Estimating Event Studies When Units Experience Multiple Events

David A. Rosenkranz*

May 2, 2022

Abstract. An event study is an empirical framework for measuring the impact of an event over time using observational data. Under no anticipation and parallel trends assumptions, difference-in-differences are known to identify the event's average treatment effect on the treated when units experience one event at most. In this paper, I introduce a new event study framework to accommodate settings where units may experience multiple events. I introduce a matching estimator which consistently and transparently estimates the average treatment effect on the treated of a single event under generalizations of the conventional no anticipation and parallel trends assumptions. I show that the matching estimator is equivalent to a weighted least squares estimator for a particular set of weights. I also introduce a parallel pre-trends test which can be used to scrutinize these assumptions in the usual sense. Finally, I demonstrate in a series of Monte Carlo simulations that the estimator and parallel pre-trends test work well for a wide range of treatment effects, including dynamic, non-stationary, and history-dependent treatment effects.

*Department of Economics, Columbia University. Email: darosenkranz@gmail.com. I thank Douglas Almond, Michael Best, Brantly Callaway, Sebastian Calonico, Simon Lee, José Luis Montiel Olea, Pedro Sant'Anna, and Ashley Swanson for helpful conversations. I am grateful to attendees of the Columbia University Department of Economics Applied Microeconomics Colloquium and Industrial Organization Colloquium. All remaining errors are mine.

An event study is an empirical framework for measuring the impact of an event over time using observational data. Dating back to at least Snow (1855), event studies are routinely used to measure the impact of news, policy changes, natural disasters, and many other events (e.g., Greenstone, Oyer, and Vissing-Jorgensen, 2006; Cengiz, Dube, Lindner, and Zipperer, 2019; and Dave et al. (2020), to name very few).

Researchers typically conduct event studies as follows. First, they articulate a theoretical model causally relating an event to an outcome of interest. In the model, no anticipation and parallel trends assumptions equilibrate difference-in-differences between treated and untreated units within some event window with the average treatment effect on the treated ("ATT," hereafter). The no anticipation assumption states that average pre-event outcomes among treated units would have been unchanged had they been untreated. The parallel trends assumption states that average post-event outcomes among treated and untreated units would have moved synchronously had the treated units been untreated. Researchers scrutinize the plausibility of the no anticipation and parallel trends assumptions by examining whether average pre-event outcomes among treated and untreated units had been moving synchronously. Second, they estimate regressions like:

$$Y_{it} = \beta D_{it} + \mathrm{FE}_i + \mathrm{FE}_t + \varepsilon_{it} \tag{1}$$

$$Y_{it} = \sum_{\tau \in \mathcal{W}} \beta_\tau D_{it\tau} + \mathrm{FE}_i + \mathrm{FE}_t + \varepsilon_{it} \tag{2}$$

where i indexes units, t indexes periods, $D_{it} := \mathbb{1}[i \text{ was treated before } t]$, $\mathcal{W}$ is an event window, and $D_{it\tau} := \mathbb{1}[i \text{ was treated in } t - \tau]$. Under no anticipation, parallel trends, and some other conditions, the OLS estimators of (1) and (2) consistently estimate ATTs of the event on Y when units experience one event at most.

Unfortunately, this approach is not applicable to settings where units experience multiple events. Intuitively, if some units experience multiple events, then "pre-event" and "post-event" are ill-defined—a period following one event may also precede another. This has several practical consequences for empirical research. First, some researchers attempt to avoid "contamination" from multiple events by excluding all (unit, event) pairs in which the unit previously experienced another event or experienced another event within the event window (e.g., Sarsons, 2017; Einav, Finkelstein, and Mahoney, 2019; Rees-Jones and Rozema, 2020).[1] Consequently, they may lose

[1]Sarsons measured how physician referrals to surgeons differentially respond to unexpected outcomes during surgery based on the surgeon's gender. When multiple unexpected surgical events occur within the same physician-surgeon pair, Sarsons "only [kept] the first instance of an event." Einav, Finkelstein, and Mahoney measured how long-term care hospitals affect health care spending and patient welfare using variation generated by LTCH entry. Since geographic markets may experience multiple LTCH entries, they "focus[ed] on the entry of the first LTCH" in their main specification because "this is where [they] expect[ed] to see the sharpest effects." Wherever multiple

variation necessary to estimate small but economically significant treatment effects, limit their estimates' external validity in settings where included events differ from excluded events, and create an unnecessary link between their sample composition and the event window. Second, some researchers include all events but make only the first event a focal event (e.g., Kleven, Landais, and Søgaard, 2019).[2] Consequently, their estimates measure the effect of starting to experience events rather than the effect of an event itself.

In this paper, I introduce an event study framework for settings where some units may experience multiple events. My framework allows researchers to incorporate variation generated by multiple events into their estimates of the ATT of a single event. It accommodates settings where units experience one event at most as a special case. First, I introduce a theoretical framework which characterizes a unit's potential outcomes across a sequence of periods as functions of the unit's treatment history. I show that the ATT of a single event is identified by generalizations of the no anticipation and parallel trends assumptions commonly used in settings where units experience one event at most. Second, I introduce a matching estimator. The estimator transparently estimates the ATT of an event by calculating the difference-in-differences between units that did and did not experience that event, but otherwise experienced similar event histories. Under the generalized no anticipation and parallel trends assumptions, the estimator is consistent whether or not ATTs vary with relative time (i.e., whether they are static or dynamic), the event period (i.e., whether they are stationary or non-stationary), or the sequence of past events (i.e., whether they are history independent or history dependent). I apply a result by Gardner (2021) to show that the matching estimator is equivalent to a weighted least squares estimator for particularly chosen weights. Third, I introduce a parallel pre-trends test which may be used to scrutinize the generalized no anticipation and parallel trends assumptions in the usual sense. Fourth, I evaluate my proposed approach in a series of Monte Carlo simulations. It performs well.

This study is most closely related to Sandler and Sandler (2014), Schmidheiny and Siegloch (2019), who each separately proposed a so-called "Multiple Dummies On" ("MDO," hereafter) estimator for settings where units experience multiple events. The MDO estimator is the OLS

LTCHs entered, the authors "truncate[d] the data just before the quarter of second LTCH entry [...] so that the post-entry results are not contaminated by further shocks[.]" Rees-Jones and Rozema measured tobacco consumption responses to "media coverage, lobbying efforts, place-based smoking restrictions, and anti-smoking appropriations" that coincide with state-level changes to cigarette taxes. Since states may experience multiple changes to cigarette taxes, they "define[d] some window of time around a tax-change event" and "consider[ed] a single event in isolation" by "examining a variable of interest across that window, comparing the 'treated' state with a tax change to 'control' states experiencing no tax change in that window."

[2]Kleven, Landais, and Søgaard measured the "impact of children on the labor market trajectories of women relative to men" around the time of childbirth. Since women may have multiple children, they measured the effect of children on female labor market trajectories "around the birth of the first child" and noted that their results should be interpreted "as a total penalty including the costs of children born after the first one."

estimator of (2), where multiple elements of $(D_{it\tau} : \tau \in \mathcal{W})$ may be equal to one for a single (i, t) observation and where the outer two elements of $(D_{it\tau} : \tau \in \mathcal{W})$ are equal to the number of events a unit experienced outside $\mathcal{W}$. In a series of Monte Carlo simulations, Sandler and Sandler showed that the MDO estimator performs well when treatment effects are static, stationary, and history independent, and when treatment effects are dynamic, stationary, and history independent, provided that $\mathcal{W}$ is wide enough to accommodate all dynamics. I contribute to this literature a matching estimator which also performs well in settings where treatment effects are non-stationary, history dependent, or dynamic beyond the range of an event window. I also contribute a theoretical framework that researchers can use to clarify their identifying assumptions and interpret their estimates. Finally, I contribute a parallel pre-trends test with more power to detect deviations from the identifying assumptions.

In recent parallel work, De Chaisemartin and D'Haultfœuille (undated) also consider how to estimate event studies in settings where units experience multiple events. They show that OLS estimators of linear regression models with multiple treatment dummies may be biased and they propose an alternative matching estimator. While we both propose estimators that compare units that experience similar treatment histories, our work differs in some respects. First, we propose different estimators. They estimate the ATT in some period t of an event that occurred in some period $e < t$ by comparing trends between units that shared the same history of events between periods 1 and $e - 1$, excluding those units that experienced another event between periods $e + 1$ and t. On the other hand, I compare trends between all units that share the same history of events. That I include all data generated between periods $e + 1$ and t within matched groups but they do not implies that my estimator uses more data to estimate long-term treatment effects within matched groups because it does not exclude data as t grows relative to e. It also implies that when some units experience another event between any two periods $t, t' > e$, their estimator may be unable to distinguish treatment effect dynamics between t and t' from differences in the samples used to estimate the t and t' ATTs. However, that their estimator compares units that share the same event history up to period e whereas my estimator compares units that share the same event history implies that their estimator will use more matched groups. Second, I also propose a regression-based weighted least squares estimator that is equivalent to my matching estimator and I describe a parallel pre-trends test. Third, I demonstrate that my estimator outperforms the MDO estimator in a series of Monte Carlo simulations. On the other hand, De Chaisemartin and D'Haultfœuille demonstrate arithmetically that OLS estimators of a large class of models may sometimes be biased.

This study is also related to the recent microeconometrics literature concerned with estimating event studies in settings where units experience one event at most (e.g., Borusyak, Jaravel, and Spiess 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; de Chaisemartin and

D'Haultfœuille, 2020; Goodman-Bacon, 2021; and Imai, Kim, and Wang, 2021). This literature has demonstrated that the OLS estimators of equations (1) and (2) do not consistently estimate ATTs when the equations misspecify the treatment effect and units experience their event in a staggered fashion. For instance, Goodman-Bacon (2021) showed that $\hat{\beta}$ is a weighted average of all possible two-unit, two-period differences-in-differences—including differences-in-differences between later-treated units and earlier-treated units before-and-after the later-treated units' event. Under no anticipation and parallel trends, such differences-in-differences extract information about the ATT when the ATT is static because the earlier-treated units' realized outcome trends are equal to the later-treated units' counterfactual trends. But if the ATT is dynamic, then the earlier-treated units' realized outcome trends include the dynamic treatment effect. Consequently, the differences-in-differences between later-treated units and earlier-treated units around the later-treated units' event bias $\hat{\beta}$.[3] Similarly problematic comparisons bias $(\hat{\beta} : \tau \in \mathcal{W})$ in settings with non-stationary treatment effects (Sun and Abraham, 2021). For instance, elements of $(\hat{\beta}_\tau : \tau < 0)$ may be non-zero under parallel trends and no anticipation. I contribute to this literature similar observations about the MDO estimator. I propose robust alternatives for settings where units may experience multiple events. My estimators transparently compute and aggregate raw difference-in-differences, enabling researchers to clearly articulate the link between their identification strategy and estimates. In settings where units experience one event at most, my proposed estimators are equivalent to the interaction weighted estimator proposed by Sun and Abraham (2021).

The remainder of this paper proceeds as follows. In section I, I introduce the theoretical framework. I define potential outcomes, express the generalized no anticipation and parallel trends assumptions, show that they are sufficient to identify ATTs, and propose formulas to aggregate the ATTs into easily interpretable causal parameters. In section II, I introduce the matching and weighted least squares estimators. In section III, I introduce the parallel pre-trends test. In section IV, I demonstrate that my estimators and parallel pre-trends test perform well in a series of Monte Carlo simulations. In section V, I discuss practical issues and extensions of the baseline approach. I conclude in section VI.

[3] To see this another way, consider that $\hat{\beta} = \sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it}\omega_{it}$, where $\omega_{it} := \frac{\tilde{D}_{it}}{\sum_{i'=1}^{N} \sum_{t'=1}^{T} \tilde{D}_{i't'}^2}$ and $\tilde{D}_{it}$ is the residual from the regression $D_{it} = \mathrm{FE}_i^D + \mathrm{FE}_t^D + \nu_{it}$. In a setting with some earlier-treated units and many later-treated units, $\mathrm{FE}_i^D + \mathrm{FE}_t^D$ may exceed 1 for earlier-treated units in later periods. Therefore, ω_{it} may be negative for earlier-treated units in later periods. That $\omega_{it} < 0$ for earlier-treated units in later periods means that $\hat{\beta}$ incorporates the corresponding Y_{it} more often in the comparison group parts of $\hat{\beta}$'s constituent differences-in-differences than in the treated group parts—even though unit i was treated before period t.

I. Model

A. Setup

Consider a setting with a population of units $\mathcal{I}$ observed over a sequence of at least two periods $\mathcal{T} := \{1, ..., T\}$. Each unit may experience an event in each period.[4] For each $(i, e) \in \mathcal{I} \times \mathcal{T}$, let H_{ie} indicate whether unit i experiences an event in period e and let $H_i := (H_{ie} : e \in \mathcal{T})$ be the unit's complete event history.[5] Let $\Omega_H \subseteq \{0, 1\}^T$ be the support of H_i. For each $(i, e) \in \mathcal{I} \times \mathcal{T}$, define $H_i^{-e} := (H_{i1}, ..., H_{ie-1}, H_{ie+1}, ..., H_{iT})$ as unit i's history of events in periods other than period e and define $\Omega_{H^{-e}}$ as the support of H_i^{-e}. I will refer to elements of $\Omega_{H^{-e}}$ as e-histories to distinguish them from the complete histories in Ω_H. For clarity, I will sometimes write H_i^{-e} as $(H_{i1}, ..., H_{ie-1}, -, H_{ie+1}, ..., H_{iT})$ and H_i as (H_i^{-e}, H_{ie}) for each $(i, e) \in \mathcal{I} \times \mathcal{T}$.

In this setting, the events may affect the units. For each $(i, t, H) \in \mathcal{I} \times \mathcal{T} \times \Omega_H$, let $Y_{it}(H)$ be unit i's potential outcome in period t were unit i to experience history H, and let $Y_{it} := Y_{it}(H_i)$ be unit i's realized outcome. For each $e \in \mathcal{T}$, let $\Omega^*_{H^{-e}} := \{H^{-e} : (H^{-e}, 1) \in \Omega_H \wedge (H^{-e}, 0) \in \Omega_H\}$ be the set of e-histories which admit both treated and untreated units in period e.[6] For each $(i, e, t, H^{-e}) \in \mathcal{I} \times \mathcal{T}^2 \times \Omega^*_{H^{-e}}$, define $\beta_{iet}(H^{-e}) := Y_{it}(H^{-e}, 1) - Y_{it}(H^{-e}, 0)$. This is the treatment effect for unit i in period t of a period e event given the e-history H^{-e}.[7] For each $(e, t, H^{-e}) \in \mathcal{T}^2 \times \Omega^*_{H^{-e}}$, define the ATT:

$$\beta_{et}^{ATT}(H^{-e}) := \mathbb{E}\left[\beta_{iet}(H^{-e}) \,\middle|\, H_i = (H^{-e}, 1)\right]$$

This is the average treatment effect in period t of a period e event given the e-history H^{-e} among units that actually experience $H_i = (H^{-e}, 1)$.

[4] I consider settings where units may experience multiple events per period in section V. Each event is permanent in the sense that a unit that is treated in a period e is forever "treated in period e." However, the effect of each event may fade over time.

[5] For example, consider the case that $T = 3$. If $H_i = (0, 1, 1)$ and $H_j = (0, 1, 0)$, then unit i experiences an event in periods 2 and 3, while unit j experiences an event only in period 2.

[6] For each period $e \in \mathcal{T}$, the set of e-histories $\Omega^*_{H^{-e}}$ is a strict subset of $\Omega_{H^{-e}}$ if and only if there exists some e-history $H^{-e} \in \Omega_{H^{-e}}$ such that $(H^{-e}, 0) \notin \Omega_H$ or $(H^{-e}, 1) \notin \Omega_H$. For such e-histories H^{-e}, treatment effects are not defined.

[7] I discuss cumulative treatment effects in section V.

B. Identification

Ideally, researchers could estimate the ATTs by directly computing $Y_{it}(H^{-e}, 1) - Y_{it}(H^{-e}, 0)$ for each $(i, e, t, H^{-e}) \in \mathcal{I} \times \mathcal{T}^2 \times \Omega^*_{H^{-e}}$. However, the fundamental problem of causal inference in this setting is that $Y_{it}(H^{-e}, 1)$ and $Y_{it}(H^{-e}, 0)$ are never both observed for the same unit—researchers cannot compute $Y_{it}(H^{-e}, 1) - Y_{it}(H^{-e}, 0)$. Instead, researchers can use event studies to estimate the ATTs under generalizations of the no anticipation and parallel trends assumptions usually made in settings where units experience one event at most.

No anticipation. For any $(t, H) \in \mathcal{T} \times \Omega_H$, let $H^{\leq t} := (H_1, ..., H_t)$ and let $H^{>t} := (H_{t+1}, ..., H_T)$. (Let $H^{>T}$ be null.) For any $(t, H) \in \mathcal{T} \times \Omega_H$ and any $(\hat{H}^{>t}, \tilde{H}^{>t})$ such that $(H^{\leq t}, \hat{H}^{>t}) \in \Omega_H$ and $(H^{\leq t}, \tilde{H}^{>t}) \in \Omega_H$:

$$\mathbb{E}\left[Y_{it}(H^{\leq t}, \tilde{H}^{>t})|H_i = H\right] = \mathbb{E}\left[Y_{it}(H^{\leq t}, \hat{H}^{>t}_B)|H_i = H\right]$$

This states that within any group of units $\{i : H_i = H\}$ that share a realized event history, the average potential outcomes in any period t following the realized sequence of events $H_i^{\leq t}$ are invariant to any future sequence of events $\tilde{H}^{>t}$ or $\hat{H}^{>t}$. In other words, it states that for any cohort of units that share a realized event history, the realized sequence of events up to period t is a sufficient statistic for the corresponding average potential outcome in period t. Consider for example the case that $T = 3$ and $\Omega_H := \{0, 1\}^3$. Then for $\{i : H_i = (0, 1, 1)\}$, the no anticipation assumption states that:

$$\begin{aligned}
\mathbb{E}\left[Y_{i1}(0,1,1)|H_i = (0,1,1)\right] &= \mathbb{E}\left[Y_{i1}(0,0,1)|H_i = (0,1,1)\right] \\
&= \mathbb{E}\left[Y_{i1}(0,1,0)|H_i = (0,1,1)\right] \\
&= \mathbb{E}\left[Y_{i1}(0,0,0)|H_i = (0,1,1)\right] \\
\mathbb{E}\left[Y_{i2}(0,1,1)|H_i = (0,1,1)\right] &= \mathbb{E}\left[Y_{i2}(0,1,0)|H_i = (0,1,1)\right]
\end{aligned}$$

The generalized no anticipation assumption is similar to the conventional no anticipation assumption researchers express in settings where units experience one event at most. For example, consider Einav, Finkelstein, and Mahoney's (2019) study of the effects of LTCH entry on health care spending. Assume for simplicity that $T = 3$ and $\Omega_H := \{0\} \times \{0, 1\}^2$. In this setting, the generalized no anticipation assumption may be stated as follows:

- For one event: Average health care spending in hospital markets that experienced exactly one LTCH entry was unaffected by the LTCH entry until the LTCH entry occurred.

  - For example, for hospital markets that experienced LTCH entry history $H_i = (0, 1, 0)$, average health care spending in period 1 was the same as it would have been had the hospital markets experienced $H_i = (0, 0, 0)$, $H_i = (0, 0, 1)$, or $H_i = (0, 1, 1)$ instead.

- For multiple events: Average health care spending in hospital markets that experienced exactly two LTCH entries was unaffected by either LTCH entry until each LTCH entry occurred.

  - For example, for hospital markets that experienced LTCH entry history $H_i = (0, 1, 1)$, average health care spending in period 1 was the same as it would have been had the hospital markets experienced $H_i = (0, 0, 0)$, $H_i = (0, 0, 1)$, or $H_i = (0, 1, 0)$ instead. In addition, average health care spending in period 2 was the same as it would have been had the hospital markets experienced $H_i = (0, 1, 0)$ instead.

Parallel post-trends. For all $(e, t, H^{-e}) \in \mathcal{T}^2 \times \Omega^*_{H^{-e}}$ such that $t \geq e \geq 2$,

$$\mathbb{E}\big[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\big|\, H_i = (H^{-e}, 1)\big] = \mathbb{E}\big[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\big|\, H_i = (H^{-e}, 0)\big]$$

Let $\mathcal{E} := \mathcal{T} \setminus \{1\}$. The parallel post-trends assumption states that for any $(e, H^{-e}) \in \mathcal{E} \times \Omega^*_{H^{-e}}$, the average change in outcomes between periods $e-1$ and $t \in \{e, ...T\}$ among units $\{i : H_i = (H^{-e}, 1)\}$ were they to experience history $H_i = (H^{-e}, 0)$ instead is the same as the contemporaneous average change in realized outcomes among units $\{i : H_i = (H^{-e}, 0)\}$. Consider for example the case that $T = 3$ and $\Omega_H = \{0, 1\}^3$. Then for units $\{i : H_i = (0, 1, 1)\}$, the parallel post-trends assumption states that:

$$\begin{aligned}
\mathbb{E}\big[Y_{i2}(0,0,1) - Y_{i1}(0,0,1)|H_i = (0,1,1)\big] &= \mathbb{E}\big[Y_{i2}(0,0,1) - Y_{i1}(0,0,1)|H_i = (0,0,1)\big] \\
\mathbb{E}\big[Y_{i3}(0,0,1) - Y_{i1}(0,0,1)|H_i = (0,1,1)\big] &= \mathbb{E}\big[Y_{i3}(0,0,1) - Y_{i1}(0,0,1)|H_i = (0,0,1)\big] \\
\mathbb{E}\big[Y_{i3}(0,1,0) - Y_{i2}(0,1,0)|H_i = (0,1,1)\big] &= \mathbb{E}\big[Y_{i3}(0,1,0) - Y_{i2}(0,1,0)|H_i = (0,1,0)\big]
\end{aligned}$$

The generalized parallel post-trends assumption is similar to the conventional parallel post-trends assumption researchers express in settings where units experience one event at most. For example, consider Sarsons's (2017) study of the effects of negative surgical events on physicians' referral patterns. Assume for simplicity that $T = 3$ and $\Omega_H := \{0\} \times \{0, 1\}^2$. In this setting, the generalized parallel post-trends assumption may be stated as follows:

- For one event: Consider surgeons who experienced exactly one negative surgical event. After the negative surgical event and but-for the negative surgical event, average referrals among these surgeons would have moved synchronously with average referrals among surgeons who experienced no negative surgical events.

- For example, consider a group of surgeons who experienced the history of negative surgical events $H_i = (0, 0, 1)$. For this group of surgeons, but for the period three event, average referrals between periods 2 and 3 would have moved synchronously with those of surgeons who also experienced e-history $H_i^{-3} = (0, 0, -)$, but who experienced $H_{i3} = 0$ instead of $H_{i3} = 1$.

• For multiple events: Consider surgeons who experienced negative surgical events in both periods 2 and 3. After each respective negative surgical event and but-for each respective negative surgical event, average referrals among these surgeons would have moved synchronously with average referrals among surgeons who did not experience that surgical event, but did experience the other.

- For example, consider a group of surgeons who experienced the history of negative surgical events $H_i = (0, 1, 1)$. For this group of surgeons, but for the period 2 event, average referrals between periods 1, 2, and 3 would have moved synchronously with those of surgeons who also experienced e-history $H_i^{-2} = (0, -, 1)$, but who experienced $H_{i2} = 0$ rather than $H_{i2} = 1$. In addition, their average referrals between periods 2 and 3 would have moved synchronously with those of surgeons who also experienced e-history $H_i^{-3} = (0, 1, -)$, but who experienced $H_{i3} = 0$ rather than $H_{i3} = 1$.

Theorem 1. Under no anticipation and parallel post-trends, it follows that for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$ such that $t \geq e$, $\beta^{ATT}_{et}(H^{-e})$ is identified by the difference-in-differences:

$$\beta^{ATT}_{et}(H^{-e}) = \mathbb{E}[Y_{it} - Y_{ie-1} | H_i = (H^{-e}, 1)] - \mathbb{E}[Y_{it} - Y_{ie-1} | H_i = (H^{-e}, 0)] \tag{3}$$

For proof, see Appendix A. ■

To illustrate, consider the case that $T = 4$ and $\Omega_H = \{0\} \times \{0, 1\}^3$. The following ATTs are identified by difference-in-differences under no anticipation and parallel post-trends:

1. For the purposes of identifying the effect of being treated in period 2, compare:

   1.1. $\{i : H_i = (0, 1, 0, 0)\}$ with $\{i : H_i = (0, 0, 0, 0)\}$ to get $\left(\beta^{ATT}_{2t}(0, -, 0, 0) : t \in \mathcal{T}\right)$

   1.2. $\{i : H_i = (0, 1, 1, 0)\}$ with $\{i : H_i = (0, 0, 1, 0)\}$ to get $\left(\beta^{ATT}_{2t}(0, -, 1, 0) : t \in \mathcal{T}\right)$

   1.3. $\{i : H_i = (0, 1, 0, 1)\}$ with $\{i : H_i = (0, 0, 0, 1)\}$ to get $\left(\beta^{ATT}_{2t}(0, -, 0, 1) : t \in \mathcal{T}\right)$

1.4. $\{i : H_i = (0,1,1,1)\}$ with $\{i : H_i = (0,0,1,1)\}$ to get $\left(\beta_{2t}^{ATT}(0,-,1,1) : t \in \mathcal{T}\right)$

2. For the purposes of identifying the effect of being treated in period 3, compare:

   2.1. $\{i : H_i = (0,0,1,0)\}$ with $\{i : H_i = (0,0,0,0)\}$ to get $\left(\beta_{3t}^{ATT}(0,0,-,0) : t \in \mathcal{T}\right)$

   2.2. $\{i : H_i = (0,1,1,0)\}$ with $\{i : H_i = (0,1,0,0)\}$ to get $\left(\beta_{3t}^{ATT}(0,1,-,0) : t \in \mathcal{T}\right)$

   2.3. $\{i : H_i = (0,0,1,1)\}$ with $\{i : H_i = (0,0,0,1)\}$ to get $\left(\beta_{3t}^{ATT}(0,0,-,1) : t \in \mathcal{T}\right)$

   2.4. $\{i : H_i = (0,1,1,1)\}$ with $\{i : H_i = (0,1,0,1)\}$ to get $\left(\beta_{3t}^{ATT}(0,1,-,1) : t \in \mathcal{T}\right)$

3. For the purposes of identifying the effect of being treated in period 4, compare:

   3.1. $\{i : H_i = (0,0,0,1)\}$ with $\{i : H_i = (0,0,0,0)\}$ to get $\left(\beta_{4t}^{ATT}(0,0,0,-) : t \in \mathcal{T}\right)$

   3.2. $\{i : H_i = (0,1,0,1)\}$ with $\{i : H_i = (0,1,0,0)\}$ to get $\left(\beta_{4t}^{ATT}(0,1,0,-) : t \in \mathcal{T}\right)$

   3.3. $\{i : H_i = (0,0,1,1)\}$ with $\{i : H_i = (0,0,1,0)\}$ to get $\left(\beta_{4t}^{ATT}(0,0,1,-) : t \in \mathcal{T}\right)$

   3.4. $\{i : H_i = (0,1,1,1)\}$ with $\{i : H_i = (0,1,1,0)\}$ to get $\left(\beta_{4t}^{ATT}(0,1,1,-) : t \in \mathcal{T}\right)$

I discuss how researchers can aggregate these parameters in section I.C.

In settings where units experience one event at most, researchers conducting event studies commonly scrutinize the plausibility of the no anticipation and parallel trends assumptions by examining whether trends among treated and untreated units had been moving synchronously before the treated units' event. In section III, I propose an approach for conducting such an exercise in settings where units may experience multiple events. In particular, I introduce a statistical test of the null hypothesis that $\beta_{et}^{ATT}(H^{-e}) = 0$ for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega_{H^{-e}}^*$ such that $t < e$. My approach relies on the following:

Parallel pre-trends. For all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega_{H^{-e}}^*$ such that $t < e$:

$$\mathbb{E}\left[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0) \,\middle|\, H_i = (H^{-e},1)\right] = \mathbb{E}\left[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0) \,\middle|\, H_i = (H^{-e},0)\right]$$

Theorem 2. Under no anticipation and parallel pre-trends, it follows that for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega_{H^{-e}}^*$ such that $t < e$, $\beta_{et}^{ATT}(H^{-e})$ is identified by the difference-in-differences in equation (3), and furthermore $\beta_{et}^{ATT}(H^{-e}) = 0$. For proof, see Appendix A. ■

The no anticipation and parallel trends assumptions have several additional implications. First, they imply that groups of treated units that share an untreated comparison group also share a counterfactual outcome path. Continuing with the foregoing example where $T = 4$ and $\Omega_H = \{0\} \times \{0,1\}^3$, parallel trends implies that average outcomes among $\{i : H_i = (0,0,0,1)\}$, $\{i : H_i =$

$(0, 0, 1, 0)\}$, and $\{i : H_i = (0, 1, 0, 0)\}$ would have moved synchronously with one another had they been untreated. As in settings where units experience one event at most, this implies that if events are assigned to groups in part based on the groups' average untreated outcome trends, then the identifying assumptions are not satisfied.

Second, the no anticipation and parallel trends assumptions imply that groups of units that share a common initial sequence of events move synchronously until they experience their first differentiating event. For example, consider the groups $A := \{i : H_i = (0, 0, 0, 0)\}$, $B := \{i : H_i = (0, 0, 1, 0)\}$, and $D := \{i : H_i = (0, 0, 1, 1)\}$. The no anticipation and parallel trends assumptions imply that average realized outcomes among A, B, and D move synchronously between periods 1 and 2. Furthermore, they imply that average realized outcomes among B and D move synchronously between periods 1, 2, and 3.

Third, no anticipation and parallel trends imply that some ATTs are linear combinations of others.[8] Define the vectors $\theta := \left(\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}\right)$ and $E := \left(\mathbb{E}[Y_{it}|H_i = H] : (t, H) \in \mathcal{T} \times \Omega_H\right)$. No anticipation and parallel trends imply that each element $\beta_{et}^{ATT}(H^{-e})$ of θ is a linear combination of four elements of E, according to equation (3). The number of linearly independent parameters $\beta_{et}^{ATT}(H^{-e})$ is equal to the rank of the matrix A satisfying:

$$A \cdot E = \theta \tag{4}$$

I return to this observation in section III.

Remark. The source of the "contamination" problem that can motivate researchers to exclude some events is a correlation between H_{ie} and H_i^{-e}. If experiencing an event in period e is correlated with experiencing events in other periods e', then the average counterfactual trend among units $\{i : H_{ie} = 1\}$ differs from the average observed trend among units $\{i : H_{ie} = 0\}$. Mathematically,

$$\begin{aligned}
&\mathbb{E}\left[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\middle|\, H_{ie} = 1\right] \\
&\quad = \sum_{H^{-e}} \mathbb{E}\left[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\middle|\, H_i = (H^{-e}, 1)\right] \mathbb{P}(H_i^{-e} = H^{-e}|H_{ie} = 1) \\
&\quad = \sum_{H^{-e}} \mathbb{E}\left[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\middle|\, H_i = (H^{-e}, 0)\right] \underbrace{\mathbb{P}(H_i^{-e} = H^{-e}|H_{ie} = 1)}_{*} \\
&\quad \neq \sum_{H^{-e}} \mathbb{E}\left[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\middle|\, H_i = (H^{-e}, 0)\right] \underbrace{\mathbb{P}(H_i^{-e} = H^{-e}|H_{ie} = 0)}_{\text{Not equal to } *}
\end{aligned}$$

[8] For example, they imply that $\left[\beta_{34}^{ATT}(0, 0, -, 0) - \beta_{33}^{ATT}(0, 0, -, 0)\right] - \left[\beta_{34}^{ATT}(0, 0, -, 1) - \beta_{33}^{ATT}(0, 0, -, 1)\right] + \beta_{44}^{ATT}(0, 0, 1, -) = \beta_{44}^{ATT}(0, 0, 0, -)$.

$$= \mathbb{E}\big[Y_{it}(H^{-e}, 0) - Y_{ie-1}(H^{-e}, 0) \,\big|\, H_{ie} = 0\big]$$

For example, if e-histories H^{-e} that include many events are relatively more likely to be realized given $H_{ie} = 1$, and if the effect of an event is positive, then the difference-in-differences between units treated in period e and units untreated in period e before-and-after period e would overestimate the ATT of the period e event. The identification strategy I introduced in this section resolves this contamination problem by comparing units that have the same sequence of events in periods $e' \neq e$.

C. Aggregating parameters

Under no anticipation and parallel trends, equation (3) identifies a large set of parameters: one $\beta_{et}^{ATT}(H^{-e})$ for every $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$. In practice, researchers may wish to aggregate these parameters to summarize their findings. I propose the following aggregates.[9]

First, I propose aggregating $\left\{\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}\right\}$ across $H^{-e} \in \Omega^*_{H^{-e}}$, as follows. For each $(e, t) \in \mathcal{E} \times \mathcal{T}$, define:

$$\beta_{et}^{ATT} := \mathbb{E}\big[\beta_{et}^{ATT}(H^{-e}) \,\big|\, H_{ie} = 1\big] = \sum_{H^{-e} \in \Omega^*_{H^{-e}}} \beta_{et}^{ATT}(H^{-e}) \mathbb{P}\big(H_i^{-e} = H^{-e} \,\big|\, H_{ie} = 1\big) \tag{5}$$

This parameter is the average effect in period t of a period e event among units $\left\{i : H_{ie} = 1\right\}$. Each $\beta_{et}^{ATT}(H^{-e})$ is weighted by the proportion of units treated in period e that experience $H_i^{-e} = H^{-e}$. For each period $e \in \mathcal{E}$, the vector of parameters $\left(\beta_{et}^{ATT} : t \in \mathcal{T}\right)$ is the average evolution of the ATT among units $\left\{i : H_{ie} = 1\right\}$.

Second, the set of parameters $\{\beta_{et}^{ATT} : (e, t) \in \mathcal{E} \times \mathcal{T}\}$ may be aggregated to construct a measure of the average evolution of the ATT. Let the backward-looking event window be W_B and the forward-looking event window be W_F such that $W_B + W_F \leq T$. Define the set $\mathcal{E}(W_B, W_F) := \{1 + W_B, ..., T - W_F + 1\}$.[10] Then for every $\tau \in \{-W_B, ..., W_F - 1\}$, define:

$$\beta_{\tau}^{ATT}(W_B, W_F) := \frac{\sum_{e \in \mathcal{E}(W_B, W_F)} \beta_{e,e+\tau}^{ATT} \mathbb{P}(H_{ie} = 1)}{\sum_{e \in \mathcal{E}(W_B, W_F)} \mathbb{P}(H_{ie} = 1)} \tag{6}$$

For each τ, the parameter $\beta_{\tau}^{ATT}(W_B, W_F)$ is the ATT of an event τ periods relative to the event among units that experienced an event during $\mathcal{E}(W_B, W_F)$. Intuitively, β_{τ}^{ATT} is a weighted average of the elements of $\{\beta_{et}^{ATT} : (e, t) \in \mathcal{E} \times \mathcal{T}\}$ for values of e, t which are exactly τ periods apart. It

[9] Callaway and Sant'Anna (2021) proposed similar aggregates in settings where units experience one event at most.

[10] In a balanced panel, each unit that experiences an event any period $e \in \mathcal{E}(W_B, W_F)$ is observed for at least W_B periods strictly before e and for at least W_F periods weakly after e.

weights each β_{et}^{ATT} by the proportion of units treated in period e. The vector $\left(\beta_{\tau}^{ATT}(W_B, W_F) : \tau \in \{-W_B, ..., W_F - 1\}\right)$ may be plotted to summarize the average evolution of the ATT.

Finally, the set of parameters $\{\beta_{et}^{ATT} : (e,t) \in \mathcal{E} \times \mathcal{T}\}$ can also be summarized by a single parameter. Define:

$$\beta^{ATT} := \sum_{e=2}^{T} \left[\sum_{t=e}^{T} \beta_{et}^{ATT} \left(\frac{1}{T - (e-1)} \right) \right] \left(\frac{\mathbb{P}(H_{ie} = 1)}{\sum_{k=1}^{T} \mathbb{P}(H_{ik} = 1)} \right), \tag{7}$$

This is a weighted average of the elements of $\{\beta_{et}^{ATT} : (e,t) \in \mathcal{E} \times \mathcal{T}\}$ among (e,t) such that $e \leq t$. Within e, each element of $\left\{\beta_{et}^{ATT} : t \in \{e, ..., T\}\right\}$ is equally weighted. Across e, the weights correspond to the fraction of units that are treated in period e. Researchers may alternatively aggregate $\{\beta_{et}^{ATT} : (e,t) \in \mathcal{E} \times \mathcal{T}\}$ into a single parameter in a way that is consistent with the vector $\left(\beta_{\tau}^{ATT}(W_B, W_F) : \tau \in \{-W_B, ..., W_F - 1\}\right)$, as follows:

$$\begin{aligned} \beta^{ATT}(W_B, W_F) &:= \sum_{e \in \mathcal{E}(W_B, W_F)} \left[\sum_{t=e}^{e+W_F-1} \beta_{et}^{ATT} \left(\frac{1}{W_F} \right) \right] \left(\frac{\mathbb{P}(H_{ie} = 1)}{\sum_{k \in \mathcal{E}(W_B, W_F)} \mathbb{P}(H_{ik} = 1)} \right) \\ &= \frac{1}{W_F} \sum_{\tau=0}^{W_F - 1} \beta_{\tau}^{ATT}(W_B, W_F) \end{aligned} \tag{8}$$

D. Illustrations

In this section, I graphically illustrate the identification strategy and aggregations introduced in sections I.B and I.C. Consider a setting with $T = 6$ and $\Omega_H = \{0,1\}^6$. Under no anticipation and parallel trends, ATTs for periods 2 through 6 are identified. Assume that we are interested in identifying the ATT of a period 3 event for units that experience the event history $(0,0,1,0,1,0)$. Under no anticipation and parallel trends, the ATT in each period $t \in \mathcal{T}$ of experiencing a period 3 event among $A := \{i : H_i = (0,0,1,0,1,0)\}$ is identified as the difference-in-differences given in equation (3) between units in A and unit in $B := \{i : H_i = (0,0,0,0,1,0)\}$, as follows:

$$\beta_{3t}^{ATT}(0,0,-,0,1,0) = \mathbb{E}\left[Y_{it} - Y_{i2} \,\middle|\, H_i = (0,0,1,0,1,0)\right] - \mathbb{E}\left[Y_{it} - Y_{i2} \,\middle|\, H_i = (0,0,0,0,1,0)\right]$$

Furthermore, $\left(\beta_{31}^{ATT}(0,0,-,0,1,0), ..., \beta_{36}^{ATT}(0,0,-,0,1,0)\right)$ is the evolution of the ATT of the period 3 event among units in A. Figure 1 plots the observed evolution of outcomes among units in A and units in B, illustrating how a comparison between these two groups can identify ATTs under no anticipation and parallel trends. In this figure, each event raises the outcome of interest by $3 \cdot (0.5)^{\text{\# of past events}}$. Figures 2 and 3 plot two additional examples for units that experience the e-histories $(0, 0, -, 0, 0, 0)$ and $(0, 1, -, 0, 1, 1)$, respectively.

Moreover, under no anticipation and parallel trends, the evolution of the ATT for units treated in period 3 who experience *any* e-history $H^{-3} \in \Omega^*_{H^{-3}}$ is identified. For any period $t \in \{1, ..., 6\}$, the identified parameters may be arranged in a table, as follows

e-history	Parameter	Weight
$(0, 0, -, 0, 0, 0)$	$\beta_{3t}^{ATT}(0, 0, -, 0, 0, 0)$	$\mathbb{P}\big(H_i^{-3} = (0, 0, -, 0, 0, 0) \,\big\vert\, H_{i3} = 1\big)$
$(1, 0, -, 0, 0, 0)$	$\beta_{3t}^{ATT}(1, 0, -, 0, 0, 0)$	$\mathbb{P}\big(H_i^{-3} = (1, 0, -, 0, 0, 0) \,\big\vert\, H_{i3} = 1\big)$
$\vdots$	$\vdots$	$\vdots$
$(1, 1, -, 1, 1, 1)$	$\beta_{3t}^{ATT}(1, 1, -, 1, 1, 1)$	$\mathbb{P}\big(H_i^{-3} = (1, 1, -, 1, 1, 1) \,\big\vert\, H_{i3} = 1\big)$

By averaging all $\beta_{3t}^{ATT}(\cdot)$ in the middle column using the probabilities in the right column as weights, researchers may then construct β_{3t}^{ATT} as defined in equation (5)—the ATT in period t of experiencing a period 3 event. By repeating this process for all t, researchers may construct $\big(\beta_{31}^{ATT}, ..., \beta_{36}^{ATT}\big)$. This is the average evolution of the ATT among units treated in period 3. Researchers may repeat this process for periods 2, 4, 5, and 6 to construct $\big(\beta_{et}^{ATT} : (e, t) \in \{2, ..., 6\} \times \{1, ..., 6\}\big)$. Figure 4 plots $\big(\beta_{et}^{ATT} : t \in \{1, ..., 6\}\big)$ for each $e \in \{2, ..., 6\}$ for some example treatment effect functions. Each line represents the average evolution of the ATT of each period e event.

These parameters may be aggregated further. Consider the parameters $\big(\beta_\tau^{ATT}(W_B, W_F) : \tau \in \{-W_B, ..., W_F - 1\}\big)$ as defined by equation (6). Continuing with the foregoing example, consider the cases that (W_B, W_F) is equal to either $(2, 2)$, $(3, 3)$, or $(2, 3)$. Then $\mathcal{E}(W_B, W_F)$ may be either $\mathcal{E}(2, 2) = \{3, 4, 5\}$, $\mathcal{E}(3, 3) = \{4\}$, or $\mathcal{E}(2, 3) = \{3, 4\}$. For each period t and for each period $e \in \mathcal{E}(W_B, W_F)$, the weight given to β_{et}^{ATT} in the construction of $\beta_{t-e}^{ATT}(W_B, W_F)$ is the share of units treated in period e among the units treated in any period $e' \in \mathcal{E}(W_B, W_F)$. For example, if $(W_B, W_F) = (2, 3)$, $\frac{|\{i: H_{ie}=3\}|}{|\{i: H_{ie}=3 \vee H_{ie}=4\}|} = \frac{1}{3}$, and $\frac{|\{i: H_{ie}=4\}|}{|\{i: H_{ie}=3 \vee H_{ie}=4\}|} = \frac{2}{3}$, then the weight given to period 3's average evolution of the ATT is $\frac{1}{3}$ and the weight given to period 4's average evolution of the ATT is $\frac{2}{3}$. Continuing with the example presented in figure 4, the average evolution of the ATT within the event window $(2, 3)$ is plotted in figure 5 for the case that $\frac{|\{i: H_{ie}=3\}|}{|\{i: H_{ie}=3 \vee H_{ie}=4\}|} = \frac{1}{3}$ and $\frac{|\{i: H_{ie}=4\}|}{|\{i: H_{ie}=3 \vee H_{ie}=4\}|} = \frac{2}{3}$. Notice that although an additional point is observed for the period 3 event after relative time zero (i.e., the point at relative time 3), researchers should not include this point in the overall average evolution of the ATT, because doing so could misleadingly show an upward-sloping dynamic treatment effect rather than a static treatment effect.

Finally, researchers can further aggregate these parameters using equations (7) or (8). Equation (8) produces an equal-weighted average of the elements of $\big(\beta_\tau^{ATT}(W_B, W_F) : \tau \in \{-W_B, ..., W_F -$

$1\})$. For example, consider figure 5. Equation (8) defines $\beta^{ATT}(2,3)$ as an equal-weighted average of $\beta_0^{ATT}(2,3)$, $\beta_1^{ATT}(2,3)$, and $\beta_2^{ATT}(2,3)$. Equation (7) produces an analogous average using all $\{\beta_{et}^{ATT} : t \geq e \geq 2\}$, not just those included in the construction of $\left(\beta_\tau^{ATT}(W_B, W_F) : \tau \in \{-W_B, ..., W_F - 1\}\right)$. However, the parameter defined in equation (7) may appear inconsistent with $\left(\beta_\tau^{ATT}(W_b, W_F) : \tau \in \{-W_B, ..., W_F - 1\}\right)$. It also implicitly assigns more weight to earlier events because earlier events have more post-event periods.

II. Estimation

A. Matching estimator

I assume that a balanced panel of N units is drawn independently from a common population. Let $\mathcal{I} := \{1, ..., N\}$ denote the set of units in the sample. For each unit $i \in \mathcal{I}$ and each history $H \in \Omega_H$, let $Y_i(H) := (Y_{it}(H) : t \in \mathcal{T})$ denote the vector of potential outcomes and let $Y_i := Y_i(H_i)$ denote the vector of observed outcomes. The data are $\{(Y_i, H_i) : i \in \mathcal{I}\}$.

For all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$, I propose the following estimator of $\beta_{et}^{ATT}(H^{-e})$:

$$\hat{\beta}_{et}^{ATT}(H^{-e}) := \frac{\sum_{i:H_i=(H^{-e},1)} Y_{it} - Y_{ie-1}}{\left|\{i : H_i = (H^{-e}, 1)\}\right|} - \frac{\sum_{i:H_i=(H^{-e},0)} Y_{it} - Y_{ie-1}}{\left|\{i : H_i = (H^{-e}, 0)\}\right|} \tag{9}$$

No anticipation, parallel trends, and the law of large numbers imply that $\hat{\beta}_{et}^{ATT}(H^{-e}) \xrightarrow{p} \beta_{et}^{ATT}(H^{-e})$ as $N \to \infty$ for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$.

Recall the vector of parameters $\theta := \left(\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}\right)$ defined in section I. Index the elements of θ with the set of indices $\mathcal{K} := \{1, ..., K\}$. Then we may write that $\theta := (\theta_k : k \in \mathcal{K})$, where each θ_k is equal to a corresponding $\beta_{et}^{ATT}(H^{-e})$. Define the vector $\hat{\theta} := \left(\hat{\theta}_k : k \in \mathcal{K}\right)$. Let the covariance matrix of $\hat{\theta}$ be denoted $V_{\hat{\theta}}$. For all $(k, H_e) \in \mathcal{K} \times \{0, 1\}$, define $N_k(H_e) := |\{i : H_i = (H^{-e}, H_e)\}|$. For each $(i, t, e) \in \mathcal{I} \times \mathcal{T} \times \mathcal{E}$, define $\Delta_{ite} := Y_{it} - Y_{ie-1}$. Then for all $(k_1, k_2) \in \mathcal{K}^2$ corresponding to some $\left((e_1, t_1, H_1^{-e_1}), (e_2, t_2, H_2^{-e_2})\right)$, the corresponding element of $V_{\hat{\theta}}$ is given by:

$$\begin{aligned}
\mathbb{C}\left[\hat{\theta}_{k_1}, \hat{\theta}_{k_2}\right] &= N_{k_1}^{-1}(1)\mathbb{C}\left[\Delta_{it_1e_1}, \Delta_{it_2e_2} | H_i = (H_1^{-e_1}, 1)\right] \mathbb{1}\left[\{i : H_i = (H_1^{-e_1}, 1) = (H_2^{-e_2}, 1)\} \neq \emptyset\right] \\
&\quad - N_{k_1}^{-1}(1)\mathbb{C}\left[\Delta_{it_1e_1}, \Delta_{it_2e_2} | H_i = (H_1^{-e_1}, 1)\right] \mathbb{1}\left[\{i : H_i = (H_1^{-e_1}, 1) = (H_2^{-e_2}, 0)\} \neq \emptyset\right] \\
&\quad - N_{k_1}^{-1}(0)\mathbb{C}\left[\Delta_{it_1e_1}, \Delta_{it_2e_2} | H_i = (H_1^{-e_1}, 0)\right] \mathbb{1}\left[\{i : H_i = (H_1^{-e_1}, 0) = (H_2^{-e_2}, 1)\} \neq \emptyset\right]
\end{aligned}$$

$$+ N_{k_1}^{-1}(0)\mathbb{C}\big[\Delta_{it_1e_1}, \Delta_{it_2e_2}|H_i = (H_1^{-e_1}, 0)\big]\,\mathbb{1}\big[\{i : H_i = (H_1^{-e_1}, 0) = (H_2^{-e_2}, 0)\} \neq \emptyset\big] \tag{10}$$

For proof, see Appendix B. Intuitively, this expression states that if each unit is drawn independently, then the covariance between any $\hat{\beta}_{e_1t_1}^{ATT}(H_1^{-e_1})$ and $\hat{\beta}_{e_2t_2}^{ATT}(H_2^{-e_2})$ depends only on the units contributing to both estimators. For example, if a unit i is among the treated group for the purposes of estimating $\hat{\beta}_{e_1t_1}^{ATT}(H_1^{-e_1})$ and among the control group for the purposes of estimating $\hat{\beta}_{e_2t_2}^{ATT}(H_2^{-e_2})$, then it will contribute to the covariance between $\hat{\beta}_{e_1t_1}^{ATT}(H_1^{-e_1})$ and $\hat{\beta}_{e_2t_2}^{ATT}(H_2^{-e_2})$ through the second summand of the foregoing expression.

Researchers can estimate $V_{\hat{\theta}}$ using the plug-in estimator $\hat{V}_{\hat{\theta}}$ constructed from equation (10) by replacing the covariance operator $\mathbb{C}$ with its sample analogue $\hat{\mathbb{C}}$. The standard error of any given $\hat{\beta}_{et}^{ATT}(H^{-e})$ is the standard deviation of a difference-in-means:

$$\hat{\sigma}_{\hat{\beta}_{et}^{ATT}(H^{-e})} := \sqrt{\frac{\left(\hat{\sigma}_{Y_{it}-Y_{ie-1}}^{1}\right)^2}{\left|\{i : H_i = (H^{-e}, 1)\}\right|} + \frac{\left(\hat{\sigma}_{Y_{it}-Y_{ie-1}}^{0}\right)^2}{\left|\{i : H_i = (H^{-e}, 0)\}\right|}}$$

where $\hat{\sigma}_{Y_{it}-Y_{ie-1}}^{1}$ is the sample standard deviation of $Y_{it} - Y_{ie-1}$ among units $\{i : H_i = (H^{-e}, 1)\}$ and $\hat{\sigma}_{Y_{it}-Y_{ie-1}}^{0}$ is the sample standard deviation of $Y_{it} - Y_{ie-1}$ among units $\{i : H_i = (H^{-e}, 0)\}$.

In section I, I introduced aggregates of the elements of $\left\{\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega_{H^{-e}}^{*}\right\}$ to summarize the average treatment effects on the treated. I propose estimating aggregates such as β_{et}^{ATT}, $\beta_{\tau}^{ATT}(W_B, W_F)$, β^{ATT}, and $\beta^{ATT}(W_B, W_F)$ with plug-in estimators (i.e., using $\hat{\beta}_{et}^{ATT}(H^{-e})$ and frequency estimators of probabilities $\mathbb{P}(\cdot)$). Consistency of these estimators follows from the continuous mapping theorem. The standard errors of the aggregate parameters can be computed with the block bootstrap.

B. Stacked regression

The foregoing matching estimator transparently estimates ATTs under the generalized no anticipation and parallel trends assumptions in settings where units experience multiple events. It is transparent in the sense that it explicitly computes the difference-in-differences in equation (3) and aggregates the elements of $(\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega_{H^{-e}}^{*})$ using easy-to-interpret weights. In this section, I introduce an equivalent regression-based estimator which may be easier to implement in practice.

Continue considering the dataset $\{(Y_i, H_i) : i \in \mathcal{I}\}$ introduced in section II.A. Let $\big\{(e, Y_i, H_i) : (e, i) \in \mathcal{E} \times \mathcal{I}$ be the "stacked" version of this dataset, wherein copies of $\{(Y_i, H_i) : i \in \mathcal{I}\}$ are appended upon one another for every $e \in \mathcal{E}$. That is, let $(e, Y_i, H_i) = (e', Y_i, H_i)$ for all $(e, e') \in \mathcal{E}^2$. In each stack e, keep only those units satisfying $H_i^{-e} \in \Omega^*_{H^{-e}}$. Consider the vector θ defined in section I.B, let k index the elements of θ, let K be the number of elements of θ, and let $\tilde{k}$ map values of (e, t, H^{-e}) to the indices of θ. Consider the regression model:

$$Y_{iet} = \sum_{k=1}^{K} \theta_k \mathbb{1}[\tilde{k}(e, t, H_i^{-e}) = k]\mathbb{1}[H_{ie} = 1] + \mathrm{FE}_{ei} + \mathrm{FE}_{etH_i^{-e}} + \varepsilon_{iet} \tag{11}$$

where $Y_{iet} = Y_{it}$ in stack e, FE_{ei} is an e-specific unit fixed effect and $\mathrm{FE}_{etH_i^{-e}}$ is an (e, H_i^{-e})-specific period fixed effect. The OLS estimator of each element of θ consistently estimates the corresponding $\beta^{ATT}_{et}(H^{-e})$.

Next, consider an event window (W_B, W_F), let $\mathcal{W} := \{-W_B, ..., W_F - 1\} \setminus \{-1\}$, and let $\mathcal{E}(W_B, W_F)$ be as defined in section I.C. Consider the subset of the foregoing stacked dataset given by $\{(e, Y_i, H_i) : (e, i) \in \mathcal{E}(W_B, W_F) \times \mathcal{I}\}$. Consider the regression models:

$$Y_{iet} = \sum_{\tau \in \mathcal{W}} \rho_\tau \mathbb{1}[i \text{ was treated in } e \text{ and } t = e + \tau] + \mathrm{FE}_{ei} + \mathrm{FE}_{etH_i^{-e}} + \varepsilon_{iet} \tag{12}$$

$$Y_{iet} = \rho \mathbb{1}[i \text{ was treated in } e \text{ and } t \geq e] + \mathrm{FE}_{ei} + \mathrm{FE}_{etH_i^{-e}} + \varepsilon_{iet} \tag{13}$$

Gardner (2021) derived the weight given to each observation in a stacked dataset by the OLS estimator of an event study regression model. The weight given to each observation within each (e, H^{-e}) is increasing in the variance of the treatment indicator within (e, H^{-e}) and the number of units within (e, H^{-e}) relative to the number of units in the full dataset. For each $(e, H^{-e}, H) \in \mathcal{E}(W_B, W_F) \times \Omega^*_{H^{-e}} \times \{0, 1\}$, define $N(e, H^{-e}, H) := |\{i : H_i = (H^{-e}, H)\}|$. For each $(e, h) \in \mathcal{E}(W_B, W_F) \times \Omega^*_{H^{-e}}$, define the weight:

$$\omega_{eh} := \underbrace{\left(\frac{\kappa_{eh}}{\sum_{e' \in \mathcal{E}(W_B, W_F)} \sum_{h' \in \Omega^*_{H^{-e'}}} \kappa_{eh}} \right)^{-1}}_{\text{Inverse of Gardner weights}} \underbrace{\left(\frac{N(e, h, 1)}{\sum_{e' \in \mathcal{E}(W_B, W_F)} \sum_{h' \in \Omega^*_{H^{-e'}}} N(e', h', 1)} \right)}_{(e,h) \text{ pairs with more treated units receive more weight}} \tag{14}$$

where

$$\kappa_{eh} := \underbrace{\left(\frac{N(e, h, 1)N(e, h, 0)}{\left(N(e, h, 1) + N(e, h, 0)\right)^2} \right)}_{\text{Variance of } H_{ie} \text{ within } (e, H^{-e})} \underbrace{\left(\frac{N(e, h, 1) + N(e, h, 0)}{\sum_{e' \in \mathcal{E}(W_B, W_F)} \sum_{h' \in \Omega^*_{H^{-e'}}} N(e', h', 1) + N(e', h', 0)} \right)}_{\text{Size of stack } (e,h) \text{ relative to the full dataset}}$$

The weighted least squares estimators of $(\rho_\tau : \tau \in \mathcal{W})$ in equation (12) and ρ in equation (13)—where each observation (i, e, t) is weighted by $\omega_{eH_i^{-e}}(W_B + W_F)^{-1}\big(N(e, H_i^{-e}, 1) + N(e, H_i^{-e}, 0)\big)^{-1}$ —consistently estimate $(\beta_\tau^{ATT}(W_B, W_F) : \tau \in \mathcal{W})$ and $\beta^{ATT}(W_B, W_F)$. The standard errors may be computed with the block bootstrap.

III. Testing parallel pre-trends

Researchers conducting event studies in settings where units experience one event at most commonly scrutinize the plausibility of the no anticipation and parallel trends assumptions by examining whether trends among treated and untreated units had been moving synchronously before the treated units' event. Researchers conducting event studies in settings where units experience multiple events may use similar reasoning to scrutinize the no anticipation and parallel trends assumptions introduced in section I. Consider the setting introduced in section I.A and assume that $T \geq 3$.

Under no anticipation and parallel pre-trends, it follows that for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$ such that $t < e$, $\beta_{et}^{ATT}(H^{-e})$ is identified by equation (3) and $\beta_{et}^{ATT}(H^{-e}) = 0$. Define the null hypothesis:

$$\mathbb{H}_0 : \quad \beta_{et}^{ATT}(H^{-e}) = 0 \qquad \text{for all } (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}} \text{ such that } t < e$$

Testing $\mathbb{H}_0$ may serve as a proxy test of the no anticipation and parallel post-trends assumptions in the sense that if for any e-history $H^{-e} \in \Omega^*_{H^e}$, average trends among units $\{i : H_i = (H^{-e}, 1)\}$ differ from average trends among units $\{i : H_i = (H^{-e}, 0)\}$ prior to period e, then they might have also differed after period e had units $\{i : H_i = (H^{-e}, 1)\}$ experienced the history $(H^{-e}, 0)$ instead.

I propose the following test of $\mathbb{H}_0$. Define the vector $\delta := \big(\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}} \text{ and } t < e\big)$, and define its estimator $\hat{\delta}$ based on equation (9) analogously. Note that δ and $\hat{\delta}$ are subvectors of θ and $\hat{\theta}$ defined in section I, respectively. As $V_{\hat{\theta}}$ was defined in section I for $\hat{\theta}$, so too define $V_{\hat{\delta}}$ for $\hat{\delta}$. Then the Wald statistic for $\mathbb{H}_0$ is given by $W := \hat{\delta}'(\hat{V}_{\hat{\delta}})^{-1}\hat{\delta}$ where $\hat{V}_{\hat{\delta}}$ is the plug-in estimator of $V_{\hat{\delta}}$. As the matrix A was defined for θ by equation (4), so too define the matrix A' for δ. Then under the null hypothesis $\mathbb{H}_0$, $W \xrightarrow{d} \chi^2_q$, where $q := \text{rank}(A')$. For a given test size α, the asymptotic critical value $\bar{W}$ of the parallel pre-trends test satisfies $1 - \alpha = G_q(\bar{W})$, where $G_q(\cdot)$ is the CDF of a χ^2_q random variable.

IV. Monte Carlo simulations

I evaluate the estimators introduced in section II and the parallel pre-trends test introduced in section III in a series of Monte Carlo simulations. The simulations differ according to their treatment effect functions: I consider static, dynamic, stationary, non-stationary, history independent, and history dependent treatment effects. I also discuss the MDO estimator proposed by Sandler and Sandler (2014) and revisited by Schmidheiny and Siegloch (2019).

To the best of my knowledge, Sandler and Sandler (2014) were the first to consider event studies in settings where units may experience multiple events. Given an event window (W_B, W_F) chosen by the researcher, they proposed the OLS estimator of the MDO regression model:

$$Y_{it} = \sum_{\tau \in \mathcal{W}} \beta_\tau^{MDO} D_{it\tau} + \mathrm{FE}_i + \mathrm{FE}_t + \varepsilon_{it}, \tag{15}$$

where β_τ^{MDO} are parameters of interest, $\mathcal{W} := \{-W_B, ..., W_F - 1\} \setminus \{-1\}$, and

$$D_{it\tau} := \begin{cases} \sum_{k=t-\tau}^{T} H_{ik} & \text{if } \tau = -W_B \\ H_{it-\tau} & \text{if } \tau \in \mathcal{W} \setminus \{-W_B, W_F - 1\} \\ \sum_{k=1}^{t-\tau} H_{ik} & \text{if } \tau = W_F - 1 \end{cases}$$

Sandler and Sandler (2014) showed that their approach performs well when treatment effects are static, stationary, and history independent. They also showed that it performs well when treatment effects are dynamic, stationary, and history independent, provided that the event window (W_B, W_F) is wide enough to include all dynamics. I show that the matching estimator introduced in section II also performs well in these settings, as well as in settings where treatment effects are non-stationary, history dependent, or dynamic beyond the range of the event window.[11] Examining estimates of $(\beta_\tau^{MDO} : \tau < 0)$ to conduct a parallel pre-trends test will fail to detect violations of the no anticipation and parallel trends assumptions at the (e, t, H^{-e})-level that average to zero when aggregated to relative time.[12]

I program the Monte Carlo simulations as follows. In each simulation, I create 50,000 units

[11]The OLS estimator of the MDO regression model consistently estimates ATTs when equation (15) correctly specifies the data-generating process. But it is biased if the ATTs are dynamic beyond the range of the event window, non-stationary, or history-dependent. For example, if each unit experiences one event at most, then the MDO regression model is equivalent to equation (2). Sun and Abraham (2021) showed that if treatment effects are non-stationary, then the OLS estimator of each $\hat{\beta}$ in (2) incorporates treatment effects $\tau' \neq \tau$ periods relative to treatment.

[12]Callaway and Sant'Anna (2021) made a similar observation for settings where units experience one event at most and the identifying assumptions are stated conditional on covariates X.

$\mathcal{I} := \{1, ..., 50,000\}$. I report Ω_H in figure 6. Figure 7 plots the corresponding distribution of events in relative time for the event window $(W_B, W_F) = (4, 4)$. It illustrates that given Ω_H and the distribution H_i over Ω_H chosen for the simulations, many units experience multiple events.[13] For each (i, t), I define $Y_{it}(H_i) = \sum_{e=1}^{T} \beta_{et}^{ATT}(H_i^{-e})H_{ie} + i + t + \varepsilon_{it}$ where $\varepsilon_{it} \overset{\text{iid}}{\sim} U[-1, 1]$. I consider several treatment effect functions. For each treatment effect function, $\beta_{iet}^{ATT}(H^{-e}) = 0$ for all $t < e$. For $t \geq e$, $\beta_{iet}^{ATT}(H^{-e})$ is given by:

6	static, stationary, and hist. indep.
$6 + 7(t-e) - 0.9(t-e)^2$	dynamic, stationary, and hist. indep.
$e^{1.5} + 7(t-e) - 0.9(t-e)^2$	dynamic, non-stationary, and hist. indep.
$\left(6 + 7(t-e) - 0.9(t-e)^2\right)0.8^{\sum_{e=1}^{t-1} H_{ie}}$	dynamic, stationary, and hist. dep.
$\left(e^{1.5} + 7(t-e) - 0.9(t-e)^2\right)0.8^{\sum_{e=1}^{t-1} H_{ie}}$	dynamic, non-stationary, and hist. dep.

In these examples, I model history dependence using a decreasing-returns-to-treatment function. That is, the magnitude of each event's treatment effect is diminishing in the number of prior events. I conduct 500 independent simulations, each consisting of one draw of $\left(H_i : i \in \mathcal{I}\right)$ and five independent draws of $\left(\varepsilon_{it} : (i, t) \in \mathcal{I} \times \mathcal{T}\right)$, one for each of the five treatment effect functions.

In figure 8 and 9-13, I compare the matching estimator introduced in section II.A, the weighted least squares estimator of the stacked regression model introduced in section II.B, the OLS estimator of the stacked regression model, and the MDO estimator. The weighted least squares estimator and the matching estimator perform well in every simulation. The MDO estimator performs well when the treatment effect is static, stationary, and history-independent. It is biased otherwise. For example, its estimates of pre-event ATTs are often non-zero. The OLS estimator of the stacked regression model performed well when the treatment effect is stationary and history-independent. When the treatment effect is non-stationary or history-dependent, the OLS estimator of the stacked regression model assigns different weights to each $\beta_{et}^{ATT}(H^{-e})$ than the matching estimator or the weighted least squares estimator of the stacked regression model.

Finally, Figure 14 plots the distribution of Wald statistics produced by the simulations of the parallel pre-trends test. The distribution of Wald statistics is approximately equal to the PDF of χ^2_{63}, as expected under the null hypothesis for these simulations.[14]

[13]Note that each unit is untreated in period 1 for simplicity. If some units were treated in period 1, then the estimator would match groups of units based on period 1 events as well as on all other events, although the effect of the period one event would not be identified.

[14]Although the simulations are designed such that there are 98 parameters in the set $\left\{\beta_{et}^{ATT}(H^{-e}) : (e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T}^e \times \Omega^*_{H^{-e}} \text{ and } t < e\right\}$, only 63 of them are linearly independent in these simulations.

V. Practical issues and extensions

A. Unmatched units

No anticipation and parallel trends identify $\beta_{et}^{ATT}(H^{-e})$ for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T} \times \Omega^*_{H^{-e}}$. However, in finite samples, there may be some $H^{-e} \in \Omega^*_{H^{-e}}$ for which either $\{i : H_i = (H^{-e}, 1)\} = \emptyset$, $\{i : H_i = (H^{-e}, 0)\} = \emptyset$, or both. For any such H^{-e}, no estimate of $\beta_{et}^{ATT}(H^{-e})$ can be produced using the matching estimator or the weighted least squares estimator. Furthermore, estimates of β_{et}^{ATT}, $\beta_{\tau}^{ATT}(W_B, W_F)$, $\beta^{ATT}(W_B, W_F)$, and β^{ATT} will not include some elements of the support of the underlying treatment effect distribution.[15] Imai, Kim, and Wang (2021) identified a similar issue for their matching estimator in settings where units experience one event at most and a unit's treated status may switch on and off. I recommend—as they do—that researchers address this issue by reporting the number and characteristics of unmatched units and discussing context-specific implications of unmatchedness in their research setting.

B. Short-lived treatment effect dynamics

In some contexts, treatment effect dynamics may be short lived in the sense that period-e treated units and untreated units begin moving synchronously again at some point after period e. In such contexts, researchers can estimate the ATT of each period e event by matching units with similar event histories "close" to period e under simple modifications of the generalized no anticipation and parallel trends assumptions. This may reduce the unmatchedness issue discussed in section V.A by reducing the dimension of the matching variable.

C. Multiple events in the same period

In some settings, units may experience multiple events within the same period. If each event has an observable distinguishing characteristic, then researchers can match treated units to untreated units with a similar leave-one-event-out identification strategy under simple modifications of the generalized no anticipation and parallel trends assumptions. That is, in each period e, researchers may match units based on their history of events in other periods, as well as on their history of events in period e, except for one event.

[15] The matching estimator and weighted least squares estimator will use variation generated by at least as many events as would the OLS estimators of (1) and (2) applied to datasets that excluded some events to avoid "contamination."

On the other hand, if within each e all events are indistinguishable, then researchers can match treated units to untreated units as follows. Let D_{ie} denote the number of events that unit i experiences in period e. Let $D_i := (D_{i1}, ..., D_{iT})$ denote unit i's event history and let D_i^{-e} denote unit i's period e event e-history. Redefine potential outcomes as functions $Y_i(D)$. For each period e, for each (D^{-e}, D_e), researchers can match units $\{i : D_i = (D^{-e}, D_e)\}$ to units $\{i : D_i = (D^{-e}, D_e - 1)\}$ to estimate the ATTs of a single event in period e under simple modifications of the generalized no anticipation and parallel trends assumptions.

Relatedly, researchers may consider alleviating the unmatchedness problem in section V.A by coarsening $\mathcal{T}$ into larger periods—e.g., reshaping their data from the quarterly level to the yearly level—and taking this approach. For example, consider the following table of quarterly event indicators for one (of several) years of data:

i	Q1	Q2	Q3	Q4
A	0	0	0	0
B	1	0	0	0
C	0	1	1	0

For simplicity, assume that A, B, and C share event histories in all other quarters. Note that B can be matched to A, but C is left unmatched with respect to both of its events. By coarsening the data to the yearly level, B can still be matched to A, and C can now be matched to B under simple modifications of the generalized no anticipation and parallel trends assumptions.

D. Cumulative treatment effects

The event study framework introduced in this paper can accommodate researchers who are interested in estimating the cumulative effects of multiple events. For ease of exposition, I continue using the language of one event per period.

In order to estimate the cumulative ATT of multiple events, researchers will typically begin by defining a set $\mathcal{G}$ of vectors of periods across which they wish to estimate cumulative ATTs. For example, if $\mathcal{T} := \{1, ..., 10\}$, then researchers who wish to estimate the effect of two consecutive events will define $\mathcal{G} := \{(1, 2), (2, 3), (3, 4), ..., (8, 9), (9, 10)\}$. For further example, researchers who wish to estimate the effect of two events separated by one period will define $\mathcal{G} := \{(1, 3), (2, 4), (3, 5), ..., (8, 10)\}$. In what follows, I use the letter e as an index of elements of $\mathcal{G}$ rather than as an index of $\mathcal{T}$. For each $e \in \mathcal{G}$, let $\underline{e}$ be the earliest coordinate of e chronologically. In other words, if $e = (5, 6, 7, 8)$, then $\underline{e} = \underline{(5, 6, 7, 8)} = 5$.

Consider the event history $H_i := (H_{i1}, ..., H_{iT})$. For every $e \in \mathcal{G}$, define the vector of event indicators $H_{ie} := (H_{it} : t \text{ is a coordinate of } e)$. For example, $H_{i(2,3)} := (H_{i2}, H_{i3})$. Define H_i^{-e} to be the vector of event indicators not included in H_{ie}. For example, $H_i^{-(2,3)} := (H_{i1}, H_{i4}, ..., H_{iT})$. Then the individual-level causal parameter of interest may be defined as:

$$\beta_{iet}(H^{-e}) := Y_{it}(H^{-e}, 1) - Y_{it}(H^{-e}, 0)$$

which should be understood as the effect in period t of experiencing an event in each period contained in e, given an e-history H^{-e}. The ATT may then be defined as:

$$\beta_{et}^{ATT}(H^{-e}) := \mathbb{E}\left[\beta_{iet}(H^{-e}) \,\middle|\, H_i = (H^{-e}, 1)\right]$$

In order to estimate the cumulative effects of multiple events, researchers can match units $\{i : H_i = (H^{-e}, 1)\}$ to units $\{i : H_i = H^{-e}, 0)\}$ under simple modifications of the generalized no anticipation and parallel trends assumptions.

VI. Conclusion

In this paper, I introduced an event study framework for settings where units experience multiple events. I introduced a theoretical model that researchers can use to clarify their identifying assumptions and interpret their estimates. I introduced a matching estimator and a weighted least squares estimator that researchers can use to estimate ATTs. I also introduced a parallel pre-trends test that researchers can use to scrutinize the identifying assumptions in the usual sense. Finally, I demonstrated in a series of Monte Carlo simulations that my estimators and parallel pre-trends test perform well for a wide range of treatment effect functions. This work will be most useful to applied researchers who wish to estimate the ATT of an event in settings where units experience multiple events without necessarily excluding some events from their sample to avoid "contamination."

VII. Bibliography

1. Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. Revisiting event study designs: robust and efficient estimation. Working paper dated August 27, 2021.

2. Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics*, vol. 225: 2021.

3. Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, 1405-1454.

4. Dave, Dhaval M., Andrew I. Friedson, Kyutaro Matsuzawa, Drew McNichols, Connor Redpath, and Joseph J. Sabia. 2020. Did President Trump's Tulsa rally reignite COVID-19? Indoor events and offsetting community effects. NBER Working Paper Series. Working paper 27522.

5. de Chaisemartin, Clement, and Xavier D'Haultfœuille. 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, vol. 110(9): 2020.

6. de Chaisemartin, Clement, and Xavier D'Haultfœuille. Undatd. Two-way fixed effects and difference-in-differences estimators with several treatments. Working paper.

7. Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2019. Long-term care hospitals: a case study in waste. NBER Working Paper No. 24946.

8. Gardner, John. 2021. Two-stage difference in differences. Unpublished manuscript dated April 2021.

9. Goodman-Bacon, Andrew. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, vol. 225: 2021.

10. Greenstone, Michael, Paul Oyer, and Annette Vissing-Jorgensen. 2006. Mandated Disclosure, Stock Returns, and the 1964 Securities Acts Amendments. *The Quarterly Journal of Economics*, pp. 399-460.

11. Imai, Kosuke, In Song Kim, and Erik Wang. Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, vol. 0(0): 2021.

12. Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, vol. 11(4): 2019.

13. Rees-Jones, Alex and Kyle T. Rozema. 2020. Price Isn't Everything: Behavioral Response Around Changes in Sin Taxes. NBER Working Paper 25958.

14. Sandler, Danielle and Ryan Sandler. 2014. Multiple event studies in public finance and labor economics: a simulation study with applications. *Journal of Economic and Social Measurement*, vol. 39, no. 1-2, pp. 31-57.

15. Sarsons, Heather. 2017. Interpreting signals in the labor market: evidence from medical referrals. Working paper.

16. Schmidheiny, Kurt, and Sebastian Siegloch. 2019. On event study designs and distributed-lag models: equivalence, generalization, and practical implications. IZA DP No. 12079.

17. Snow, John. 1855. On the mode of communication of cholera. London: Churchill, January 1855.

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

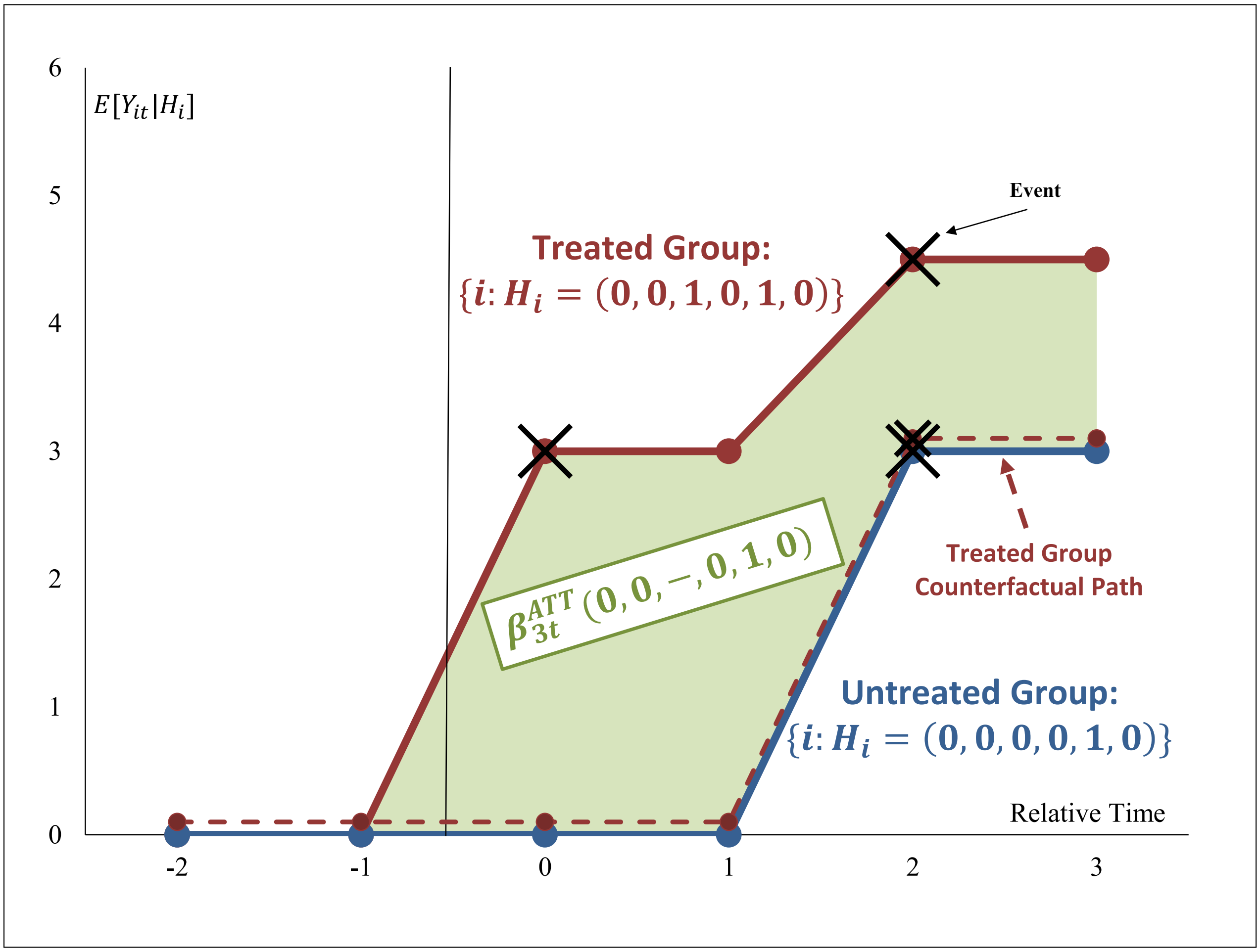

Fig. 1. This figure illustrates the identification strategy described in section I.B. In this example, the ATT of being treated in period 3 given the e-history $(0, 0, -, 0, 1, 0)$ is identified by the difference-in-differences between $\{i : H_i = (0, 0, 1, 0, 1, 0)\}$ and $\{i : H_i = (0, 0, 0, 0, 1, 0)\}$ before-and-after period 3. For clarity, group and period fixed effects are omitted and the counterfactual path for treated units has been offset slightly from the observed path of the untreated units. See the discussion near page 13. See figures 2 and 3 for additional examples.

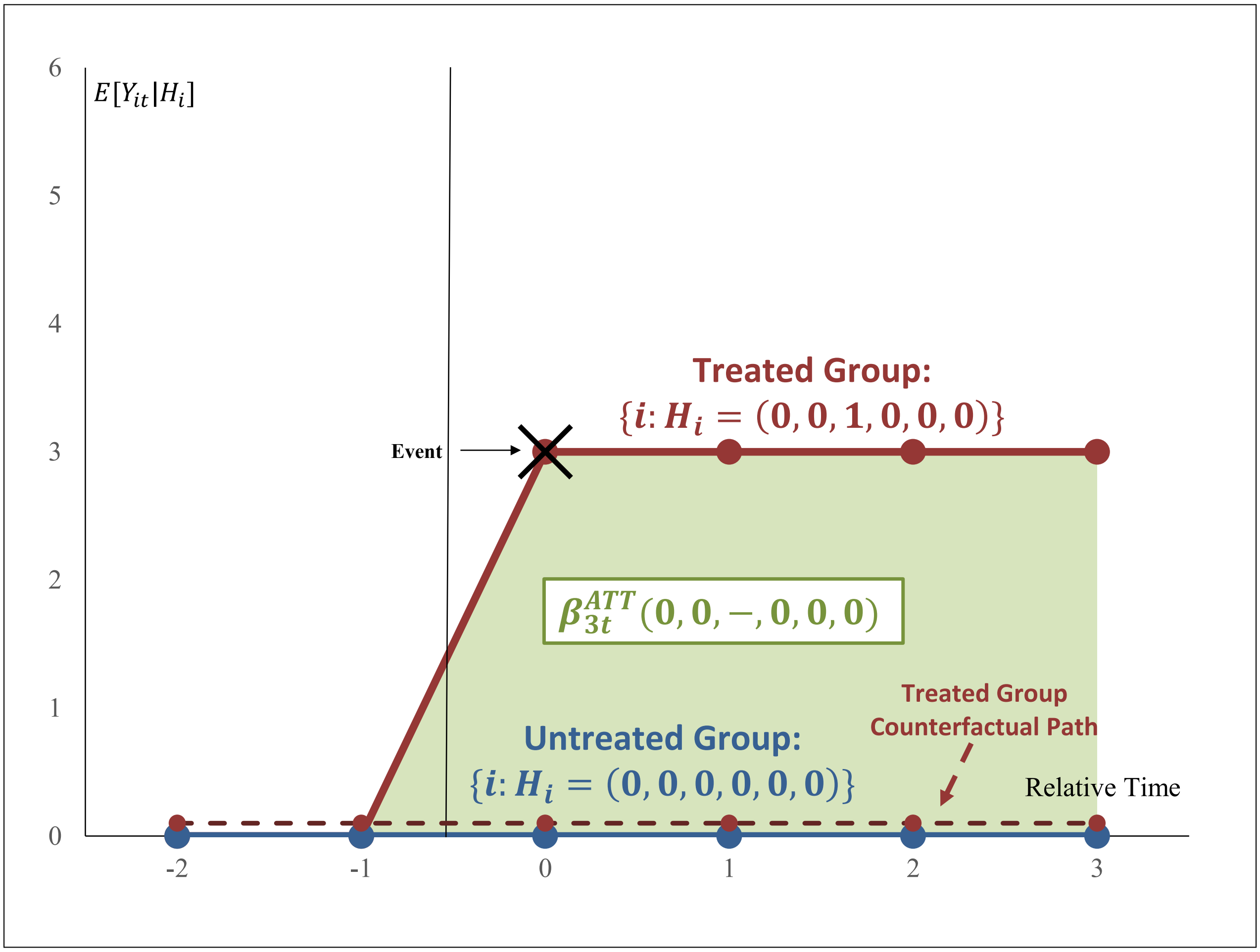

Fig. 2. This figure illustrates the identification strategy described in section I.B. In this example, the ATT of being treated in period 3 given the e-history $(0, 0, -, 0, 0, 0)$ is identified by the difference-in-differences between $\{i : H_i = (0, 0, 1, 0, 0, 0)\}$ and $\{i : H_i = (0, 0, 0, 0, 0, 0)\}$ before-and-after period 3. For clarity, group and period fixed effects are omitted and the counterfactual path for treated units has been offset slightly from the observed path of the untreated units. See the discussion near page 13. See figures 1 and 3 for additional examples.

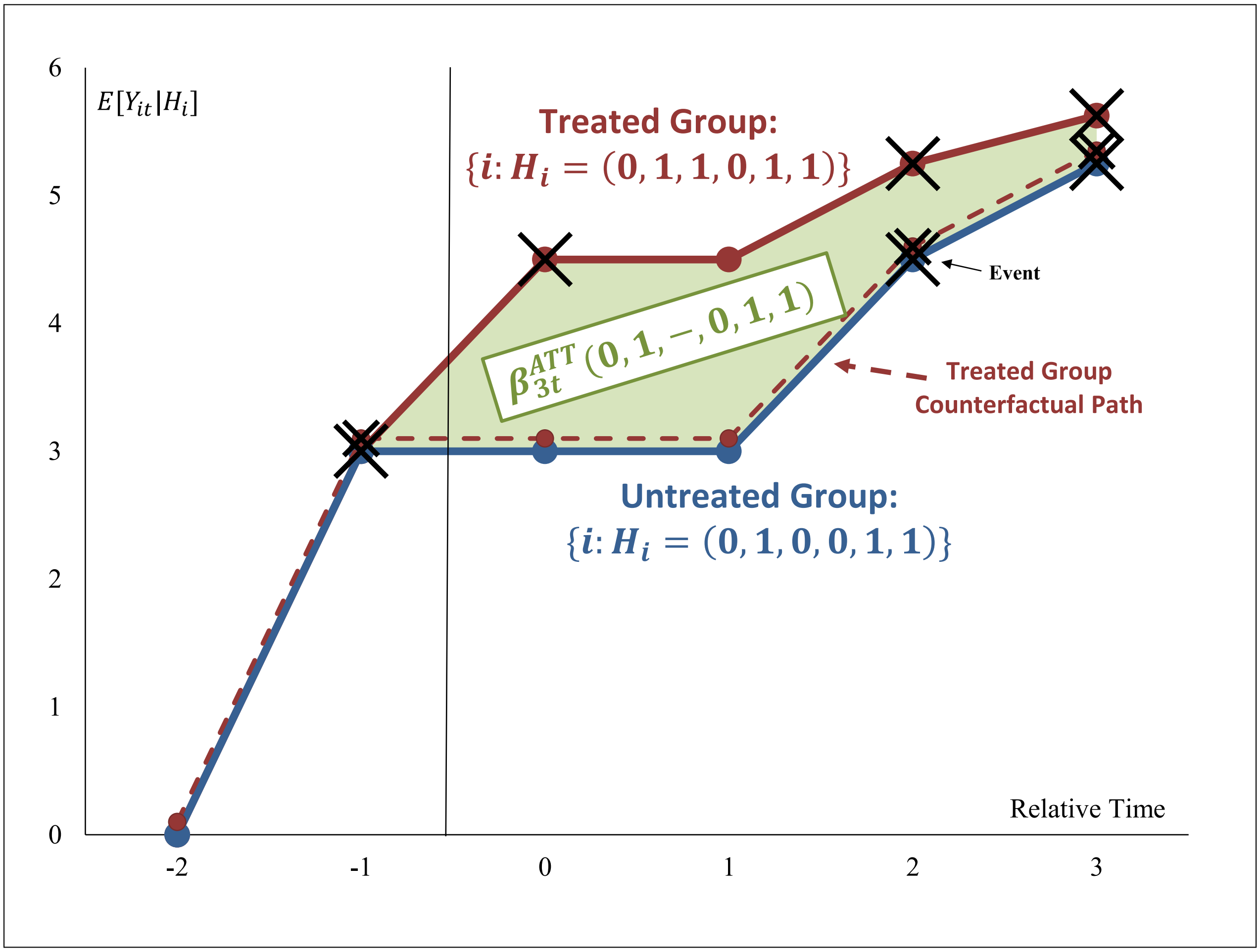

Fig. 3. This figure illustrates the identification strategy described in section I.B. In this example, the ATT of being treated in period 3 given the e-history $(0, 1, -, 0, 1, 1)$ is identified by the difference-in-differences between $\{i : H_i = (0, 1, 1, 0, 1, 1)\}$ and $\{i : H_i = (0, 1, 0, 0, 1, 1)\}$ before-and-after period 3. For clarity, group and period fixed effects are omitted and the counterfactual path for treated units has been offset slightly from the observed path of the untreated units. See the discussion near page 13. See figures 1 and 2 for additional examples.

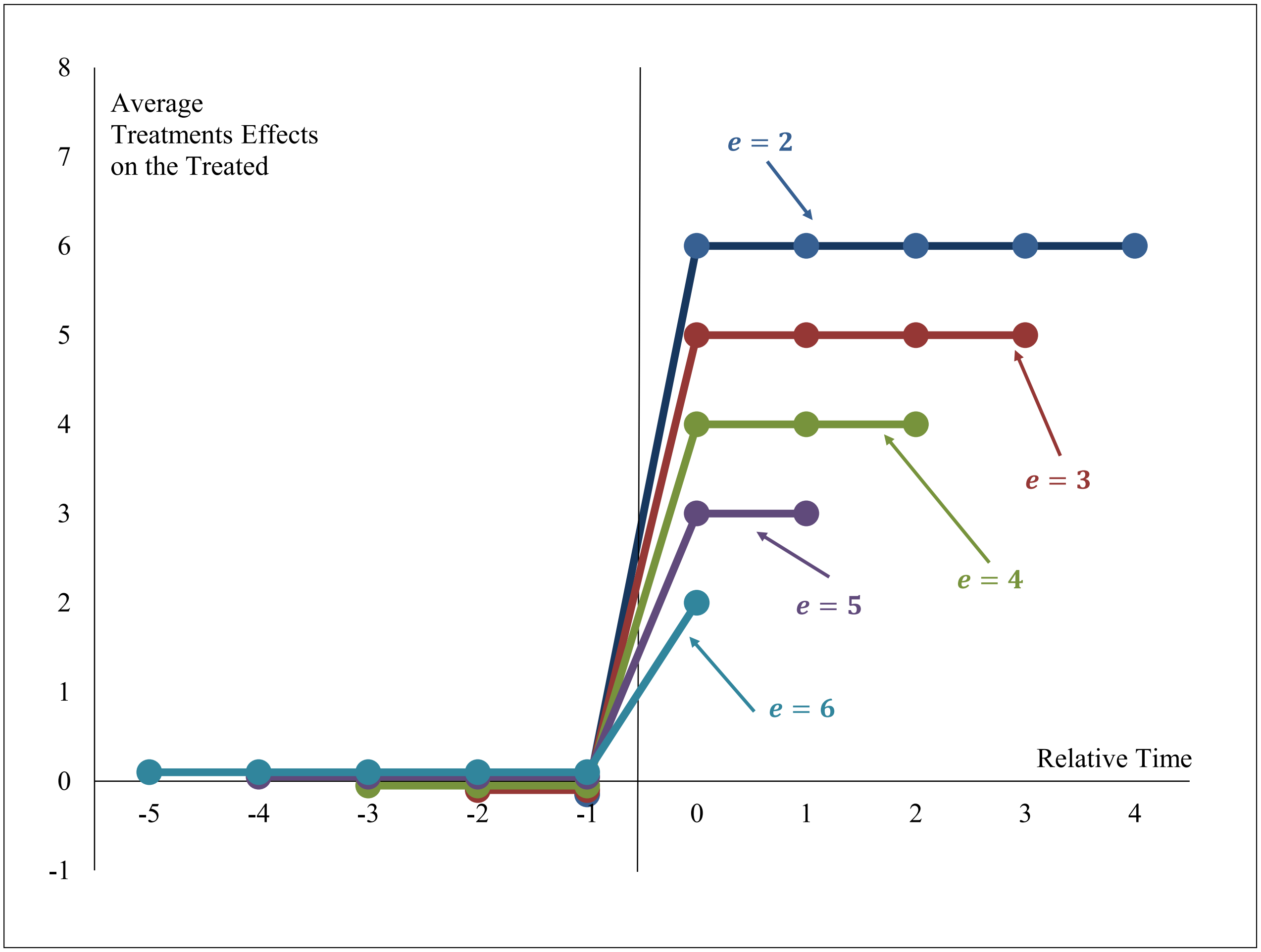

Fig. 4. This figure illustrates the average evolution of the ATT $\left(\beta_{et}^{ATT} : t \in \mathcal{T}\right)$ for each $e \in \mathcal{E}$ given static and non-stationary treatment effects. In this example, the ATTs decline with e. The pre-event parameters are set to be slightly different from zero so that each line in the figure can be distinguished in the pre-event period. See the discussion near page 13.

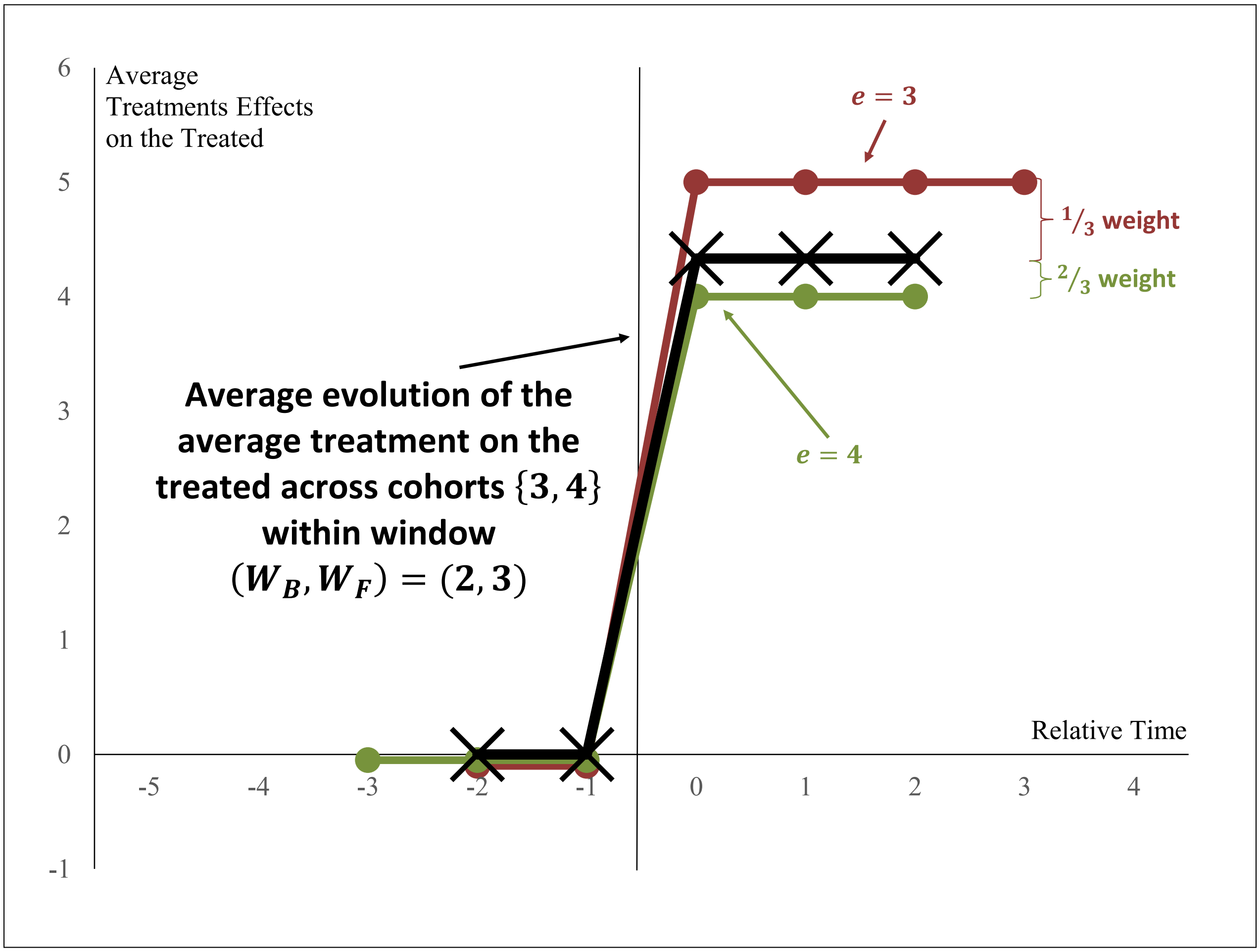

Fig. 5. This figure illustrates how $\left(\beta_\tau^{ATT}(2,3) : \tau \in \{-2, 2\}\right)$ may be constructed from $\{\beta_{et}^{ATT} : (e,t) \in \mathcal{T}^2 \text{ and } e \geq 2\}$ according to equation (6). The pre-event parameters are set to be slightly different from zero so that each line in the figure can be distinguished in the pre-event period. See the discussion near page 13.

ID	1	2	3	4	5	6	7	8	9	10	Prob. Mass
	Event Periods										
1.	-	-	-	-	-	-	-	-	-	-	0.010
2.	-	1	-	-	-	-	-	-	-	-	0.010
3.	-	-	1	-	-	-	-	-	-	-	0.010
4.	-	-	-	1	-	-	-	-	-	-	0.010
5.	-	-	-	-	1	-	-	-	-	-	0.010
6.	-	-	-	-	-	1	-	-	-	-	0.010
7.	-	-	-	-	-	-	1	-	-	-	0.010
8.	-	-	-	-	-	-	-	1	-	-	0.010
9.	-	-	-	-	-	-	-	-	1	-	0.010
10.	-	-	-	-	-	-	-	-	-	1	0.010
11.	-	1	-	1	-	-	-	-	-	-	0.200
12.	-	-	-	-	1	1	-	-	-	-	0.200
13.	-	-	-	-	-	-	1	-	-	1	0.200
14.	-	-	-	-	-	-	1	1	-	-	0.200
15.	-	-	-	1	-	-	1	1	-	-	0.020
16.	-	1	-	1	1	-	-	-	-	-	0.020
17.	-	-	1	-	1	1	-	-	-	-	0.020
18.	-	-	1	-	-	-	1	-	-	1	0.020
19.	-	1	1	-	-	-	1	-	-	1	0.020
										Probability mass with zero events	0.010
										Probability mass with one event	0.090
										Probability mass with multiple events	**0.900**

Fig. 6. This table presents the population distribution of H_i over Ω_H in the Monte Carlo simulations. Note that units that draw histories 11-19 experience multiple events. See the discussion near page 19.

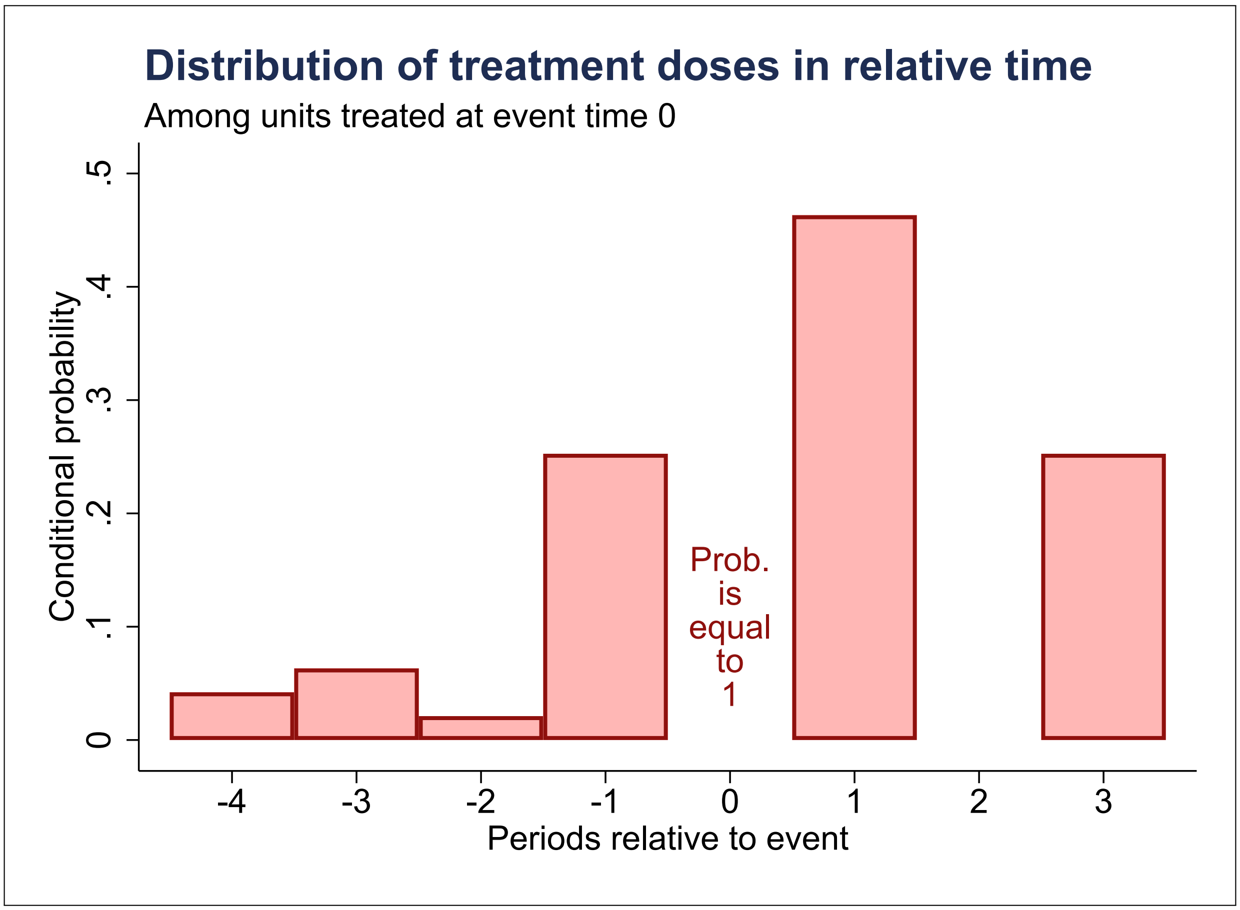

Fig. 7. This figure presents the distribution of experiencing an event in $\tau \in \{-4, ..., 3\}$ conditional on experiencing an event in $\tau = 0$ in the Monte Carlo simulations. See the discussion near page 19.

	Treatment Effect Function				
	Static	**Dynamic**			
	Stationary	**Stationary**		**Non-Stationary**	
Lag	**Hist. Ind.**	**Hist. Ind.**	**Hist. Dep.**	**Hist. Ind.**	**Hist. Dep.**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Matching estimator/weighted least squares estimator of stacked reg.				
-4	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00
-1	-	-	-	-	-
0	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00
	MDO estimator				
-4	0.00	0.03	0.68	0.00	0.34
-3	0.00	0.05	0.52	0.00	0.25
-2	0.00	0.01	1.16	0.08	0.01
-1	-	-	-	-	-
0	0.00	0.00	6.49	0.12	17.65
1	0.00	0.02	5.59	0.11	4.08
2	0.00	0.05	7.10	0.07	3.36
3	0.00	0.08	6.17	0.29	0.17
	Unweighted OLS estimator of stacked regression				
-4	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00
-1	-	-	-	-	-
0	0.00	0.00	1.49	0.42	5.20
1	0.00	0.00	1.48	0.67	2.02
2	0.00	0.00	1.48	0.62	1.93
3	0.00	0.00	1.48	0.21	0.01

Fig. 8. This table compares the mean-squared error of the matching estimator, the MDO estimator, and the OLS estimator of the stacked regression model in the Monte Carlo simulations. See the discussion near page 19.

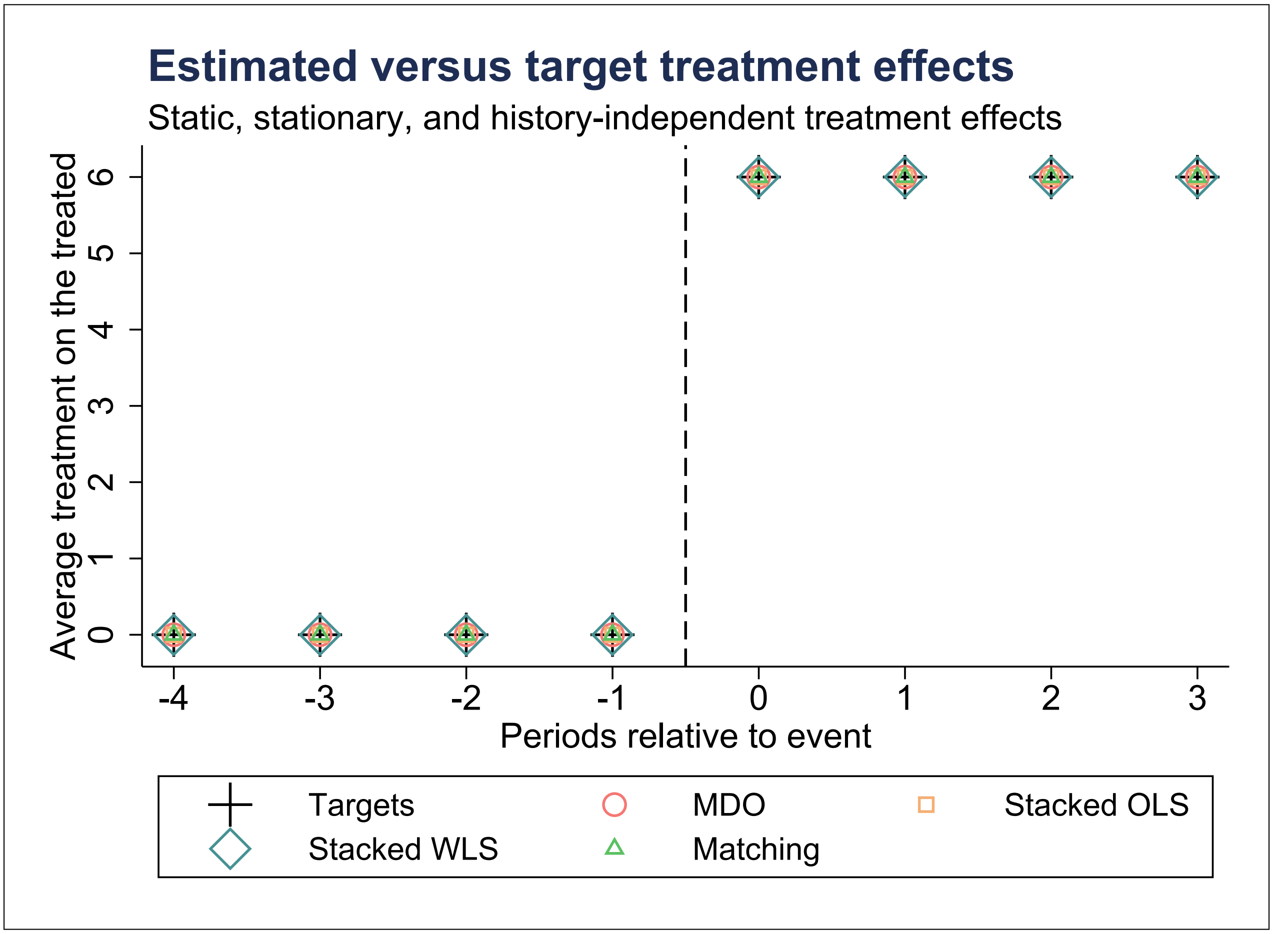

Fig. 9. This figure compares the performance of the MDO estimator, the OLS estimator of the stacked regression model, the weighted least squares estimator of the stacked regression model, and the matching estimator across 500 independent simulations for estimating $(\beta_\tau^{ATT}(4,4) : \tau \in \{-4, ..., 3\})$. See the discussion near page 19.

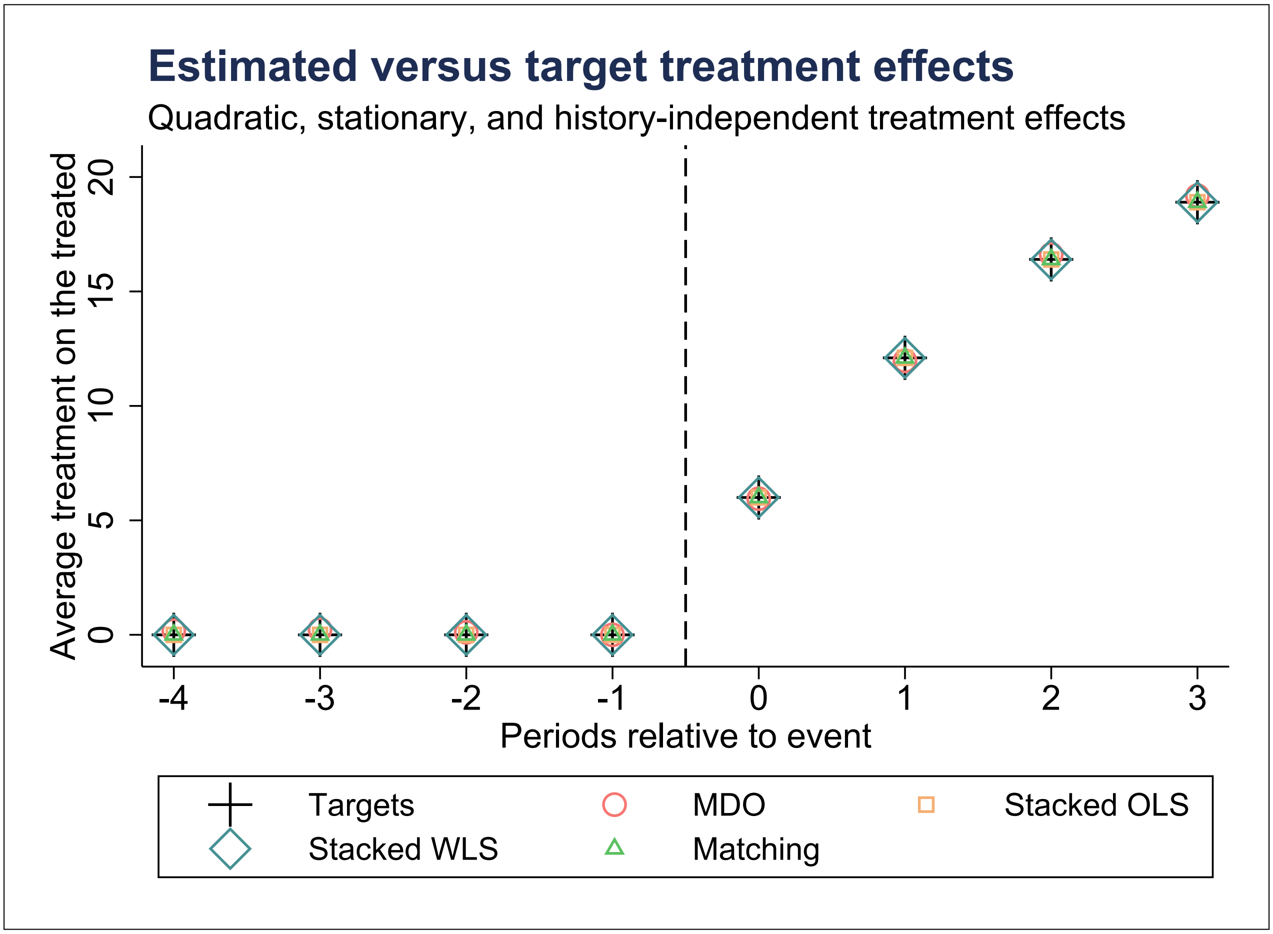

Fig. 10. This figure compares the performance of the MDO estimator, the OLS estimator of the stacked regression model, the weighted least squares estimator of the stacked regression model, and the matching estimator across 500 independent simulations for estimating $(\beta_\tau^{ATT}(4,4) : \tau \in \{-4, ..., 3\})$. See the discussion near page 19.

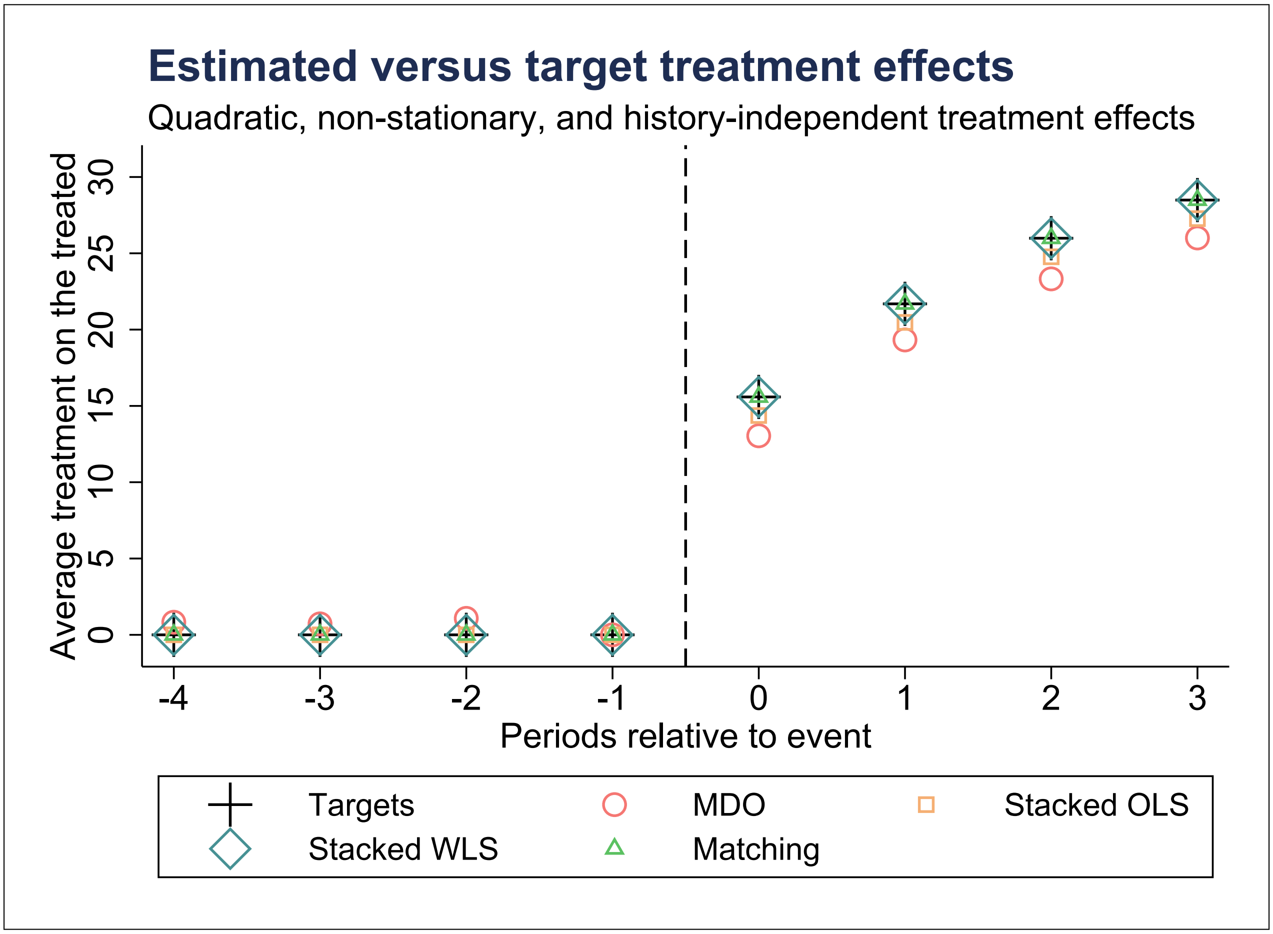

Fig. 11. This figure compares the performance of the MDO estimator, the OLS estimator of the stacked regression model, the weighted least squares estimator of the stacked regression model, and the matching estimator across 500 independent simulations for estimating $(\beta^{ATT}_{\tau}(4, 4) : \tau \in \{-4, ..., 3\})$. See the discussion near page 19.

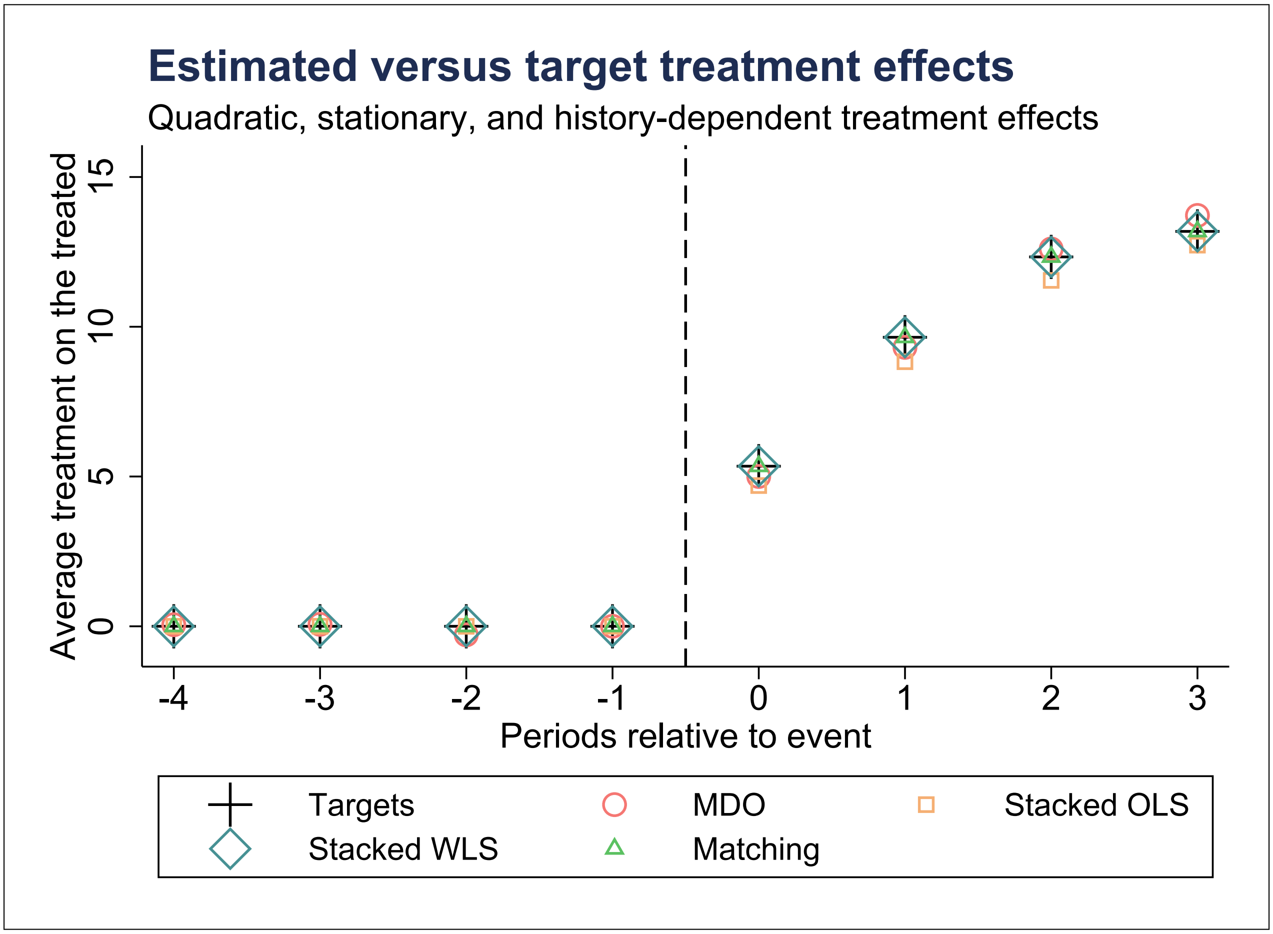

Fig. 12. This figure compares the performance of the MDO estimator, the OLS estimator of the stacked regression model, the weighted least squares estimator of the stacked regression model, and the matching estimator across 500 independent simulations for estimating $(\beta_\tau^{ATT}(4,4) : \tau \in \{-4, \ldots, 3\})$. See the discussion near page 19.

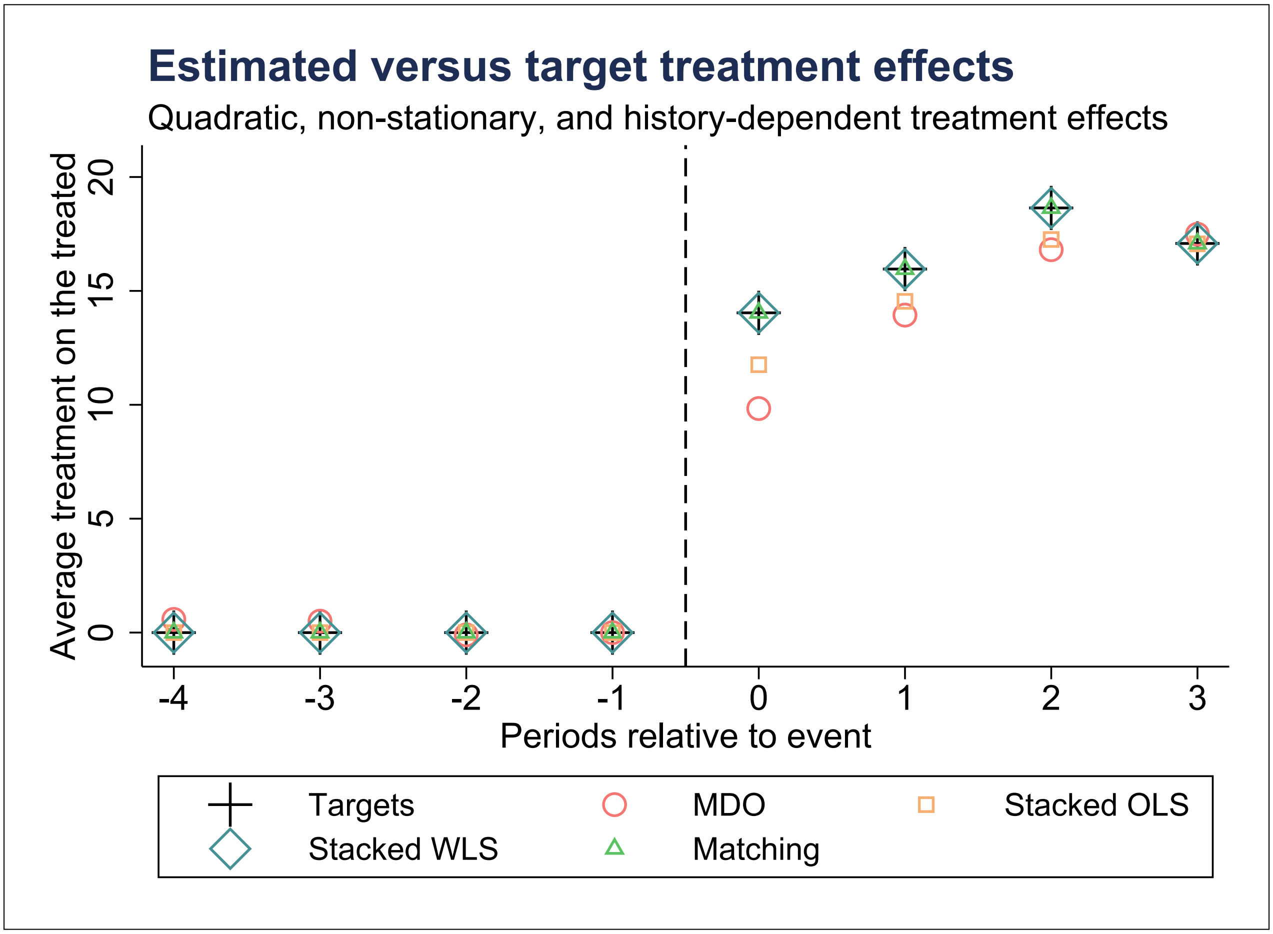

Fig. 13. This figure compares the performance of the MDO estimator, the OLS estimator of the stacked regression model, the weighted least squares estimator of the stacked regression model, and the matching estimator across 500 independent simulations for estimating $(\beta_\tau^{ATT}(4,4) : \tau \in \{-4, ..., 3\})$. See the discussion near page 19.

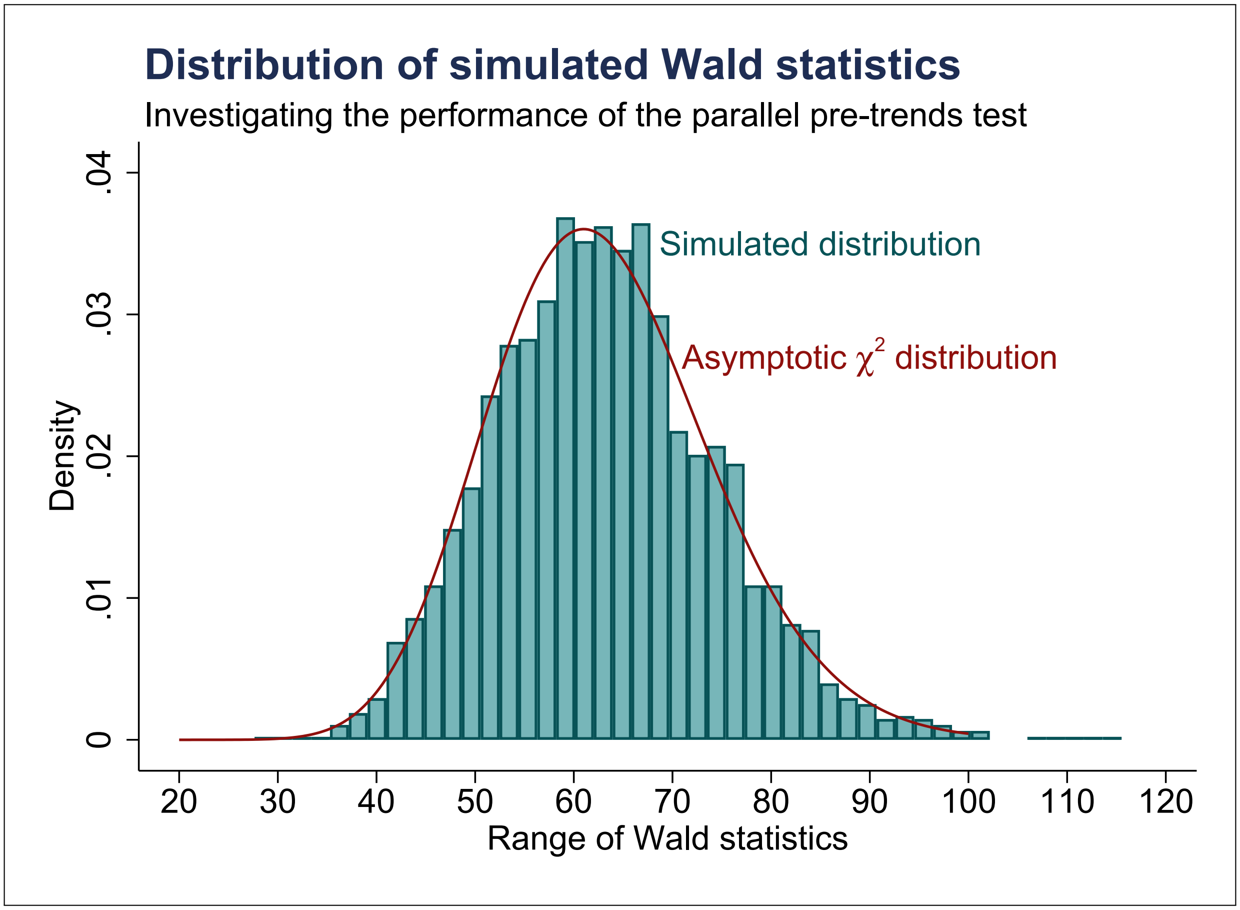

Fig. 14. The parallel pre-trends test introduced in section III is based on a Wald statistic that converges in distribution to a chi-squared random variable as $N \longrightarrow \infty$. The figure plots the distribution of Wald statistics across 500 independent simulations, showing that they are approximately distributed by χ^2_{63}, where $63 = \text{rank}(A')$ for the matrix A' described in section III. See the discussion near page 19.

VIII. Appendix

A. Proofs of Theorems 1 and 2

Theorems 1 and 2 state that under no anticipation and parallel trends assumptions, the average treatment effects on the treated $\beta_{et}^{ATT}(H^{-e})$ are equal to difference-in-differences given by equation (3) for all $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T}^e \times \Omega^*_{H^{-e}}$, and furthermore that $\beta_{et}^{ATT}(H^{-e}) = 0$ for all $t < e$.

Proof. Consider any $(e, t, H^{-e}) \in \mathcal{E} \times \mathcal{T}^e \times \Omega^*_{H^{-e}}$. Then,

$$
\begin{aligned}
\beta_{et}^{ATT}(H^{-e}) &:= \mathbb{E}[Y_{it}(H^{-e},1) - Y_{it}(H^{-e},0)|H_i = (H^{-e},1)] \\
&= \mathbb{E}[Y_{it}(H^{-e},1)|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it}(H^{-e},0)|H_i = (H^{-e},1)] \\
&= \mathbb{E}[Y_{it}(H^{-e},1)|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},1)] \\
&\quad - \mathbb{E}[Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},1)] \\
&= \mathbb{E}[Y_{it}(H^{-e},1)|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},0)] \\
&\quad - \mathbb{E}[Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},1)] \quad \text{by parallel trends} \\
&= \mathbb{E}[Y_{it}(H^{-e},1)|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},0)] \\
&\quad - \mathbb{E}[Y_{ie-1}(H^{-e},1)|H_i = (H^{-e},1)] \quad \text{by no anticipation} \\
&= \mathbb{E}[Y_{it}(H^{-e},1) - Y_{ie-1}(H^{-e},1)|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it}(H^{-e},0) - Y_{ie-1}(H^{-e},0)|H_i = (H^{-e},0)] \\
&= \mathbb{E}[Y_{it} - Y_{ie-1}|H_i = (H^{-e},1)] - \mathbb{E}[Y_{it} - Y_{ie-1}|H_i = (H^{-e},0)]
\end{aligned}
$$

Furthermore, if $t < e$, then

$$
\begin{aligned}
\beta_{et}^{ATT}(H^{-e}) &:= \mathbb{E}[Y_{it}(H^{-e},1) - Y_{it}(H^{-e},0)|H_i = (H^{-e},1)] \\
&= \mathbb{E}[Y_{it}(H^{-e},0) - Y_{it}(H^{-e},0)|H_i = (H^{-e},1)] \quad \text{by no anticipation} \\
&= 0
\end{aligned}
$$

B. Derivation of the Covariance of Any Two $\hat{\beta}^{ATT}_{et}(H^{-e})$

Consider a sample of N units drawn independently from a common population producing a dataset $\{(Y_i, H_i) : i \in \mathcal{I}\}$ with the characteristics described in section II. Consider any $(k_1, k_2) \in \mathcal{K}^2$ indexing some $\big((e_1, t_1, H_1^{-e_1}), (e_2, t_2, H_2^{-e_2})\big)$. Define $\mathcal{I}_{1,1} := \{i : H_i = (H_1^{-e_1}, 1)\}$, $\mathcal{I}_{1,0} := \{i : H_i = (H_1^{-e_1}, 0)\}$, $\mathcal{I}_{2,1} := \{i : H_i = (H_2^{-e_2}, 1)\}$, and $\mathcal{I}_{2,0} := \{i : H_i = (H_2^{-e_2}, 0)\}$. Further define $N_{m,n} := |\mathcal{I}_{m,n}|$. Then the covariance $\mathbb{C}\big[\hat{\beta}^{ATT}_{e_1,t_1}(H_1^{-e_1}), \hat{\beta}^{ATT}_{e_2,t_2}(H_2^{-e_2})\big]$ is given by

$$
\begin{aligned}
&\mathbb{C}\Bigg[N_{1,1}^{-1}\sum_{i\in\mathcal{I}_{1,1}}(Y_{it_1}-Y_{ie_1-1}) - N_{1,0}^{-1}\sum_{i\in\mathcal{I}_{1,0}}(Y_{it_1}-Y_{ie_1-1}), N_{2,1}^{-1}\sum_{i\in\mathcal{I}_{2,1}}(Y_{it_2}-Y_{ie_2-1}) - N_{2,0}^{-1}\sum_{i\in\mathcal{I}_{2,0}}(Y_{it_2}-Y_{ie_2-1})\Bigg]\\
&= N_{1,1}^{-1}N_{2,1}^{-1}\sum_{i\in\mathcal{I}_{1,1}}\sum_{j\in\mathcal{I}_{2,1}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{jt_2}-Y_{je_2-1}\big]\\
&\quad - N_{1,1}^{-1}N_{2,0}^{-1}\sum_{i\in\mathcal{I}_{1,1}}\sum_{j\in\mathcal{I}_{2,0}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{jt_2}-Y_{je_2-1}\big]\\
&\quad - N_{1,0}^{-1}N_{2,1}^{-1}\sum_{i\in\mathcal{I}_{1,0}}\sum_{j\in\mathcal{I}_{2,1}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{jt_2}-Y_{je_2-1}\big]\\
&\quad + N_{1,0}^{-1}N_{2,0}^{-1}\sum_{i\in\mathcal{I}_{1,0}}\sum_{j\in\mathcal{I}_{2,0}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{jt_2}-Y_{je_2-1}\big]\\
&= N_{1,1}^{-1}N_{2,1}^{-1}\sum_{i:i\in\mathcal{I}_{1,1}\wedge i\in\mathcal{I}_{2,1}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\big]\\
&\quad - N_{1,1}^{-1}N_{2,0}^{-1}\sum_{i:i\in\mathcal{I}_{1,1}\wedge i\in\mathcal{I}_{2,0}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\big]\\
&\quad - N_{1,0}^{-1}N_{2,1}^{-1}\sum_{i:i\in\mathcal{I}_{1,0}\wedge i\in\mathcal{I}_{2,1}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\big]\\
&\quad + N_{1,0}^{-1}N_{2,0}^{-1}\sum_{i:i\in\mathcal{I}_{1,0}\wedge i\in\mathcal{I}_{2,0}}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\big] \quad \text{by } Y_i \perp Y_j \quad \forall (i,j) \text{ s.t. } i \neq j\\
&= N_{1,1}^{-1}N_{2,1}^{-1}|\mathcal{I}_{1,1}\cap\mathcal{I}_{2,1}|\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\,\big|\, i\in\mathcal{I}_{1,1}\wedge i\in\mathcal{I}_{2,1}\big]\\
&\quad - N_{1,1}^{-1}N_{2,0}^{-1}|\mathcal{I}_{1,1}\cap\mathcal{I}_{2,0}|\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\,\big|\, i\in\mathcal{I}_{1,1}\wedge i\in\mathcal{I}_{2,0}\big]\\
&\quad - N_{1,0}^{-1}N_{2,1}^{-1}|\mathcal{I}_{1,0}\cap\mathcal{I}_{2,1}|\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\,\big|\, i\in\mathcal{I}_{1,0}\wedge i\in\mathcal{I}_{2,1}\big]\\
&\quad + N_{1,0}^{-1}N_{2,0}^{-1}|\mathcal{I}_{1,0}\cap\mathcal{I}_{2,0}|\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}\,\big|\, i\in\mathcal{I}_{1,0}\wedge i\in\mathcal{I}_{2,0}\big] \quad \text{by common pop.}\\
&= \begin{cases} 0 & \text{if } \mathcal{I}_{1,1}\cap\mathcal{I}_{2,1} = \emptyset\\ N_{k_1,1}^{-1}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}|H_i = (H_1^{-e_1},1)\big] & \text{if } \mathcal{I}_{1,1}\cap\mathcal{I}_{2,1} \neq \emptyset \end{cases}\\
&\quad - \begin{cases} 0 & \text{if } \mathcal{I}_{1,1}\cap\mathcal{I}_{2,0} = \emptyset\\ N_{k_1,1}^{-1}\mathbb{C}\big[Y_{it_1}-Y_{ie_1-1}, Y_{it_2}-Y_{ie_2-1}|H_i = (H_1^{-e_1},1)\big] & \text{if } \mathcal{I}_{1,1}\cap\mathcal{I}_{2,0} \neq \emptyset \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
&- \begin{cases} 0 & \text{if } \mathcal{I}_{1,0} \cap \mathcal{I}_{2,1} = \emptyset \\ N_{k_1,0}^{-1} \mathbb{C}\left[Y_{it_1} - Y_{ie_1-1}, Y_{it_2} - Y_{ie_2-1} | H_i = (H_1^{-e_1}, 0)\right] & \text{if } \mathcal{I}_{1,0} \cap \mathcal{I}_{2,1} \neq \emptyset \end{cases} \\
&+ \begin{cases} 0 & \text{if } \mathcal{I}_{1,0} \cap \mathcal{I}_{2,0} = \emptyset \\ N_{k_1,0}^{-1} \mathbb{C}\left[Y_{it_1} - Y_{ie_1-1}, Y_{it_2} - Y_{ie_2-1} | H_i = (H_1^{-e_1}, 0)\right] & \text{if } \mathcal{I}_{1,0} \cap \mathcal{I}_{2,0} \neq \emptyset \end{cases}
\end{aligned}
$$

where the last equality follows from the fact that if $\mathcal{I}_{1,1} \cap \mathcal{I}_{2,1} \neq \emptyset$, then $(H_1^{-e_1}, 1) = (H_2^{-e_2}, 1)$, and then $\{i : H_i = (H_1^{-e_1}, 1)\} = \{i : H_i = (H_2^{-e_2}, 1)\}$, and similar for the other three pairs $(\mathcal{I}_{1,1}, \mathcal{I}_{2,0})$, $(\mathcal{I}_{1,0}, \mathcal{I}_{2,1})$, and $\mathcal{I}_{1,0}, \mathcal{I}_{2,0})$.