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

Sarah Abraham† and Liyang Sun‡  

July 3, 2019  

Abstract  

In event studies where the timing of treatment varies across units, researchers often estimate dynamic treatment effects using fixed effects regressions that include leads and lags of the treatment. We show that these regressions could produce causally uninterpretable results as they assign non-convex weights to cohort-specific treatment effects. In particular, this invalidates the Granger causality test in which coefficients on leads provide evidence for lack of or existence of pre-trends. We propose alternative estimators, and illustrate the shortcomings of fixed effects estimators in comparison to our proposed estimators through an empirical application on the consequences of hospitalization.  

Keywords: Differences-in-Differences, Fixed Effects, Granger Causality Test  

*We are grateful to Isaiah Andrews, Amy Finkelstein, Anna Mikusheva, and Heidi Williams for their guidance and support. We thank Alberto Abadie, Jonathan Cohen, Nathan Hendren, Peter Hull, Guido Imbens, Kevin Kainan Li, Whitney Newey, James Poterba, and Pedro H. C. Sant’Anna 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/.  

†Cornerstone Research, 699 Boylston St, Boston, MA 02116.  
‡Department of Economics, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139. Corresponding author; lsun20@mit.edu
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 by including leads and lags of treatment as regressors. Event studies are used to estimate the average dynamic effects of discrete shocks and nontransient treatments. For example, Dobkin et al. (2018) estimated the effect of a hospitalization on retiring individuals over time using an event study. While identifying assumptions underlying DID have been developed, there is no formal framework for thinking about identifying assumptions for event studies.

In this paper, we formalize event studies by generalizing the potential outcome framework. We note that units’ exposure to the treatment can be summarized by the onset of their exposure, which we call “event time”. This discrete variable defines the potential outcomes, and categorizes units into multiple cohorts based on when they first receive the treatment.

For event studies to consistently estimate an average dynamic treatment effects, we find three identifying assumptions using definitions derived under the potential outcome framework. The first assumption is a generalized form of parallel trends assumption. The second assumption requires no anticipation of the treatment. The third assumption imposes no variation in treatment effects across cohorts. Although previous literature has discussed each assumption in a separate context, we focus on the role of these three assumptions on event studies.

While violations of the first and second assumptions invalidate event studies, we find violation of the third assumption, namely treatment effect heterogeneity, makes event studies harder to interpret. The implication of treatment effect heterogeneity in event studies parallels several recent econometrics papers focusing on DID specifications (Athey and Imbens, 2018; Borusyak and Jaravel, 2017; de Chaisemartin and D’Haultfœuille, 2018; Goodman-Bacon, 2018; Imai and Kim, 2019; Słoczyński, 2018).

We propose an alternative method that is robust to variation in treatment effects across cohorts (i.e. violation of the third assumption). This method estimates the dynamic effect for each cohort, and then calculates the average of cohort-specific estimates. We derive a consistent variance
estimator for this alternative estimator. Researchers can easily construct a pointwise confidence interval for average dynamic treatment effects using this variance estimator.

We illustrate its use using data from Dobkin et al. (2018). While event studies estimates can fall out of convex hull of the underlying treatment effects, estimates using our alternative method, by construction, are weighted averages of the underlying treatment effects with weights representative of cohort share.

In the next section, we formally state the three identifying assumptions for event studies. Section 3 derives the estimands of event studies allowing for failure of identifying assumptions. 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.

2 Setting

The setting we consider consists of a random sample of \( N \) independent units who first receive treatment at different times. 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 \( t \).

To estimate the treatment effects in an event study framework, researchers often use linear two-way (unit and time) fixed effects (FE) regressions with leads and lags of treatment. These regressions often take the following form

\[
Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-K}^{-2} \mu_l D_{i,t}^l + \sum_{l=0}^{L} \mu_l D_{i,t}^l + \nu_{i,t}
\]

where \( D_{i,t}^l := \mathbf{1}\{t - E_i = l\} \) is an indicator for being \( l \) time periods relative to \( i \)'s initial treatment (\( l = 0 \) is the year of initial treatment). We include both unit and time fixed effects, \( \alpha_i \) and \( \lambda_t \). We are interested in the assumptions needed for the \( \mu_l \) parameters to estimate the average treatment effects of units who receive treatment at different times.
Note that we do not include all possible relative time indicators in specification (1). We need to exclude at least two relative period indicators due to multicollinearities. These collinearities are discussed by Borusyak and Jaravel (2017). In terms of which relative time to exclude, common practice focuses on interpretability, normalizing relative to the period prior to treatment, and also excluding indicators for periods distant from the treatment. In our notation, we drop $D_{i,t}^{-1}$, and exclude indicators for periods more than $K$ periods before treatment and more than $L$ periods after treatment.

2.1 A potential outcome framework

To study assumptions for specification (1) to identify average treatment effects, we need formal definitions of potential outcomes. The original potential outcome framework (Neyman-Rubin model) was proposed for binary treatment in cross-sectional studies. We generalize this framework to event studies where we may categorize units into multiple cohorts based on when they first receive the treatment.

In event studies, the treatment path of a unit can be uniquely characterized by the time period of the initial treatment, which we call “event time”. We denote this discrete random variable by $E_i = \min\{t : D_{i,t} = 1\}$. This is because 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 $s < t$.

We thus define “treatment effect” for a given event time $e$ as the difference between the baseline outcome and the potential outcome: $Y_{i,t}^e - Y_{i,t}^\infty$. Here $Y_{i,t}^e$ denotes the potential outcome in period $t$ under a hypothetical treatment path $e$; $Y_{i,t}^\infty$ denotes the outcome if unit $i$ never receives treatment, which we call the “baseline outcome”. We observe each unit under only a single treatment path $E_i$, so the observed outcome for unit $i$ is simply $Y_{i,t} = Y_{i,t}^{E_i}$. 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:

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

(2)

Additionally, we define the cohort-specific average treatment effects on the treated in relative time.\(^1\)

**Definition 1.** We define the cohort-specific average treatment effects on the treated \( l \) periods from initial treatment as 

\[ \text{CA TT}_{e,l} = E[ Y_{i,e+l}^{e} - Y_{i,e+l}^{\infty} \mid E_i = e]. \]

Each \( \text{CA TT}_{e,l} \) represents the average treatment effect \( l \) periods from initial treatment for the cohort of units first treated at time \( e \). Recall that event time allows us to define cohort \( e \) as the set of units for which \( E_i = e \). We denote units never treated as \( E_i = \infty \). 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. Relative time allows us to compare across cohorts while holding their exposure to the treatment constant.

### 2.2 Identifying assumptions

With the above definitions, we formalize three identifying assumptions for event studies. The first assumption is a generalized form of parallel trends assumption. The second assumption requires no anticipation of the treatment. The third assumption imposes no variation across cohorts.

**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

---

\(^1\)Callaway and Sant’Anna (2018) propose a similar definition, “group-time average treatment effect”, which states \( \text{CA TT}_{e,t} \) in calendar time.
Assumption 2. (No anticipatory behavior at \( l \) periods prior to treatment.) For \( l < 0 \), 
\[
Y^{e}_{i,e+l} = Y^{\infty}_{i,e+l}
\]
for all \( e \in supp(E_i) \).

This assumption requires potential outcomes \( l \) periods before treatment to be equal to the baseline outcome. If this assumption holds for all \( l < 0 \), there are no pre-trends since \( CATT_{e,l} = 0 \) for all \( e \in supp(E_i) \) and all \( l < 0 \). 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 represent 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 \( Y^{1}_{i,0} \) and we cannot assume \( Y^{1}_{i,0} \) equals \( Y^{\infty}_{i,0} \).

Assumption 3. (Treatment effect homogeneity). For each lag of treatment \( l \geq 0 \), \( CATT_{e,l} \) does not depend on cohort \( e \) and is equal to \( CATT_{l} \).

Assumption 3 does not preclude dynamic treatment effects. Cohorts can share the same profile of non-constant treatment effect.

Assumption 3 is violated when different cohorts have different profiles of dynamic treatment effect. In event studies, heterogeneity in treatment effects across cohorts could arise for many reasons. For example, cohorts may differ in their covariates, which affect how they respond to treatment. We will explore a concrete example in our application: if treatment effects differ with age and 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).
3 Estimators from linear two-way FE regression

We show what specification (1) estimates when each of the three identifying assumptions fail. While violations of the first two identifying assumptions invalidate event studies, a violation of the last identifying assumption only makes event studies harder to interpret.

Recall that specification (1) takes the following form

\[ Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-K}^{L} \mu_l D_{i,t}^{l} + \nu_{i,t} \]  

(3)

where \( D_{i,t}^{l} := 1(t - E_i = l) \) is an indicator for being \( l \) time periods relative to \( i \)’s initial treatment \((l = 0 \text{ is the year of initial treatment})\). We denote by \( \mu_l \) the estimands of specification (1), i.e. the population regression coefficients. Their corresponding estimators are denoted by \( \hat{\mu}_l \) respectively.

3.1 Interpreting the coefficients under no assumptions

First we show without any assumptions, we can write \( \mu_l \) as a linear combination of differences in trends.

**Proposition 1.** The population regression coefficient on the indicator for \( l \) periods from treatment is a linear combination of differences in trends at relative time \( l \) as well as at other relative times \( l' \neq l \),

\[
\mu_l = \sum_{e} \omega_{e,l}^{(l)} \left( E[Y_{i,e+l} - Y_{i,0}^{\infty} | E_i] - E[Y_{i,e+l}^{\infty} - Y_{i,0}^{\infty}] \right) + \sum_{l' \neq l} \sum_{e} \omega_{e,l'}^{(l)} \left( E[Y_{i,e+l'} - Y_{i,0}^{\infty} | E_i] - E[Y_{i,e+l'}^{\infty} - Y_{i,0}^{\infty}] \right). 
\]

(4)

The weights \( \omega^{(l)} \) associated with each \( \mu_l \) sum up to zero, but we defer a more detailed interpretation of the weights to Proposition 2.

The above proposition is a direct result of fixed effects models.
3.2 Interpreting the coefficients under parallel trends assumption only

Second we show with Assumption 1, we can write $\mu_l$ as a weighted sum of $CATT$s in Proposition 2. This is because under Assumption 1, we can add causal interpretation to the above expression as

$$E[Y_{i,e+l} - Y_{i,0}^\infty | E_i] - E[Y_{i,e+l}^\infty - Y_{i,0}^\infty] = CATT_{e,l} + E[Y_{i,t}^\infty - Y_{i,0}^\infty] - E[Y_{i,t} - Y_{i,0}]$$

for $t = e + l$.

The weights are in general hard to interpret and need not be positive without additional assumptions that restrict treatment effects to be homogeneous. To our knowledge, no paper has yet studied the linear two-way dynamic FE specification when treatment effects are heterogeneous.

Specifically, the FE estimands for $l$ periods relative to treatment can be written as non-convex averages of not only $CATT_{e,l}$ from that period, but also $CATT_{e,l'}$ from other periods. This implies that $\mu_l$ could fall outside the convex hull of $CATT_{e,l}$, which means it is possible for $\mu_l$ to be of the opposite sign to all $CATT_{e,l}$, the cohort-specific average treatment effects from $l$ periods after initial treatment.

**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 $CATT_{e,l'}$ from other relative times $l' \neq l$,

$$\mu_l = \sum_e \omega_{e,l}^{(1)} CATT_{e,l} + \sum_{l' \neq l} \sum_e \omega_{e,l}^{(l)} CATT_{e,l'}.$$  \hspace{1cm} (5)

The weights on $CATT_{e,l}$ sum up to one $\sum_e \omega_{e,l}^{(l)} = 1$. The weights on $CATT_{e,l'}$ from each other relative times $l' \neq l$ that is included in the regression sum up to zero $\sum_e \omega_{e,l'}^{(l)} = 0$. For the rest relative times $l' \neq l$ that are excluded from the regression, we have $\sum_{l' < -K, > L} \sum_e \omega_{e,l'}^{(l)} = -1$.

The weight $\omega_{e,l'}^{(l)}$ is the population regression coefficient in front of $D_{i,t}^{l'}$ from regressing $1 \{ E_i = e \} \cdot D_{i,t}^{l'}$ on the relative time indicators included in (1) i.e. $\{ D_{i,t}^{l'} \}$ and two-way fixed effects.
3.3 Interpreting the coefficients under parallel trends and no anticipation assumptions

Even when we further assume no pre-trends, another implication of Proposition 2 is that specification (1) can yield non-zero pre-trend estimates. This finding invalidates the Granger causality test, in which non-zero coefficients on leads provide evidence for existence of pre-trends. We state these results more formally in Proposition 3 below.

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

$$\mu_l = \sum_{l' \geq 0} \sum_e \omega_{e,l'}^{(l)} \text{CATT}_{e,l'},$$

(6)

where for each $l'$ included in the regression, the weights sum up to zero $\sum_e \omega_{e,l'}^{(l)} = 0$. For the rest relative times $l' \neq l$ that are excluded from the regression, we have $\sum_{l' > L} \sum_e \omega_{e,l'}^{(l)} = -1$.

**Invalidity of Granger causality test.** Under the no anticipatory behavior assumption, cohort-specific treatment effects prior to treatment are all zero: $\text{CATT}_{e,l} = 0$ for all $l < 0$. Therefore, any linear combination of these $\text{CATT}_{e,l}$ is also zero. However, due to the influence of post-treatment $\text{CATT}_{e,l' \geq 0}$, this estimand $\mu_l$ is not necessarily zero.

In short, when effects are not homogenous across cohorts, it is problematic to interpret non-zero estimates on $D_{l,t}^l$ for $l < 0$ as evidence for pre-trends because variation in post-treatment $\text{CATT}_{e,l}$ can affect the lead FE estimators. In Online Appendix B we present through simulation two examples where FE estimates do not appropriately summarize underlying pre-trends. See Callaway and Sant’Anna (2018) for alternative tests for pre-trends that do not suffer from this drawback.
3.4 Interpreting the coefficients under all three assumptions

If we further assume homogenous treatment effects as stated in Assumption 3, $CATT_{e,l}$ is constant across $e$ for a given $l$. Adding this assumption to Proposition 3, we find the FE dynamic estimators do recover causally interpretable estimates. The lagged FE estimand $\mu_l$ for $l \geq 0$ is equal to $CATT_{e,l}$ minus a linear combination of $CATT_{e,l'}$ for $l'$ excluded from the regression.

**Proposition 4.** If Assumption 1 (parallel trends), Assumption 2 for all $l < 0$ (no anticipatory behavior), and Assumption 3 (treatment effect homogeneity) all hold, the population regression coefficient on the indicator for $l$ periods after the treatment is equal to $CATT_l$, minus a linear combination of $CATT$ from the excluded relative periods

$$\mu_l = CATT_l - \sum_{l' \geq L} \omega_{l'}^{(l)} CATT_{l'}$$

(7)

The weights for relative periods $l' \geq L$ that are excluded from the regression sum up to minus one, i.e. $\sum_{l' > L} \omega_{l'}^{(l)} = -1$.

4 Alternative estimation method

We propose a new estimation method that is robust to failure of the last identifying assumption, namely when treatment effects vary across cohorts. 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 with heterogeneous treatment effects, $CATT_{e,l}$ vary by $e$ so the error term $\epsilon_{i,t}$ 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 $E_i$. Since the relative time indicator $D_{i,l}$ is a deterministic function of $E_i$, the error term is correlated with the regressor $D_{i,l}$ as well. This observation hints at using an interacted specification saturated in relative time indicators $D_{i,l}$ 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.

4.1 Description of the alternative method

Our proposed method estimates the average treatment effect for each cohort, and for each relative time \( l \), it averages over the cohort-specific estimates associated with relative time \( l \). 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.

4.1.1 Difference-in-differences for \( CATT_{e,l} \)

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

\[
\hat{\delta}_{e,l} = \frac{\mathbb{E}_N[(Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i = e\}]}{\mathbb{E}_N[1\{E_i = e\}]} - \frac{\mathbb{E}_N[(Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i \in C\}]}{\mathbb{E}_N[1\{E_i \in C\}]} \tag{8}
\]

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 5.** 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 Assumption 2 for all \( l < 0 \), the DID estimator
is an unbiased and consistent estimator for $CATT_{e,l}$.

It is also possible to relax the parallel trends assumption to allow the timing of treatment to depend on covariates. One can estimate $CATT_{e,l}$ consistently based on 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 $CATT_{e,l}$ 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 2 of Online Appendix A). First, we estimate $CATT_{e,l}$ using a linear two-way fixed effects specification that interacts relative time indicators with cohort indicators:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \leq 1} \delta_{e,l}(1\{E_i = e\} \cdot D_{i,t}^j) + \epsilon_{i,t}$$

on observations from $t = 0, \ldots, T - 1$ and $e = 1, \ldots, T$. (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 $CATT_{0,l}$ 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 $CATT_{e,l}$ estimates from step 1 to form average treatment effect estimates with weight estimates from step 2.

The coefficient estimator $\widehat{D}_{e,l}$ from Regression (9) is a consistent DID estimator for $CATT_{e,l}$ in the form of (8) with $s = e - 1$ and $C = \{T\}$. (Among all possible DID estimators for $CATT_{e,l}$, we propose Regression (9) because it is a natural extension to the dynamic FE specification (1) by saturating in relative time indicators and cohort indicators.) Thus, the IW estimator is consistent for a weighted average of $CATT_{e,l}$ 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 $\text{CATT}_{e,l}$.

With a few standard assumptions (which we present as Assumption 4 in Online Appendix A) on Regression (9), we can show that each IW estimator is asymptotically normal and derive its asymptotic variance. The large sample approximation allows us to estimate the variance of IW estimators directly without relying on bootstrap as in Callaway and Sant’Anna (2018).

4.3 Comparing FE and IW estimators

Assuming homogeneous treatment effects $\text{CATT}_{e,l}$ is constant across $e$ for a given $l$, and can be consistently estimated by both the FE and IW estimators. The FE estimators may have smaller variance when they are consistent because specification (1) estimate fewer coefficients. In particular, the FE estimators are more efficient when regressors in specification (1) are linear transformations of regressors in specification (9) and errors are homoskedastic and serially uncorrelated. In the presence of heterogeneous treatment effects, only the saturated specification (9) 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.

5 Applications

We illustrate the empirical importance of our findings by estimating specification (1) using data from Dobkin et al. (2018). Dobkin et al. (2018) use an event study approach to examine the economic consequences of hospital admissions for adults. While they consider a slightly different specification, their data provides a good context to illustrate issues with two-way dynamic FE specifications.

The issues we identify on two-way FE regressions do not directly apply to results in Dobkin et al. (2018). In particular, they balance their analysis sample in relative times, keeping only observations up to three waves prior to the hospitalization and three waves after the hospitalization. (Relative time in the original sample ranges from nine waves prior to the hospitalization to nine
waves after the hospitalization.) By restricting the number of relative time periods, the weights on $CATT$ from excluded periods are mechanically zero, which eliminate the issues we illustrate below.

### 5.1 Setting and data

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

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

Here $i$ indexes an individual, and $t$ indexes survey wave ($T = 4$) and is normalized to zero for wave 7, the first wave in our sample. 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. We have a slightly lower fraction of white in our sample, but otherwise have a similar sample to Dobkin et al. (2018).

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

### 5.2 Assumptions and estimation

**Parallel trends (Assumption 1).** Hospitalization is likely to be earlier among sicker individuals with high out-of-pocket medical spending and low labor earnings, even when restricted to individuals who were ever hospitalized. Thus, it is not plausible that the baseline outcome $Y_{i,t}^\infty$ is mean independent of the timing of hospitalization. The parallel trends assumption is more plausible as it allows the timing to depend on unobserved time-invariant characteristics such as chronic disease.

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

**Treatment effect heterogeneity (Assumptions 3).** For the two outcomes we study, we think the assumption of 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 estimate the following two specifications without survey weights. We estimate the linear two-way dynamic FE specification (1) as

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{l \neq -1} \mu_l D_{i,l}^t + u_{i,t}$$  \hspace{1cm} (10)

for \( t = 0, \ldots, 4 \). We estimate the saturated specification (9) as

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \in \{1,2,3\}} \sum_{l \neq -1} \delta_{e,l} 1\{E_i = e\} \cdot D_{i,l}^t + \epsilon_{i,t}$$  \hspace{1cm} (11)

for \( t = 0, \ldots, 3 \). We drop \( t = 4 \) from (11) because everyone has been hospitalized by \( t = 4 \), and \( CATT_{e,l} \) in \( t = 4 \) are not identified. For the IW estimators \( \nu_l \), our proposed alternative estimators to the FE estimators \( \mu_l \), we take weighted averages of \( \delta_{e,l} \) with weights that are sample share of each cohort \( e \) across cohorts that are observed \( l \) periods after hospitalization. We estimate the above specifications with \( Y_{i,t} \) 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 \( H_0 : \mu_{-3} = \mu_{-2} = 0 \) or \( H_0 : \delta_{3,-3} = \delta_{3,-2} = \delta_{2,-2} = 0 \).

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 might not be in the convex hull of the \( CATT_{e,l} \) estimates and thus do not have a
causal interpretation.

Recall by Proposition 2, the FE estimand $\mu_l$ is

$$\sum_e \omega_{e,l}^{(l)} CATT_{e,l} + \sum_{l' \neq l} \sum_e \omega_{e,l'}^{(l)} CATT_{e,l'}. \quad (12)$$

The point estimate $\hat{\mu}_l$ is the sample analog of expression (12).

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 $\omega_{e,l}^{(0)}$ by regressing $1\{E_i = e\} \cdot D_{l,t}^1$ on the relative time indicators in specification (10) i.e. $\{D_{l,t}^1\}_{l=-2}^3$ and two-way fixed effects. The coefficient estimator of $D_{l,t}^0$ in such regression, $\tilde{\omega}_{e,l}^{(0)}$, consistently estimates $\omega_{e,l}^{(0)}$. The point estimate $\hat{\mu}_0$ is the product of these estimated weights $\tilde{\omega}_{e,l}^{(0)}$ and $\tilde{\delta}_{e,l}$. 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. We found no evidence of anticipatory behavior (compared to relative period $-1$), so in the limit, the FE estimator $\hat{\mu}_0$ should not be affected by pre-trends since $CATT_{e,l} = 0$ for all $l < 0$; however in finite samples, their estimates $\hat{\delta}_{e,l}$ 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.

While under non-convex weighting, $\hat{\mu}_0$ still falls in the convex hull of its underlying $CATT_{e,0}$ estimates, $\hat{\mu}_{-2}$ turn out to be outside the convex hull of its underlying $CATT_{e,-2}$ estimates as shown in Table 2. This illustrates again that the FE estimator could lead to causally uninterpretable results.

6 Conclusions

This paper analyzes the behavior of commonly used event studies under heterogeneous treatment effects. We derive the estimands for the linear two-way dynamic FE specifications underlying event studies and show that they are linear combinations of $CATT$. However, the weights can be non-convex, and the FE estimator associated with $l$ periods relative to initial treatment may pick
up spurious terms consisting of treatment effects from periods other than \( l \). This could lead FE estimands to fall outside the convex hull of the underlying \( CATT \) under heterogeneous treatment effects.

Given these negative results on FE estimators, we propose “interaction-weighted” (IW) estimators for estimating dynamic treatment effects. The IW estimators are formed by first estimating \( CATT_{c,l} \) with a regression saturated in cohort and relative time indicators, and then averaging estimates of \( CATT_{c,l} \) across \( e \) at a given \( l \). These \( CATT_{c,l} \) are identified under parallel trends and no anticipation assumptions. 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 \( CATT_{c,l} \) 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. We calculate weights for the FE estimates in this example, and show that the FE estimates can fall outside the convex hull of the underlying \( CATT \) estimates, thus are causally uninterpretable. IW estimates, on the other hand, are weighted averages of the underlying \( CATT \) estimates with weights representative of cohort share.
References


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 $\tilde{\nu}_l$ in circles against relative wave $l$, with their respective pointwise 95% confidence intervals. Both are estimates for the effect of hospitalization at relative wave $l$. The outcome variable is out-of-pocket medical spending in Panel A and labor earnings in Panel B respectively.
Figure 2: Weights $\hat{\omega}_{e,l}^{(0)}$ on cohort specific effect estimates $\hat{\delta}_{e,l}$ 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 $\hat{\delta}_{e,l}$’s, estimates for cohort-specific effects $CATT_{e,l}$’s from all cohorts $e$ and relative waves $l$. This figure plots the weight $\hat{\omega}_{e,l}^{(0)}$ associated with each $\hat{\delta}_{e,l}$ in forming the FE estimate $\hat{\mu}_0$. 
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>656</td>
<td>56</td>
<td>2.29</td>
</tr>
<tr>
<td>Male</td>
<td>656</td>
<td>0.456</td>
<td>0.498</td>
</tr>
<tr>
<td>Year of admission</td>
<td>656</td>
<td>2,007</td>
<td>2.11</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>656</td>
<td>0.122</td>
<td>0.327</td>
</tr>
<tr>
<td>Black</td>
<td>656</td>
<td>0.151</td>
<td>0.358</td>
</tr>
<tr>
<td>White</td>
<td>656</td>
<td>0.742</td>
<td>0.438</td>
</tr>
<tr>
<td>Other race</td>
<td>656</td>
<td>0.107</td>
<td>0.309</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>656</td>
<td>0.05</td>
<td>0.219</td>
</tr>
<tr>
<td>Private</td>
<td>656</td>
<td>0.715</td>
<td>0.452</td>
</tr>
<tr>
<td>Medicare</td>
<td>656</td>
<td>0.072</td>
<td>0.259</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>656</td>
<td>3,302</td>
<td>9,024</td>
</tr>
<tr>
<td>Wave 8 ($E_i \geq 2$)</td>
<td>404</td>
<td>2,355</td>
<td>8,132</td>
</tr>
<tr>
<td>Wave 9 ($E_i \geq 3$)</td>
<td>228</td>
<td>2,056</td>
<td>3,532</td>
</tr>
<tr>
<td>Wave 10 ($E_i = 4$)</td>
<td>65</td>
<td>2,044</td>
<td>4,379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Earnings</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 7 ($E_i \geq 1$)</td>
<td>656</td>
<td>43,810</td>
<td>67,950</td>
</tr>
<tr>
<td>Wave 8 ($E_i \geq 2$)</td>
<td>404</td>
<td>38,944</td>
<td>58,601</td>
</tr>
<tr>
<td>Wave 9 ($E_i \geq 3$)</td>
<td>228</td>
<td>36,274</td>
<td>56,768</td>
</tr>
<tr>
<td>Wave 10 ($E_i = 4$)</td>
<td>65</td>
<td>29,037</td>
<td>46,289</td>
</tr>
</tbody>
</table>

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

23
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>-3</td>
<td>149</td>
<td>591</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(792)</td>
<td>(1273)</td>
<td>(1273)</td>
</tr>
<tr>
<td>-2</td>
<td>203</td>
<td>356</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(480)</td>
<td>(700)</td>
<td>(967)</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>3,013</td>
<td>2,975</td>
<td>2,826</td>
</tr>
<tr>
<td></td>
<td>(511)</td>
<td>(523)</td>
<td>(1038)</td>
</tr>
<tr>
<td>1</td>
<td>888</td>
<td>494</td>
<td>825</td>
</tr>
<tr>
<td></td>
<td>(664)</td>
<td>(562)</td>
<td>(912)</td>
</tr>
<tr>
<td>2</td>
<td>1,172</td>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td></td>
<td>(983)</td>
<td>(1011)</td>
<td>(1011)</td>
</tr>
<tr>
<td>3</td>
<td>1,914</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(1426)</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>-3</td>
<td>-2,642</td>
<td>-8,228</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(3504)</td>
<td>(6594)</td>
<td>(6594)</td>
</tr>
<tr>
<td>-2</td>
<td>-5,089</td>
<td>-7,830</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(3005)</td>
<td>(4139)</td>
<td>(6349)</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>-1,225</td>
<td>-1,175</td>
<td>455</td>
</tr>
<tr>
<td></td>
<td>(2742)</td>
<td>(2866)</td>
<td>(5593)</td>
</tr>
<tr>
<td>1</td>
<td>-7,508</td>
<td>-5,886</td>
<td>-1,670</td>
</tr>
<tr>
<td></td>
<td>(4312)</td>
<td>(2866)</td>
<td>(6500)</td>
</tr>
<tr>
<td>2</td>
<td>-11,102</td>
<td>-10,670</td>
<td>-10,670</td>
</tr>
<tr>
<td></td>
<td>(5976)</td>
<td>(6155)</td>
<td>(6155)</td>
</tr>
<tr>
<td>3</td>
<td>-8,780</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(8332)</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.
Online Appendix

A Proofs

Proof of Propositions 1, 2, 3 and 4

Proof. Partialing out the unit and time fixed effects, Regression (1) is

\[ \hat{Y}_{i,t} = \sum_{l=-K}^{L} \mu_l \hat{D}_{i,t}^l + \sum_{l=0}^{L} \mu_l \hat{D}_{i,t} = Y_{i,t} + u_{i,t} \]

where \( \hat{X}_{i,t} \) is time- and cross-sectional demeaned version of \( X_{i,t} \) i.e. \( \hat{X}_{i,t} = X_{i,t} - \bar{X}_{i,\cdot} \). Note that for any \( \hat{X}_{i,t} \) and \( \hat{Z}_{i,t} \), demeaned versions of \( X_{i,t} \) and \( Z_{i,t} \) respectively, we have \( E[\hat{X}_{i,t} \hat{Z}_{i,t}] = E[X_{i,t} \hat{Z}_{i,t}] = E[\hat{X}_{i,t} Z_{i,t}] \). The population regression coefficient on \( D_{i,t}^l \) from specification (1), i.e. the probability limit of the coefficient estimator \( \hat{\mu}_l \), is then

\[ \mu_l = \mathbf{e}_l' \left( \sum_{t=0}^{T} E[\mathbf{D}_{i,t} \mathbf{D}_{i,t}'] \right)^{-1} \sum_{t=0}^{T} E[\mathbf{D}_{i,t} Y_{i,t}] \]  

(13)

where \( \mathbf{D}_{i,t} = \begin{pmatrix} D_{i,t}^{-K} & \cdots & D_{i,t}^{-2} & D_{i,t}^0 & \cdots & D_{i,t}^L \end{pmatrix}' \) is a column vector collecting \( \hat{D}_{i,t}^l \), and \( \mathbf{e}_l \) is a column vector with 1 in the entry corresponding to the entry of \( \hat{D}_{i,t}^l \) in \( \mathbf{D}_{i,t} \), and 0 otherwise.

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

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

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

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

Under parallel trends, we have \( E[Y_{i,t}^\infty - Y_{i,0}^\infty | E_t] = E[Y_{i,t}^\infty - Y_{i,0}^\infty] \). Then the difference in trends \( E[Y_{i,t}^\infty - Y_{i,0}^\infty | E_t] - E[Y_{i,t}^\infty - Y_{i,0}^\infty] \) is equal to average treatment effects \( E[Y_{i,t} - Y_{i,0}^\infty | E_t] \). The expression for
population regression coefficient \( \mu_l \) simplifies from equation (13) to

\[
\mu_l = e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} E[Y_{i,t} - Y_{i,t}^\circ | E_i]]
\]  

(14)

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 \( \text{CATTE}_{e,l'} \):

\[
E[Y_{i,t} - Y_{i,t}^\circ | E_i] = \sum_{l'=-T}^{T} \sum_{e=0}^{T} D_{i,t}^{l'} \cdot 1\{E_i = e\} \text{CATTE}_{e,l'}
\]

Plug the above expression in (14), the coefficient \( \mu_l \) is equal to

\[
\mu_l = e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'} \cdot 1\{E_i = e\}] \text{CATTE}_{e,l'}
\]

\[
= \sum_{l'\neq 0} e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'} \cdot 1\{E_i = e\}] \text{CATTE}_{e,l'}
\]

\[
= \omega_{e,l'}^{(l)}
\]

The superscript \( l \) in \( \omega_{e,l'}^{(l)} \) indexes \( \mu_l \). The subscript \( l' \) in \( \omega_{e,l'}^{(l)} \), indexes \( \text{CATTE}_{e,l'} \). We next show three properties of these weights in the expression of \( \mu_l \), the population regression coefficient on the indicator for relative period \( l \): 1) the weights on \( \text{CATTE}_{e,l} \) sum up to one i.e. \( \sum_e \omega_{e,l}^{(l)} = 1 \); 2) for each relative time \( l' \neq l \) included among the indicators, the weights on \( \text{CATTE}_{e,l'} \) sum up to zero i.e. for \( l' \geq -S, \neq -1, \leq F \), we have \( \sum_e \omega_{e,l'}^{(l)} = 0 \); 3) for relative times \( l' \) excluded among the indicators, the weights on \( \text{CATTE}_{e,l'} \) sum up to minus one across all excluded relative time \( l' \) and cohorts \( e \) i.e. for \( l' < -S, = -1, > F \), we have \( \sum_{l' < -S, = -1, > F} \sum_e \omega_{e,l'}^{(l)} = -1 \).

To see that 1) \( \sum_{e=0}^{T} \omega_{e,l}^{(l)} = 1 \), note that the sum of weights is equal to

\[
\sum_{e=0}^{T} \omega_{e,l}^{(l)} = \sum_{e=0}^{T} e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'} \cdot 1\{E_i = e\}]
\]

\[
= e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'} \cdot \sum_{e=0}^{T} 1\{E_i = e\}]
\]

\[
= e_i \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} \tilde{D}_{i,t}^\top] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'}]
\]

It is thus the population regression coefficient on \( D_{i,t}^l \) from regressing \( D_{i,t}^l \) on \( \tilde{D}_{i,t} \) and the unit and
time fixed effects, which is just one. Similarly, for each \( l' \neq l \) but included among the indicators, the sum of weights for \( CATT_{e,l'} \) is population regression coefficient on \( D_{i,t}^{l'} \) from regressing \( D_{i,t}^{l} \) on \( D_{i,t} \) and the unit and time fixed effects, which is zero. To see that the weights on \( CATT_{e,l'} \) from all excluded relative time \( l' \) and cohorts \( e \) add up to \(-1\), note that the sum of these weights is equal to

\[
\sum_{l' < \bar{s}, \bar{e} > \bar{f}} \sum_{e} \omega_{e,l'}^{(l)} = e_i^T \sum_{l' < \bar{k}, \bar{e} > \bar{L}} \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'}] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'}] \]

\[
= e_i^T \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l}] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} \sum_{l' < \bar{k}, \bar{e} > \bar{L}} D_{i,t}^{l'}] \]

\[
= e_i^T \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l}] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} (1 - \sum_{K \leq l' \leq L, l' \neq 1} D_{i,t}^{l'})] \]

\[
= - e_i^T \sum_{K \leq l' \leq L, l' \neq 1} \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l}] \right)^{-1} \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'}] = -1
\]

where the third equality follows from \( \sum_{-T \leq t \leq T} D_{i,t}^{l'} = 1 \). The last line simplifies to \(-1\) because a) \( \sum_t E[\tilde{D}_{i,t}] = 0 \) due to demeaning; b) each term in the summation

\[
\left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l}] \right)^{-1} \left( \sum_{t=0}^{T} E[\tilde{D}_{i,t} D_{i,t}^{l'}] \right)
\]

equals the population regression coefficient on \( D_{i,t} \) from regressing \( D_{i,t}^{l'} \) on \( D_{i,t} \) and the unit and time fixed effects, which is a column vector with one in the entry corresponding to \( D_{i,t}^{l'} \). Summing over \( l' \), the above expression is equal to a column vector of ones.

Finally, we note that the weight \( \omega_{e,l'}^{(l)} \) is equal to the population regression coefficient on \( D_{i,t}^{l} \) from regressing \( D_{i,t}^{l'} \cdot 1 \{E_i = e\} \) on relative time indicators and two-way fixed effects. There are certain combinations of \( e \) and \( l' \) that makes \( \omega_{e,l'}^{(l)} = 0 \): if \( e + l' < 0 \) or \( > T \), then \( D_{i,t}^{l'} \cdot 1 \{E_i = e\} \) is always zero. Regression coefficient on \( D_{i,t}^{l} \) is then zero. For other combinations of \( e \) and \( l' \), the weight \( \omega_{e,l'}^{(l)} \) is usually non-zero.
Proof of Proposition 5

Proof. Provided that the DID estimator is well-defined, the DID estimator is an unbiased and consistent estimator for \( E[Y_{i,e+l} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+l} - Y_{i,s} \mid E_i \in C] \).\(^2\) To show that it is an unbiased and consistent estimator for \( CATT_{e,l} \), we show \( E[Y_{i,e+l} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+l} - Y_{i,s} \mid E_i \in C] = CATT_{e,l} \).

Since \( s < e \) and \( c > e + l \), we have

\[
E[Y_{i,e+l} - Y_{i,s} \mid E_i = e] - E[Y_{i,e+l} - Y_{i,s} \mid E_i \in C] = E[Y_{i,e+l}^e - Y_{i,s}^e \mid E_i = e] - E[Y_{i,e+l}^c - Y_{i,s}^c \mid E_i = c] - \sum_{c \in C} Pr \{ E_i = c \} E[Y_{i,e+l}^c - Y_{i,s}^c \mid E_i = c] + E[Y_{i,e+l}^c - Y_{i,s}^c \mid E_i = c] - E[Y_{i,e+l}^e - Y_{i,s}^e \mid E_i = e] - E[Y_{i,e+l}^e - Y_{i,s}^e \mid E_i = e]
\]

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

Definition 2. The IW estimator for a weighted average of \( CATT_{e,l} \) is constructed via the following three steps:

\(^2\)Take the first term for an example. For unbiasedness, note that by the Law of Iterated Expectations

\[
E[\frac{\mathbb{E}_N(\sum_{i=1}^n (Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i = e\})}{\mathbb{E}_N(1\{E_i = e\})}] = E\left[ E[\frac{\mathbb{E}_N(\sum_{i=1}^n (Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i = e\})}{\mathbb{E}_N(1\{E_i = e\})} \mid E_i = e] \right]
\]

\[
= E[\frac{E[\sum_{i=1}^n (Y_{i,e+l} - Y_{i,s}) \cdot 1\{E_i = e\}]}{\mathbb{E}_N(1\{E_i = e\})}] = E[(Y_{i,e+l} - Y_{i,s}) \mid E_i = e] \cdot \mathbb{E}_N(1\{E_i = e\})
\]

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, it converges in probability to \( E[Y_{i,e+l} - Y_{i,s} \mid E_i = e] \).
Step 1 Estimate the $\text{CATT}_{e,t}$ using the interacted specification.

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

$$= \alpha_i + \lambda_t + B_{i,t}^T \delta + \epsilon_{it}$$

(16)

on observations from $t = 0, \ldots, T - 1$ 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 $D_{it}^{-1}$.

Here $B_{i,t}$ is a column vector collecting $1\{E_i = e\} \cdot D_{it}^l$ for each cohort $1 \leq e \leq T - 1$ with relative time spanning $-e \leq l \leq T - 1 - e, l \neq -1$. Similarly, $\delta$ is a column vector collecting the coefficients $\delta_{e,l}$ on $1\{E_i = e\} \cdot D_{it}^l$. The matrix notation is used later to derive the asymptotic variance of IW estimators.

Step 2 Estimate the weights for the dynamic estimator.

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

Define $\hat{\mathbf{f}}$ 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 $\hat{\mathbf{f}}^*$ the probability limit of $\hat{\mathbf{f}}$, 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 $l = 0$,

$$\hat{\mathbf{f}}^0 = \begin{pmatrix} N_1 / N_1 + N_2 & 0 \\ 0 & N_2 / N_1 + N_2 \end{pmatrix}$$

and its probability limit is

$$\mathbf{f}^0 = \begin{pmatrix} Pr\{E_i = 1 \mid 1 \leq E_i \leq 2\} & 0 \\ 0 & Pr\{E_i = 2 \mid 1 \leq E_i \leq 2\} \end{pmatrix}$$

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

---

3We need to drop time period $T$ because $\text{CATT}_{e,T}$ are not identified. We need to exclude cohort 0 from the sample because $\text{CATT}_{0,l}$ are not identified.
**Step 3** Compute the IW estimator as the weighted sum of $\delta_{e,l}$ (estimated in Step 1) using weights (estimated in Step 2).

To form an estimator alternative to the dynamic FE estimator $\hat{\mu}_l$ from Regression (1), we can use

$$\tilde{v}_l := \sum_{e=1}^{T-1-l} \frac{N_e}{\sum_{e=1-l}^{T-1-l} N_e} \hat{\delta}_{e,l} = vec(\tilde{f})^\top \tilde{\delta}.$$ 

Note that $\hat{\delta}_{e,l}$ from Regression (9) is a DID estimator for $CATT_{e,l}$ in the form of (8) with $s = e - 1$ and $C = \{T\}$, which is consistent.\(^4\) Thus, as we state more formally below, $\tilde{v}_l$ is consistent for a weighted average of $CATT_{e,l}$ 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 $CATT_{e,l}$. One can also form other weighted averages of $\hat{\delta}_{e,l}$ 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 4) on Regression (9), we can show that the IW estimators are asymptotically normal. We use the notation $\tilde{X}_{i,t}$ to denote double-demeaning $X_{i,t} - \overline{X}_{e,t} + \overline{X}$, where $\overline{X}_e = \sum_{t=0}^{T-1} X_{i,t} / T$, $\overline{X}_t = E[X_{i,t}]$ and $\overline{X} = \sum_{t=0}^{T-1} E[X_{i,t}] / T$.

**Assumption 4.** (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.

---

\(^4\)Among all possible DID estimators for $CATT_{e,l}$, we propose Regression (9) because it is a natural extension to the dynamic FE specification (1) by saturating in relative time indicators and cohort indicators.
4. Denote by $\mathbf{B}$ the data matrix, whose rows consist of $\mathbf{B}_{i,t}^\top$, double-demeaned version of $\mathbf{B}_{i,t}^\top$. Assume $\mathbf{B}$ has full rank. If $\mathbf{B}$ is reduced-rank because cohort $e$ is empty, then discard regressors involving $\mathbf{1}\{E_i = e\}$.

Denote by $\delta$ the probability limit of $\widehat{\delta}$, which is a vector of $\text{CATT}_{e,l}$. 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 6.** (Consistency and asymptotic normality of the IW estimators for dynamic treatment effects). Under Assumptions 1, 2 and 4, the IW estimator converges in probability to

$$\widehat{\nu}_l \to_p \sum_{e=1}^{T-1} \Pr\{E_i = e \mid 1 - l \leq E_i \leq T - 1 - l\} \text{CATT}_{e,l} = \text{vec}(f^\top) \delta.$$  

The asymptotic distribution of this estimator is

$$\sqrt{N}(\widehat{\nu}_l - \text{vec}(f^\top) \delta) \to_d N\left(0, \delta^\top \Sigma_f \delta + \Sigma_l\right)$$

for $\Sigma_f$ the asymptotic variance of $\sqrt{N}(\text{vec}(\widehat{f}) - \text{vec}(f)) \to_d$ where $\widehat{f}$ is the weight matrix estimator and

$$V_B = \sum_{t=0}^{T-1} E[\mathbf{B}_{i,t} \mathbf{B}_{i,t}^\top] \quad \Sigma_l = \text{vec}(f^\top) \text{Var} \left(\sum_{t=0}^{T-1} \mathbf{B}_{i,t} \mathbf{e}_{i,t}^2 \mathbf{B}_{i,t}^\top\right) V_B^{-1} \text{vec}(f).$$

**Proof of Proposition 6**

*Proof.* We first show the asymptotic normality of the weights. Recall that $\widehat{f}$ is a matrix with its $(t,e)^{th}$ entry equal to $\mathbf{1}\{t - e = l\} \cdot N_e / \sum_{e=1}^{T-1} N_e$. Here $\mathbf{1}\{t - e = l\}$ indicates when cohort $e$ experiences exactly $l$ periods of treatment and $N_e / \sum_{e=1}^{T-1} N_e$ is equal to the sample share of units in cohort $e$ among units that experience at least $l$ periods of treatment. Denote by $\mathbf{f}$ the probability limit of $\widehat{f}$, which is a matrix with its $(t,e)^{th}$ entry equal to $\mathbf{1}\{t - e = l\} \cdot Pr(E_i = t - l \mid 1 - l \leq E_i \leq T - 1 - l)$ since $\sum_{e=1}^{T-1} N_e / \sum_{e=1}^{T-1} N_e \to_p \sum_{e=1}^{T-1} N_e / \sum_{e=1}^{T-1} N_e = \Pr\{E_i = e \mid 1 - l \leq E_i \leq T - 1 - l\}$ by the Law of Large Numbers and Slutsky’s theorem. Note that $\sum_{e=1}^{T-1} N_e / \sum_{e=1}^{T-1} N_e$ is also the regression coefficient estimator from the following cross-sectional regression

$$\mathbf{1}\{E_i = e\} = \mathbf{\beta} \mathbf{1}\{1 - l \leq E_i \leq T - 1 - l\} + \eta(e).$$
with population regression coefficient equal to $\beta = Pr \{ E_i = e \mid 1 - l \leq E_i \leq T - 1 - l \}$. Then by OLS asymptotics which holds as $E_i$ are iid by assumption and $\eta(e_i)$ is bounded, we have

$$\sqrt{N} \left( \frac{N_e}{\sum_{i=l}^{T-1-l} N_e} - Pr \{ E_i = e \mid 1 - l \leq E_i \leq T - 1 - l \} \right) \rightarrow_d N \left( 0, \frac{E[1 \mid 1 - l \leq E_i \leq T - 1 - l]^2 \eta(e_i)^2}{E[1 \mid 1 - l \leq E_i \leq T - 1 - l]^2} \right).$$

Note that $1 \{ 1 - l \leq E_i \leq T - 1 - l \}^2 = 1 \{ 1 - l \leq E_i \leq T - 1 - l \}$ so the asymptotic variance is equal to

$$E[\eta(e_i)^2 \mid 1 - l \leq E_i \leq T - 1 - l] Pr \{ 1 - l \leq E_i \leq T - 1 - l \}$$

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

Similarly, for a pair of cohorts with $e \neq e'$, $\frac{N_e}{\sum_{i=l}^{T-1-l} N_e}$ and $\frac{N_e}{\sum_{i=l}^{T-1-l} N_e}$ are asymptotically correlated with covariance $E[\eta(e_i) \eta(e'_i) \mid 1 - l \leq E_i \leq T - 1 - l] / Pr \{ 1 - l \leq E_i \leq T - 1 - l \}$. Thus, $vec(\mathbf{f})$ has asymptotic distribution $\sqrt{N} \left( vec(\mathbf{f}') - vec(\mathbf{f}) \right) \rightarrow_d N \left( 0, \Sigma_{f'f} \right)$. $\Sigma_{f'f}$ is a matrix with diagonal entries equal to

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

and off-diagonal entries equal to

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

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

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

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

Lastly, by the delta method, we have

$$\sqrt{N} \left( vec(\mathbf{f})^\top \tilde{d} - vec(\mathbf{f})^\top \delta \right) \rightarrow_d N \left( 0, \delta^\top \Sigma_{f'f} \delta + \Sigma_t \right),$$

where $\Sigma_t = vec(\mathbf{f}')^\top V^{-1}_B Var \left( \sum_{t=0}^{T-1} \mathbf{B}_{i,t} \mathbf{e}_{i,t}^2 \mathbf{B}_{i,t}^T \right) V^{-1}_B vec(\mathbf{f})$. This follows because $vec(\mathbf{f})$ and $\tilde{d}$ are uncorrelated: the asymptotic covariance between $\frac{N_e}{\sum_{i=l}^{T-1-l} N_e}$ and $\tilde{d}$ is equal to

$$E[1 \{ 1 - l \leq E_i \leq T - 1 - l \} \eta(e_i), \mathbf{B}_{i,t} \mathbf{e}_{i,t}] = \frac{E[1 \{ 1 - l \leq E_i \leq T - 1 - l \}^2]}{E[1 \{ 1 - l \leq E_i \leq T - 1 - l \}^2]}. $$

Since $\mathbf{B}$ and $\eta(e_i)$ are functions of $E_i$, we have $E[\mathbf{B} \eta(e_i), \mathbf{e}_{i,t} \mid 1 - l \leq E_i \leq T - 1 - l] = E[\mathbf{B} \eta(e_i), E[\mathbf{e}_{i,t} \mid 1 - l \leq E_i \leq T - 1 - l]]$. 

32
\( E_i, 1 - l \leq E_i \leq T - 1 - l \). Furthermore, specification (9) is saturated in \( E_i \) and relative time so \( E[\hat{\epsilon}_{i,t} \mid E_i] = 0 \). This proves Proposition 6.

\[\square\]

### B Simulation design

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

\[
Y_{i,t} = i + t + \delta_{\varepsilon_{e,l}} \cdot \mathbb{1}\{E_i = e\} \cdot D_{i,t}^l + \epsilon_{i,t}
\]

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 \( \epsilon_{i,t} \sim N(0, 1) \). First, we set \( \epsilon_{e,l<0} = 0 \) for the case of no anticipatory behavior. For each simulation, we estimate the following FE specification

\[
Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-2,0,1,2} \mu_l D_{i,t}^l + \epsilon_{i,t}
\]

for \( t = 0, \ldots, 3 \) as well as the saturated specification

\[
Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \in \{1,2\}} \sum_{l \neq -1} \delta_{\varepsilon_{e,l}} \cdot \mathbb{1}\{E_i = e\} \cdot D_{i,t}^l + \epsilon_{i,t}
\]

for \( t = 0, 1, 2 \).

While \( \delta_{a,l<0} = 0 \) and hence \( CATT_{e,l<0} = 0 \) for all \( e \), the FE estimates of \( \hat{\mu}_{l=2} \) from specification (18) are usually negative as shown in Figure 3, which would suggest a negative pre-trend when none exists. In contrast, the IW estimates of \( \hat{\nu}_{l=2} \) based on \( CATT \) estimates from specification (19) are indistinguishable from zero, correctly suggesting a lack of pre-trend.
**Figure 3: Simulated Distribution of Lead Estimates with No Anticipatory Behavior**

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

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 (18). The estimates of $\hat{\mu}_{-2}$ are now indistinguishable from zero as shown in Figure 4, which would suggest a lack of pre-trend when one does exist. In contrast, the IW estimates of $\hat{\nu}_{-2}$ based on $CATT$ estimates from specification (19) are usually negative around -3, correctly suggesting a pre-trend.

**Figure 4: Simulated Distribution of Lead Estimates with Anticipatory Behavior**

Notes: This figure plots the histogram of $\hat{\mu}_{-2}$ and $\hat{\nu}_{-2}$, the FE and IW estimates 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.