_{i,t}^0 \\ \vdots \\ D_{i,t}^T \end{pmatrix}^T \begin{pmatrix} CATT_{E_i, -T+1} - CATT_{E_i, -T} \\ \vdots \\ CATT_{E_i, 0} - CATT_{E_i, -1} \\ \vdots \\ CATT_{E_i, T} - CATT_{E_i, T-1} \end{pmatrix}.
$$

\[^{10}\text{As before, even though } l \text{ is summed over its possible range, we need to drop one } D_{i,t}^l \text{ due to multicollinearity. Another multicollinearity that arises in (3) does not appear here because unit fixed effects are dropped.}\]
We can set $CATT_{E_i,-T}$ through $CATT_{E_i,-1}$ to zero if the no anticipation assumption holds. For relative times $\tilde{l}$ excluded from (3), we assume $CATT_{E_i,\tilde{l}}$ are zero. For indices that are out of range, $e + l < 0$ or $e + l > T$, the above expression is well-defined even though $CATT_{e,l}$ is undefined because the corresponding $D_{i,t}^l = 0$ for all $t$. Similarly, if there are never-treated units with $E_i > T$, $D_{i,t}$ is set to a zero vector for them so the above expression is well-defined regardless of their $CATT_{e,l}$.

Using this expression, the expression for population regression coefficients $\mu_l$ simplifies to

$$
\left( \sum_{t=1}^{T} E[D_{i,t} D_{i,t}^T] \right)^{-1} \left( \sum_{t=1}^{T} E[D_{i,t} D_{i,t}^T (E_i)] \right)
$$

Following the same argument in the proof for Proposition 2, we can write $\mu_l$ as a weighted average

$$
\sum_{0 \leq e \leq T-l} \omega_{e,l}^l (CATT_{e,l} - CATT_{e,l-1}) + \sum_{l' \neq l} \sum_{0 \leq e \leq T-l'} \omega_{e,l'} (CATT_{e,l'} - CATT_{e,l'-1}),
$$

where $\sum_{-l_0 \leq e \leq T-l_0} \omega_{e,l}^l = 1$ and for each $l' \neq l$, $\sum_{-l' \leq e \leq T-l'} \omega_{e,l'} = 0$. □

**Special Case: Homogeneous Treatment Effects Across Cohorts.** In the special case of homogeneous effects across cohorts, $CATT_{e,l}$ is constant across $e$ for a given $l$ and in particular is equal to $CATT_l$. The population regression coefficients thus recover causally interpretable estimates; expression (18) simplifies so that $\mu_l = CATT_l - CATT_{l-1}$, the effect on growth in outcome during the $l^{th}$ periods from treatment.

### C Results on FE models with only time fixed effects

Consider the following variation to FE models.

$$
Y_{i,t} = \lambda_t + \gamma D_{i,t} + \xi_{i,t}
$$

(19)

$$
Y_{i,t} = \lambda_t + \sum_{l=-T}^{T} \mu_l D_{i,t}^l + \nu_{i,t}
$$

(20)

These differ from the main FE models discussed in our paper only in that the unit fixed effects are dropped. These models are commonly used when instead of the parallel trends assumption, researchers assume that the timing of treatment is random.
Assumption 6. (Random treatment timing) Baseline outcome is mean independent of the event time 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 is equal to \( E[Y_{i,t}^\infty] \).

Below we show similar negative results on this version of FE models: specifications (19) and (20) estimate some weighted averages of \( CATT_{e,l} \). The weights in specifications (19) are non-negative but these weights do not correspond to any intuitive weights as they are proportional to treatment variance. The weights in specifications (20) can be negative.

**Proposition 10.** Denote \( D_{.,e+l} \equiv \text{Pr}\{E_i \leq e + l\} \) and \( \hat{D}_{i,t} = D_{i,t} - D_{.,t} \). Under Assumption 6 and 2, the population regression coefficient \( \gamma \) from (19) is

\[
\gamma = \sum_{e=0}^{T} \sum_{l=0}^{T-e} \omega_{e,l} CATT_{e,l} \]

where the weight is non-negative \( \omega_{e,l} \equiv \text{Pr}\{E_i = e | E_i \leq e + l\} \frac{D_{.,e+l}(1-D_{.,e+l})}{\sum_{l'=1}^{T} E[D_{.,l}^2]} \) and these weights sum up to one.

**Proof.** The proof of this proposition is almost identical to the proof of Proposition 8, so for brevity we do not included it here.

**Proposition 11.** Under Assumption 6 only, the population regression coefficient on the indicator for \( l \) periods from treatment from (20) is a weighted average of \( CATT_{e,l} \) for all \( e \) as well as \( CATT_{e,l'} \) for all \( e \) and \( l' \neq l \),

\[
\mu_l = \sum_{0\leq e \leq T-l} \omega_{e,l} CATT_{e,l} + \sum_{l'\neq l} \sum_{0\leq e \leq T-l'} \omega_{e,l'} CATT_{e,l'} , \tag{21}
\]

where \( \sum_{-l \leq e \leq T-l} \omega_{e,l} = 1 \) and for each \( l' \neq l \), \( \sum_{-l' \leq e \leq T-l'} \omega_{e,l'} = 0 \).

The weight \( \omega_{e,l'} \) is the projection coefficient on \( D_{i,t} \) from regressing \( \mathbf{1}\{E_i = e\} \cdot D_{i,t}^{l'} \) on time fixed effects and the rest of relative time indicators in (20), \( \{D_{i,t}^{l'}\}_{l' \neq l} \).

**Proof.** The proof of this proposition is almost identical to the proof of Proposition 9, so for brevity we do not included it here.
D Simulation design

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

$$Y_{i,t} = i + t + \sum_{e=1}^{3} \sum_{l=1-e}^{T-e} \delta_{e,l} \mathbb{1}\{E_i = e\} \cdot D_{i,t}^l + \epsilon_{i,t}$$

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 $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$. First, we set $\delta_{e,t<0} = 0 \ \forall e$ for the case of no anticipatory behavior. For each simulation, we estimate

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-1}^{2} \mu_l D_{i,t}^l + \epsilon_{i,t} \hspace{1cm} (22)$$

While $\delta_{u,-1} = 0 \ \forall u$ and hence $CATT_{e,-1} = 0$ for all $e$, the estimates of $\hat{\mu}_{-1}$ are usually negative as shown in Figure 3, which would suggest a negative pre-trend when none exists.

Figure 3: Simulated Distribution of Lead Estimates with No Anticipatory Behavior

Notes: This figure plots the histogram of $\hat{\mu}_{-1}$, the FE estimate 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.

We then set $\delta_{e,-1} = 3 \ \forall e$ to reflect anticipatory behavior in the period right before treatment, and thus $CATT_{e,-1} = 3$ for all $e$. For each simulation, we re-estimate specification (22). The estimates of $\hat{\mu}_{-1}$ are now indistinguishable from zero as shown in Figure 4, which would suggest a lack of pre-trend when one does exist.
Figure 4: Simulated Distribution of Lead Estimates with Anticipatory Behavior

Notes: This figure plots the histogram of $\hat{\mu}_{-1}$, the FE estimate for the treatment effect one period before treatment, across 1000 simulated samples. The true underlying treatment effects one period before treatment are set to be three for all cohorts in these simulations.

E Replicating Dobkin et al. (2018)

While our empirical application also analyzes the economic consequences of hospitalizations, a direct comparison between our results and Dobkin et al. (2018) is not straightforward even when we consider their two-way fixed effects specifications. Our estimation differs from Dobkin et al. (2018) in the following five ways which we discuss further below.

1. Dobkin et al. (2018) restrict their analysis sample to observations up to three waves prior to the hospitalization and three waves after the hospitalization when relative time in the original sample ranges from nine waves prior to the hospitalization to nine waves after the hospitalization.

2. Dobkin et al. (2018) include in their analysis all waves of HRS (waves 1-11) and do not restrict to a sample that is balanced in calendar time.

3. Dobkin et al. (2018) focus on a subsample of non-elderly insured individuals (aged 50 to 59).

4. Dobkin et al. (2018) use a linear time trend in relative time instead of including leads indicators ($D_{i,t}^l$ with $l < 0$).
5. Dobkin et al. (2018) weight their estimates using HRS survey weights. We believe difference (1), the exclusion of some relative times, goes the furthest in explaining the differences in our estimates. In their dynamic specification, the only relative time indicators included are $D_{i,t}^l$ for $-3 \leq l \leq 3$ and the modified dynamic specification is

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-3,l\neq-1}^{3} \mu_l D_{i,t}^l + \epsilon_{i,t}$$

on a sample of observations within three waves from hospitalization: $|E_i - t| \leq 3$ for each $t$. To understand the implications of this restriction on population regression coefficients $\mu_l$, we can take a similar approach in the proof for Proposition 2 and write the $\mu_l$ as

$$(\mu_l)_{-3 \leq t \leq 3, l \neq -1} = \left( \sum_{t=0}^{T} E[\ddot{D}_{i,t}D_{i,t}^l | |E_i - t| \leq 3] \right)^{-1} \left( \sum_{t=0}^{T} E[\ddot{D}_{i,t}Y_{i,t} | |E_i - t| \leq 3] \right)$$

where $\ddot{D}_{i,t} = \begin{pmatrix} \ddot{D}_{i,t}^{-3} \\ \vdots \\ \ddot{D}_{i,t}^{3} \end{pmatrix}$ and again, $\dddot{X}_{i,t}$ is time- and cross-sectional demeaned version of $X_{i,t}$. Note that due to the sample restriction, we need to condition on $|E_i - t| \leq 3$ at each $t$. Following a similar argument as in the proof for Proposition 2, we see that $\mu_l$ is a weighted average of $CATT_{e,l}$ for all $e$ as well as $CATT_{e,l'}$ for all $e$ and $l' \neq l$, $-3 \leq l' \leq 3$. Excluding these relative times eliminates the influence on estimates of $\mu_l$ from simultaneously estimating $CATT_{e,l'}$ for $l' < -3$ and $l' > 3$. Thus, if they were to estimate a dynamic specification on the original sample with all possible relative time indicators, the estimates $\mu_{-3 \leq l \leq 3, l \neq -1}$ could change as they are affected by $CATT_{e,l'}$ for $l' < -3$ and $l' > 3$. We think this sample restriction is the main issue that prevents a direct comparison between our approaches.

The other four difference have a smaller influence on our estimates. Note that we do not use survey weights because it would complicate the estimation of weights $\omega_{e,l}$ in the expression of $\mu_l$. This means we include additional observations with zero survey weights that are excluded from estimation in Dobkin et al. (2018).

To corroborate our explanations, we can replicate exactly the specifications of Dobkin et al. (2018) and then see how estimates change with each difference. In Table 3, Column
(1) replicates the results of Dobkin et al. (2018). In Column (2), we reintroduce time fixed effects instead of a linear trend (difference 4) and do not use survey weights (difference 5), which generally reduces the magnitude of the estimates. In Column (3), we further relax their sample restriction from difference (1) by including all observation and thus all relative time indicators. Comparing results in Column (2) and (3) allows us to focus on the impact of including observations from additional relative time.
Table 3: Comparison with Dobkin et al. (2018)

(a) Out-of-pocket Medical Spending

<table>
<thead>
<tr>
<th>$l$ wave relative to index hospitalization</th>
<th>Dobkin et al. (2018) $(1)$</th>
<th>(4) and (5) $(2)$</th>
<th>Difference $(3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,352</td>
<td>1,936</td>
<td>2,147</td>
</tr>
<tr>
<td></td>
<td>(288)</td>
<td>(224)</td>
<td>(195)</td>
</tr>
<tr>
<td>1</td>
<td>1,434</td>
<td>959</td>
<td>1,350</td>
</tr>
<tr>
<td></td>
<td>(372)</td>
<td>(297)</td>
<td>(256)</td>
</tr>
<tr>
<td>2</td>
<td>1,395</td>
<td>963</td>
<td>1,528</td>
</tr>
<tr>
<td></td>
<td>(513)</td>
<td>(386)</td>
<td>(326)</td>
</tr>
<tr>
<td>3</td>
<td>1,706</td>
<td>1,261</td>
<td>2,030</td>
</tr>
<tr>
<td></td>
<td>(626)</td>
<td>(508)</td>
<td>(440)</td>
</tr>
<tr>
<td>Observations</td>
<td>13,286</td>
<td>13,879</td>
<td>20,077</td>
</tr>
</tbody>
</table>

(b) Labor Earnings

<table>
<thead>
<tr>
<th>$l$ wave relative to index hospitalization</th>
<th>Dobkin et al. (2018) $(1)$</th>
<th>(4) and (5) $(2)$</th>
<th>Difference $(3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-3,440</td>
<td>-1,528</td>
<td>-4,814</td>
</tr>
<tr>
<td></td>
<td>(1970)</td>
<td>(1087)</td>
<td>(2483)</td>
</tr>
<tr>
<td>1</td>
<td>-10,842</td>
<td>-7,284</td>
<td>-13,904</td>
</tr>
<tr>
<td></td>
<td>(3146)</td>
<td>(1661)</td>
<td>(5217)</td>
</tr>
<tr>
<td>2</td>
<td>-14,467</td>
<td>-10,250</td>
<td>-20,362</td>
</tr>
<tr>
<td></td>
<td>(4106)</td>
<td>(2286)</td>
<td>(7757)</td>
</tr>
<tr>
<td>3</td>
<td>-17,418</td>
<td>-13,541</td>
<td>-27,069</td>
</tr>
<tr>
<td></td>
<td>(5083)</td>
<td>(2907)</td>
<td>(10306)</td>
</tr>
<tr>
<td>Observations</td>
<td>13,286</td>
<td>13,879</td>
<td>20,077</td>
</tr>
</tbody>
</table>

Notes: The number of observations increases from Column (1) to (2) because we include observations with zero survey weights. The number of observations increases from Column (2) to (3) because we include observations from all relative time periods, not just observations within three relative time periods from hospitalization. Although not apparent, after applying the linear transformation described in online Appendix C of Dobkin et al. (2018), results in Column (1) match with results in Column (2) of Table 4 in Dobkin et al. (2018). Results in Column (3) are of larger magnitudes compared to (2) but have the reasonable signs. The magnitude becomes larger because the latter specification assigns non-zero weight to treatment effects from relative time periods beyond three. Thus, these estimates
are affected by treatment effects from these relative time periods. By restricting the number of relative time periods as in Column (2), these weights are mechanically zero.

Recall that in the main text, we show the dynamic specification with all possible relative time indicators return estimates of the opposite sign. One reason there is no sign reversion here is the underlying weights in Column (3) are much smaller. For example, for the estimate of treatment effects right after hospitalization \( \hat{\mu}_0 \), the weights from other relative time periods \( l \neq 0 \) are negligible as \( \max_{e,l} |\omega_{e,l}^0| = 0.21 \), compared to a weight of 3 in the main text. This means \( \hat{\mu}_0 \) is not affected as much by treatment effects from other relative time periods \( l \neq 0 \).

**F Including covariates and doubly robust scores**

When there are covariate-specific time trends, we need to modify the unconditional parallel trends assumption to the following one. Denote the confounding covariates by \( X_{i,t} \).

**Assumption 7.** Conditional parallel trends in baseline outcome. \( E[ Y_{i,t}^\infty - Y_{i,s}^\infty | E_i = e, X_{i,t} ] \) is the same for all \( e \) and all \( s,t \) and is equal to \( E[ Y_{i,t}^\infty - Y_{i,0}^\infty | X_{i,t} ] \).

We maintain the no anticipation assumption, but do not impose any restrictions on treatment effects.

The causal parameter of interest is still

\[
\theta_0 := \text{CATT}_{e,l}.
\]

Suppose we know the covariate-specific time trends \( g_0^\infty (X_{i,t}) := E[ Y_{i,t}^\infty - Y_{i,0}^\infty | X_{i,t} ] \). Then we can form a consistent estimator for \( \theta_0 \) by \( E_N [ Y_{i,t} - Y_{i,0} - g^\infty (X_{i,t}) | E_i = t - l ] \). This estimator is an extension to Heckman et al. (1997).

Suppose we know the probability of being in cohort \( e = t - l \) conditional on \( X_{i,t} \), denoted by \( m_0 (X_{i,t}) := Pr \{ E_i = t - l | X_{i,t} \} \), and the probability of not having received treatment by time \( t \) conditional on \( X_{i,t} \), denoted by \( n_0 (X_{i,t}) := Pr \{ E_i > t | X_{i,t} \} \), as well as the share of cohort \( e \), \( m_0 := Pr \{ E_i = t - l \} \). Then we can form a consistent estimator for \( \theta_0 \) by

\[
E_N \left[ \frac{1 \{ E_i = t - l \}}{m_0} (Y_{i,t} - Y_{i,0}) - \frac{1 \{ E_i > t \}}{m_0 \cdot n_0 (X_{i,t})} m_0 (X_{i,t}) (Y_{i,t} - Y_{i,0}) \right].
\]
This inverse propensity score reweighted estimator is an extension to Abadie (2005). A similar version has been proposed by Callaway and Sant’Anna (2018).

Collect

\[ \eta_0 (X_{i,t}) := \left( E[Y_{i,t}^\infty - Y_{i,0}^\infty \mid X_{i,t}], Pr \{ E_i = t - l \mid X_{i,t} \}, Pr \{ E_i > t \mid X_{i,t} \}, Pr \{ E_i = t - l \} \right)^T. \]

Let \( \eta (X_{i,t}) := (g^\infty (X_{i,t}), m (X_{i,t}), n (X_{i,t}), m)^T \) denote the nuisance parameter with true value \( \eta_0 (X_{i,t}) \). In practice, it is likely that we make mistakes in estimating \( \eta (X_{i,t}) \).

We may employ the following doubly robust score as a special case to the doubly robust score derived for cross-sectional models. Define

\[ \psi (W_i, \theta, \eta) := \frac{1}{m} \left\{ E_i = t - l \right\} (Y_{i,t} - Y_{i,0} - g^\infty (X_{i,t})) - \frac{1}{m} \left\{ E_i > t \right\} m (X_{i,t}) (Y_{i,t} - Y_{i,0} - g^\infty (X_{i,t})) - \theta \frac{1}{m} \left\{ E_i = t - l \right\} \]

where \( W_i := (Y_{i,t}, Y_{i,0}, X_{i,t}) \). To see that, replace the cross-sectional outcome \( Y_i \) in Chernozhukov et al. (2017) with the long difference \( Y_{i,t} - Y_{i,0} \).

We can show that the score satisfies the identification condition \( E[\psi (W_i, \theta_0, \eta_0)] = 0 \) and the Neyman orthogonality condition \( \partial_\eta E[\psi (W_i, \theta_0, \eta)] \bigg|_{\eta=\eta_0} = 0 \). This score is robust to small mistakes in \( \eta (X_{i,t}) \). The estimation can be done by \( K \)-fold cross-fitting as described in Chernozhukov et al. (2017).

For \( X_{i,t} = 1 \), we are back to the unconditional parallel trends assumption. The DID estimator with \( s = 0 \) and \( C = \{ c : t < c \} \) is the root \( \theta \) of the above doubly robust score in the sample, \( \mathbb{E}_N [\psi (W_i, \theta, \hat{\eta})] = 0 \), where \( \hat{\eta} \) is the sample analog of \( \eta_0 \).