Causal Inference in Financial Event Studies*

Paul Goldsmith-Pinkham
Yale University & NBER

Tianshu Lyu
Yale University

May 14, 2025

Abstract

We study financial event studies—empirical settings where asset market returns assess the impact of information or policy changes. We show that abnormal return estimators are sensitive to factor model misspecification, making them inconsistent estimators of causal effects. We demonstrate that staggered event timing can mitigate this issue in short-horizon studies but not in long-horizon analyses, where misspecification bias accumulates over time. Synthetic control methods emerge as a solution that avoids these issues by directly modeling counterfactual security paths without requiring correct specification of the factor structure. Our empirical applications to political connections (Acemoglu et al., 2016) and S&P 500 index inclusion (Greenwood & Sammon, 2025) demonstrate the practical implication of these methodological insights, particularly in settings that lack both short horizons and random event timing.

*Contact: paul.goldsmith-pinkham@yale.edu We thank Nick Barberis, Stefano Giglio and Will Goetzmann for helpful discussions.

1 Introduction

Financial event studies are among the most widely used empirical tools in financial economics, serving as a critical methodology for assessing how financial markets react to new information. Since the pioneering work of Fama et al. (1969), researchers have employed event studies to examine the economic impact of corporate announcements, regulatory changes, and macroeconomic news across a variety of settings. The central premise of these studies is straightforward: if markets efficiently incorporate information, asset prices should adjust to reflect the economic consequences of an event. By measuring abnormal returns around event dates, researchers can quantify the market's assessment of an event's effect on firm value.

Despite their widespread use and intuitive appeal, the identification assumptions underlying financial event studies have rarely been examined through the lens of modern causal inference. This gap is particularly concerning given that similar empirical methodologies in other fields—such as difference-in-differences designs—have undergone substantial methodological scrutiny in recent years (e.g., De Chaisemartin and d'Haultfoeuille (2020), Sun and Abraham (2021), and Goodman-Bacon (2021)). Moreover, the econometric framework typically used in financial event studies relies on correctly specified factor models to generate counterfactual returns, yet little attention has been paid to how factor model misspecification influences causal estimates.

In this paper, we bridge this gap by developing a potential outcomes framework for financial event studies and clarifying the conditions under which these studies identify causal effects. Our approach accommodates both short-horizon and long-horizon event windows, allowing us to analyze identification challenges that arise when studying immediate market reactions versus extended price responses. This distinction is crucial, as many influential studies in the literature examine long-horizon returns to investigate phenomena such as post-earnings announcement drift (Bernard & Thomas, 1989), long-run merger performance (Savor & Lu, 2009), and gradual market responses to other corporate events (Kwon & Tang, 2022).

Our analysis yields several important insights. First, we demonstrate that abnormal return estimators in financial event studies can be understood as attempting to estimate the average treatment effect on the treated (ATT) by constructing counterfactual returns for treated securities. The typical approach—using a factor model estimated during a pre-event window to predict counterfactual returns during the event window—relies on both limited anticipation of the event and correct specification of the factor structure. When these conditions fail, abnormal returns may provide misleading estimates of the true causal effect, especially when there is only one time period that the event occurs.

Second, we show that the bias in abnormal return estimators due to factor model misspecification is generally negligible in short-horizon studies with multiple event timings, but the bias can accumulate substantially in long-horizon analyses. This result provides formal justification for the common intuition among empirical researchers that model specification matters little for studies of immediate price reactions but becomes critically important for extended windows. The key insight is that while expected returns from factor exposures are typically small over short horizons (e.g., daily or weekly), they compound over longer horizons and can generate substantial bias if the factor structure is misspecified.

Third, we clarify how the assignment of events across securities and time affects identification. While randomized event timing can help mitigate concerns about factor-related bias in short-horizon studies, it does not guarantee identification in long-horizon analyses. We show that even with random timing, misspecified factor models can produce substantial bias in long-horizon studies if omitted factors exhibit drift or risk premia.

Fourth, we demonstrate that synthetic control methods (Abadie et al., 2010; Abadie & Cattaneo, 2021; Xu, 2017) offer a natural solution to the identification challenges in financial event studies. By directly modeling the counterfactual security path without requiring correct specification of the factor structure, synthetic control approaches can significantly reduce bias in both short-horizon with single events and long-horizon settings. This application represents a novel contribution to the financial event studies literature, where synthetic control methods have been underutilized despite their natural fit for the empirical task.

We illustrate our theoretical results through two empirical applications. First, we revisit Acemoglu et al. (2016), who study the role of political connections on stock returns following Timothy Geithner's nomination as Treasury Secretary. We show that the standard abnormal return approach may overstate the effect of political connections, as the event occurred during a period of market turbulence when factor model misspecification is particularly consequential. Using synthetic control methods, we show that the results from the paper disappear. This highlights the importance of abnormal return misspecification with single events.

Second, we examine S&P 500 index inclusion effects over several decades, following Greenwood and Sammon (2025). We find that, consistent with our analytic results, short-horizon price reactions are relatively robust to estimation method. However, we find that long-horizon pre-announcement drifts—often interpreted as evidence of anticipation or information leakage—are partially attributable to factor model misspecification. By implementing synthetic control approaches and propensity score matching based on firm characteristics, we can account for approximately half of the pre-announcement drift, suggesting that previous studies may have overstated the magnitude of anticipatory price movements.

Our findings have important implications for both the interpretation of existing financial event studies and the design of future research. Many influential findings in the literature rely on long-horizon abnormal returns to study phenomena such as post-announcement drift, merger performance, and gradual information diffusion. Our results suggest that some of these findings may reflect factor model misspecification rather than genuine causal effects. At the same time, we provide a constructive way forward through synthetic control methods, which can significantly reduce bias while maintaining the intuitive appeal of the event study approach.

2 A simple example

In this section, we illustrate how the assignment of events in a financial event study framework affects the bias of estimators, using a stylized two factor case.

2.1 Model with Two Factors

Let $i = 1, \ldots, N$ index securities and $t = 1, \ldots, T$ index discrete time. Each security i may or may not experience an event at time T_i (possibly varying by i). If a security never experiences a return, we denote $T_i = \infty$. For this section, we assume the following "true" model for returns:

$$R_{i,t} \ = \ \alpha_i \ + \ \beta_{i,1} \, F_{1,t} \ + \ \beta_{i,2} \, F_{2,t} \ + \ \tau_{t-T_i} D_i \ + \ \varepsilon_{i,t}, \tag{1}$$

where:

- $F_{1,t}$ and $F_{2,t}$ are two risk factors
- $\beta_{i,1}, \beta_{i,2}$ are constant factor loadings for security i,
- $D_i = 1(T_i \neq \infty)$, whether the event has occurred for security i at time t,
- τ_{t-T_i} is the effect of the event $t - T_i$ periods from the event,
- $\varepsilon_{i,t}$ is a mean-zero idiosyncratic error.

As we discuss further in Section 3.1, we will assume that $\tau_\kappa = 0$ for $\kappa < 0$ or some larger lag. In other words, the event has no impact on prices before it occurs. This assumption is necessary for the abnormal returns assumption we will discuss below.

2.2 The abnormal returns estimator using one factor

We know that the asset has a true model of two factors, but imagine a researcher studying this asset only include the first factor $F_{1,t}$ in its "benchmark" model when estimating abnormal returns, following Campbell et al. (1997), Brown and Warner (1985), and Kothari and Warner

(2007) or other guidance. This may because the second factor is unobserved/unmeasured, or for reasons of parsimony. Let

$$\widehat{R}^{(0)}_{i,t} \;=\; \widehat{\alpha}_i \;+\; \widehat{\beta}_{i,1}\, F_{1,t}$$

be the predicted return under no treatment (estimated from pre-event data). Due to the omitted factor,

$$\hat{\beta}_{i,1} \rightarrow_p \beta_{i,1} + \beta_{i,2}\frac{Cov(F_{2,t}, F_{1,t})}{Var(F_{1,t})}. \tag{2}$$

The *abnormal return* on the day of the event for a single firm treated in period s is

$$AR_{i,0} \;=\; R_{i,s} \;-\; \widehat{R}^{(0)}_{i,s} \;\approx\; \tau_0 \;+\; \beta_{i,2}\left(F_{2,s} - \frac{Cov(F_{2,t}, F_{1,t})}{Var(F_{1,t})}F_{1,s}\right) + \varepsilon_{it} \tag{3}$$

As a result, our estimator contains three terms: the firm treatment effect on the day of the event, the omitted variable bias from the second factor, adjusted for the factor correlation, and the idiosyncratic noise for the firm in period s. Note that if $F_{2,t}$ and $F_{1,t}$ are uncorrelated or independent, the omitted variable bias term simplifies to just $\beta_{i,2}F_{2,t}$. Regardless, because $F_{2,t}$ is omitted from the benchmark, if it is non-zero on average *after* the event, the difference $R_{i,t} - \widehat{R}^{(0)}_{i,t}$ will be biased relative to the true event effect τ_0.

Often, we will observe many treated firms at a single time period s, such that we consider the average abnormal return:

$$\overline{AR}_{s,0} = n_s^{-1}\sum_{i:T_i=s} AR_{i,0} \;\approx\; \tau_0 \;+\; \underbrace{\beta_{s,2}\left(F_{2,s} - \frac{Cov(F_{2,t}, F_{1,t})}{Var(F_{1,t})}F_{1,s}\right)}_{\text{asymptotic factor bias}} + \;\; \underbrace{n_s^{-1}\sum_{i:T_i=s}\varepsilon_{is}}_{\text{goes to zero with } n_s} \tag{4}$$

where $\beta_{s,2} = n_s^{-1}\sum_{i:T_i=s}\beta_{i,2}$ is the factor loadings for the average portfolio made up of securities experiencing the event in period s. While averaging across these many units helps average out the idiosyncratic error as $n_s \to \infty$, it has no impact on the omitted factor.

Intuitively, this result highlights that have many *cross-sectional units* does not help alleviate common shocks. Econometrically, this estimator is inconsistent, since it converges to a random variable – the value will vary as a function of the omitted factor times the factor loading for the firms experiencing the event (Andrews, 2005).

Of course, intuitively, in many applications the factor loadings are often not too large, and the underlying risk premia are, on average, typically small relative to τ_0. For example, the one-day index inclusion effect is estimated to be somewhere between 1-4%, depending on the time period. By comparison, the market return is, on average, 0.05%, two orders of magnitude smaller than the treatment effects. However, there are many periods when the market return can be far larger, such as during periods of market volatility. As a result, this bias can be quite large. We will demonstrate this issue in simulations in Section 5 and empirically in Section 6.1.

One natural question is whether averaging over *many* events can solve this issue, analogous to a staggered difference-in-difference design. As we show in Section 3, the misspecified abnormal returns estimator will still be biased, even when the factor premia F_t are as-if randomly assigned:

$$\overline{AR}_{\kappa=0} \approx \tau_0 \ + \ E(\beta_{i,2}|T_i = s)\left(E(F_{2,s}) - \frac{Cov(F_{2,t}, F_{1,t})}{Var(F_{1,t})} E(F_{1,s}) \right). \tag{5}$$

Again, this abnormal returns estimator has a bias term hanging around, but with enough events, this will average to a constant. This bias is the average effect of the omitted risk factor on the returns of a portfolio of firms experiencing the event.

The astute finance reader will note that this bias term is typically *quite* small in practice, since the average factor premium at the daily level is typically quite small relative to overall stock movements. Hence, while the abnormal return estimator is biased, the bias is nearly negligible for a short-horizon. For example, imagine the the factors were uncorrelated, and $E(F_{2,s})$ was 0.02 percent. Then, for a factor loading of 1.5, this bias would be 0.03 percent, which is quite small relative to many abnormal returns reported in the literature. However, if the factors are not as-if randomly drawn (e.g. the treated firms on a given day's factor loadings are correlated with the omitted factor on that day), then this bias could be larger.

An implication of this small bias is that the abnormal returns estimator is often quite close to the true treatment effect, even when the factor model is misspecified. Moreover, this bias could be small even for a model that ignores both factors, consistent with the simulation evidence in Brown and Warner (1985) that the form of the abnormal return estimator has limited effects on the estimates.[1] Of course, this small bias assumption crucially relies on the idea that the effect size, τ_0, is large. The presence of this bias will make smaller effects harder to detect.

However, this small bias can be larger for longer-horizon event studies, as we will discuss in Section 4.2. As we cumulate the abnormal returns over time, the bias term will accumulate, potentially leading to substantial bias in the estimated treatment effect. Intuitively, the bias documented above will be scaled by the length of the period: if the factor premium is 0.02 percent per day, then the bias will be proportional to 0.02 percent for a one-day event, 0.04 percent for a two-day event, and so on. For a 250 day period, the magnitude of the bias will

[1]The simulations in Brown and Warner (1985) are such that the event days are exactly randomly assigned across time: "Each time a security is selected, a hypothetical event day is generated. Events are selected with replacement, and are assumed to occur with equal probability on each trading day from July 2, 1962, through December 31, 1979."

be scaled by 5 percent which is quite large relative to many treatment effects. It could be negative if the factor loading is negative, or positive – the sign is ambiguous.

In all of the previous results, the inconsistency and bias of the estimator comes from the misspecified returns model. A challenge faced by many empirical practicioners is that the set of relevant factors is unknown or unobserved. In the next section, we will discuss how synthetic control methods can help alleviate this issue by directly modeling the counterfactual security path without requiring correct specification of the factor structure. We additionally discuss how random assignment of the event can help alleviate these issues.

3 Identification

This section formalizes the setup of our event study in the language of potential outcomes. We begin by introducing the basic notation (Section 3.1), defining potential returns and treatment indicators for each security over time. We then specify the causal estimands of interest (Section 3.2), clarifying what it means to identify a treatment effect in an event study context. Finally, we discuss how these causal quantities relate to traditional event study methods based on "abnormal returns" and factor model adjustments.

3.1 Setup

Let $i = 1, 2, \ldots, N$ index securities (e.g., stocks), and let t index discrete time periods (e.g., daily or monthly observations), $t = 1, 2, \ldots, T$. Let $D_{i,t}$ be a binary variable that identicates whether an event has occured. We assume the event is irreversible:

Assumption 1 (Irreversibility of Treatment). $D_{i,1} = 0, \quad D_{i,t} = 1 \quad \text{if } D_{i,t-1} = 1 \forall i.$

This assumption is in line with many financial event studies, where an event (e.g., an earnings announcement, merger, or policy change) occurs at a single point in time and does not reverse. It is also implied by Equation (1), and is a common assumption in many difference-in-difference settings (Callaway & Sant'Anna, 2021).

Let the treatment timing T_i denote when the event occurs for security i,

$$T_i = \begin{cases} t & \text{if security } i \text{ is treated (event occurs) at time } t, \\ \infty & \text{if never treated.} \end{cases} \tag{6}$$

In many empirical settings, there is a single event date t_0 (e.g., an earnings announcement, merger announcement, or policy change) for all securities, so $T_i = t_0$ (and ∞ otherwise), which is isomorphic to writing $T_i \in \{0, 1\}$. Let $\mathcal{C}$ denote the set of securities where $T_i = \infty$, and the set of possible event timings $\mathcal{S} \subseteq \{1, \ldots, T\}$.

Now, we setup the potential outcomes framework for our returns. Let $R_{i,t}(s)$ be the potential return for security i at time t if it has the event occur in period s, and $R_{i,t}(\infty)$ the potential return in the absence of any event. Because a security cannot be both treated and untreated, we only observe one of the potential returns for each (i, t):

$$R_{i,t} = R_{i,t}(\infty) + \sum_{s=2}^{T} (Y_{i,t}(s) - Y_{i,t}(\infty)) 1(T_i = s). \tag{7}$$

Next, we turn to a crucial modeling assumption about the structure of expected returns, based on a long literature in asset pricing (Chamberlain & Rothschild, 1983; Connor, 1984).

Assumption 2 (Linear Factor Model of Expected Returns)**.** *We assume that in the absence of the event, each security i's return follows a linear factor model with intercept α_i, K time varying factors $\mathbf{F}_t$ and factor weights β_i, such that*

$$E\left(R_{it}(\infty) \mid T_i = s, \mathbf{F}_t\right) = \alpha_s + \beta_s \mathbf{F}_t, \tag{8}$$

where $\alpha_s = E(\alpha_i | T_i = s), \beta_s = E(\beta_i | T_i = s)$.

Note that this assumption is quite strong. For example, it does not allow for changing factor loadings (Barberis et al., 2005). It also does not allow for the market to *anticipate* an event (rationally) in the future if the event does not eventually occur. In future work, we will relax this assumption (this issue is considered in a series of papers in the finance literature, e.g. Prabhala (1997), that consider conditional events).

There may be cases where assignment of T_i in time and across units can be useful for identification. We will consider two types of assignment mechanisms: random timing and random assignment. Let $\boldsymbol{X}_i = (\alpha_i, \beta_i)$ be the vector of security-specific parameters. Let $\boldsymbol{F} = (F_1, \ldots, F_T)$ be the vector of factor realizations.

Definition 1. Let $p_t(\boldsymbol{X}_i, \boldsymbol{F}) = \Pr(T_i = t | \boldsymbol{X}_i, \boldsymbol{F})$ be the probability that firm i experiences the event at time t, given its characteristics X_i and the vector of factor realizations. We refer to this as the *timing propensity score.*

We can consider special versions of the timing propensity score:

Assumption 3 (Random Assignment)**.** $p_t(\boldsymbol{X}_i, \boldsymbol{F}) = p_t(\boldsymbol{F})$*, implying that which firms experience events is independent of their characteristics $\boldsymbol{X}_i$. This is analagous to a randomized intervention.*

Assumption 4 (Random Timing)**.** ***Random Timing:*** $p_t(\boldsymbol{X}_i, \boldsymbol{F}) = p_t(\boldsymbol{X}_i)$*, implying that event timing is unrelated to market factors $\boldsymbol{F}$.*

We will not require Assumption 3 for identification, but it will be useful for understanding the properties of our estimators. Moreover, we will show (unsurprisingly) that random assignment implies that simple difference-in-means estimators are unbiased relative to the a (misspecified) abnormal return estimator.

3.2 Average Treatment Effect Estimands for Event Studies

We are interested in identifying causal effects of the event (treatment) on returns. Commonly, researchers focus on the *average treatment effect on the treated* (ATT) over some event window. We will build these treatment effect parameters from the building block of individual treatment effects:

$$\tau_i(s,t) = R_{i,t}(s) - R_{i,t}(\infty), \tag{9}$$

the difference between the treated and untreated returns in period s for firm i that is treated in period t. Then, we can define an ATT:[2]

$$\tau(s,t)^{ATT} = E(\tau_i(s,t) \mid T_i = s) = E(R_{i,t}(s) \;-\; R_{i,t}(\infty) \mid T_i = s). \tag{10}$$

Often, researchers will combine these ATTs to study effects relative to an event time. This can be written as:

$$\theta_\kappa = \sum_{s\in\mathcal{S}} w_s \tau(s, s+\kappa)^{ATT}, \tag{11}$$

where w_s denotes the relative weight put on each event timing. The weights w_s can be chosen based on the propensity score structure. Under random timing, a natural choice is $w_s = \frac{N_s}{\sum_{s'\in\mathcal{S}} N_{s'}}$, where N_s is the number of firms with $T_i = s$. This corresponds to the approach in Callaway and Sant'Anna, 2021, where weights are based on the relative density of each treatment timing group. This combines the different event timings.

Generally, there may be a vector of treatment effects that a researcher is interested in, such as the ATT in each period from the event to H periods after, or some functional transformation, such as the *cumulative* treatment effect over the period (which is analogous to the cumulative abnormal return commonly studied in the literature): $\theta_H^{CATT} = \sum_{\kappa=0}^{H} \theta_\kappa$.

Generally, the ATE will be unidentified in our setting, but we can identify the ATT. Note that the ATT requires us to estimate the counterfactual outcome return for the treated group:

$$\tau(s,t)^{ATT} \;=\; \underbrace{E(R_{i,\,T_i+\kappa}|T_i \neq \infty)}_{\text{observed}} \;-\; \underbrace{E(R_{i,\,T_i+\kappa}(0)|T_i \neq \infty)}_{\text{counterfactual}}. \tag{12}$$

We now turn to the relevant assumptions for identifying these treatment effects.

[2]Note we can define equivalent ATEs, but they are not identified, and hence we ignore them.

3.3 Identifying Assumptions

In addition to our model of factor returns, we will need to make an assumption on when the events can begin to have an effect.

Assumption 5 (Limited Anticipation)**.** *For all securities i and all time periods $t < T_i - \delta$, the potential returns are unaffected by the future treatment:*

$$R_{i,t}(T_i) = R_{i,t}(\infty) \quad \forall t < T_i - \delta \tag{13}$$

Remark 1. Since we can write

$$\begin{aligned} E(R_{i,t}(T_i) \mid T_i = s, \boldsymbol{F}_t) &= E(R_{i,t}(\infty) \mid T_i = s, \boldsymbol{F}_t) + \tau(t,s)^{ATT} & (14) \\ &= \alpha_s + \beta_s \boldsymbol{F}_t + \tau(s,t)^{ATT}, & (15) \end{aligned}$$

Assumption 5 implies that $\tau(s,t)^{ATT} = 0$ for all $t < T_i - \delta$. This means that the event has no impact on the returns of the treated group prior to the event.

This assumption implies that the treated group cannot have an impact from the announcement prior to the date of the release. There is obvious evidence in the finance literature of hidden information leaking out, with prices responding beforehand (e.g. Schwert (1996)). Indeed, this is often pointed to evidence for the strong version of the efficient markets hypothesis. Hence, limited anticipation will be necessary to set a benchmark for when leakage has not yet occurred. This will allow the researcher to identify the periods in which we can estimate the counterfactual returns. This is the assumption necessary to use the pre-event estimation window commonly used in financial event studies (Campbell et al., 1997; Kothari & Warner, 2007).

However, it is important to distinguish between selection into the treatment (e.g. $\{R_{it}(s)\}_{s\in\mathcal{S}}$ being correlated T_i) and anticipation of the treatment. The former is quite plausible, as we see in our analysis of the S&P 500 index inclusion effect in Section 6.2 – firms that are growing and having a large market cap are more likely to be selected into the S&P. The latter will bias our estimates of the true treatment effect, and can be caused by market participants anticipating the event. We will discuss this further in future work.

3.4 Estimators

Now consider the following estimators. First, consider the canonical abnormal returns model used in finance research (Campbell et al., 1997; Brown & Warner, 1985). We define an observed set of $\boldsymbol{F}^o_t$, and then estimate (using OLS), $(\hat{\alpha}_i, \hat{\beta}_i)$ using data prior to $T_i - \delta$:

$$R_{it} = \alpha_i + \beta_i \boldsymbol{F}_t^o + \varepsilon_{it},\ t < T_i - \delta. \tag{16}$$

Hence, $\hat{\alpha}_i, \hat{\beta}_i$ are the squared minimizers of a linear model for a stock's returns, using observed factors (which may include no factors, or a single factor, or many).

Definition 2 (Abnormal returns estimator)**.** Define the predicted value of R_{it} for a given $\boldsymbol{F}_t^o$ and estimated $\hat{\alpha}, \hat{\beta}$ as $\hat{R}_{it} = \hat{\alpha}_i + \hat{\beta}_i \boldsymbol{F}_t^o$. Hence,

$$AR_{it} = R_{it} - \hat{R}_{it} \tag{17}$$

and

$$\tau^{AR}(s,t) = E(AR_{i,t}|T_i = s) = E(R_{i,t}|T_i = s) - E(\hat{R}_{it}|T_i = s) \tag{18}$$

We next consider two alternative methods to estimation. The first is a simple difference-in-means estimator:

Definition 3 (Difference-in-means estimator)**.** The difference-in-means estimator is defined as:

$$\hat{\tau}^{cont}(s,t) = E(R_{i,t}|T_i = s) - E(R_{i,t}|i \in \mathcal{C}) \tag{19}$$

$$\hat{\theta}_\kappa^{cont} = \sum_{s \in \mathcal{S}} w_s \hat{\tau}^{cont}(s, s+\kappa). \tag{20}$$

This estimator is the difference between the average return of the treated group and the average return of the untreated group. If the control group includes all equities, and is weighted by market cap, then this estimator is equivalent to the "market-adjusted-return model" (Campbell et al., 1997; Brown & Warner, 1985).

Second, we consider a synthetic control estimator (Abadie & Cattaneo, 2021) that uses the pre-event data to construct a synthetic control group:

Definition 4 (Synthetic control estimator)**.** Let $\overline{R}_{s,t} = E(R_{it}|T_i = s)$ be the average return of the treated group at time t for event timing s. The synthetic control estimator is defined as:

$$\hat{\tau}^{synth}(s,t) = \overline{R}_{s,t} - \sum_{j:j \in \mathcal{C}} \hat{\omega}_j R_{j,t} \hat{\theta}_\kappa^{synth} \qquad = \sum_{s \in \mathcal{S}} w_s \hat{\tau}^{synth}(s, s+\kappa). \tag{21}$$

where $\hat{\omega}_j$ is chosen using a synthetic control estimator to minimize

$$\hat{\omega} = \arg\min_{\omega} \sum_{t < s-\delta} (\overline{R}_{s,t} - \sum_{j:j \in \mathcal{C}} \omega_j R_{j,t})^2. \tag{22}$$

Note that for our estimator, we focus on using the portfolio of securities treated in period s as a single unit ($\overline{R}_{s,t}$) rather than the individual returns. This approach is debated within the synthetic control literature, but in the financial setting, studying a portfolio of firms in this way is quite natural.

We will need to make the assumption that there exists a portfolio from the set of control securities that can exactly replicate the returns of the treated group. This is a strong assumption, but it is often made in the synthetic control literature. These weights, however, do not need to be known.

Assumption 6. *There exists a set of weights $\{\omega_s^*\}_{j\in\mathcal{C}}$ for each event period such that*

$$\overline{R}_{s,t} = \sum_{j:j\in\mathcal{C}} \omega_{s,j}^* R_{j,t} \quad \forall t < s - \delta. \tag{23}$$

In practice, this assumes that there are a set of stocks who construct an exact replicating portfolio for the set of treated of stocks. This is quite unlikely if there are only a few stocks in the treated stocks. However, if there are many treated firms, this is more likely to hold. In future work, we can also consider the case where we allow for imperfect fit, as in Abadie and L'hour (2021) and Ben-Michael et al. (2021, 2022).

It is also important to note that in many synthetic control approaches, the weights are required to be non-negative. This is done to ensure that the synthetic control is a convex combination of the control units. This is a restrictive assumption in finance – it would disallow the opportunity to short stocks, and significantly lower the likelihood of generating a matching portfolio and spanning the set of risk factors. Hence, this assumption is not necessary for our estimator, and we will not impose it. In fact, many other synthetic style methods (such as matrix completion methods) should work well in this setting as well. We defer this to future work.[3]

We can now consider our first set of results.

Proposition 1. *Let Assumptions 1, 2 and 5 hold. Then,*

1. *In finite samples, the estimators differ from their target estimands as a function of how*

[3] In Baker, Gelbach, et al. (2020), the authors explore the benefits of these methods for single firm event studies and find that synthetic control and matrix completion can do better than abnormal returns using Monte Carlo simulations. However, they do not explore the analytic properties of these methods for identifying the ATT.

closely the constructed counterfactual portfolio matches the treated portfolio:

$$\tau^{AR}(s,t) - \tau(s,t)^{ATT} = (\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_t - \hat{\beta}_s \boldsymbol{F}_t^o) \tag{24}$$

$$\hat{\tau}^{cont}(s,t) - \tau(s,t)^{ATT} = (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\boldsymbol{F}_t \tag{25}$$

$$\hat{\tau}^{synth}(s,t) - \tau(s,t)^{ATT} = (\alpha_s - \hat{\alpha}_s^{synth}) + (\beta_s - \hat{\beta}_s^{synth})\boldsymbol{F}_t, \tag{26}$$

where $\hat{\alpha}_s = E(\hat{\alpha}_i|T_i = s)$, $\hat{\beta}_s = E(\hat{\beta}_i|T_i = s)$, $\alpha_s = E(\alpha_i|T_i = s)$, $\beta_s = E(\beta_i|T_i = s)$, $\alpha_\infty = E(\alpha_i|T_i = \infty)$, $\beta_\infty = E(\beta_i|T_i = \infty)$. $\hat{\beta}_s^{synth} = \sum_{j:j\in\mathcal{C}} \hat{\omega}_j \beta_j$, *and* $\hat{\alpha}_s^{synth} = \sum_{j:j\in\mathcal{C}} \hat{\omega}_j \alpha_j$.

2. *If* $n_s, n_c, T \to \infty$, *then asymptotically, the synthetic control estimator is unbiased, but the other two estimators are inconsistent and converge to a random variable that is a function of the factor realizations:*

$$\tau^{AR}(s,t) - \tau(s,t)^{ATT} \to_p (\alpha_s - \tilde{\alpha}_s) + (\beta_s \boldsymbol{F}_t - \tilde{\beta}_s \boldsymbol{F}_t^o) \tag{27}$$

$$\hat{\tau}^{cont}(s,t) - \tau(s,t)^{ATT} \to_p (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\boldsymbol{F}_t \tag{28}$$

$$\hat{\tau}^{synth}(s,t) - \tau(s,t)^{ATT} \to_p 0. \tag{29}$$

3. *If Assumption 3 holds, then the difference-in-means estimator is asymptotically unbiased, even with* T *fixed:*

$$\hat{\tau}^{cont}(s,t) - \tau(s,t)^{ATT} \to_p 0. \tag{30}$$

4. *If* $\boldsymbol{F}_t^o = \boldsymbol{F}_t \;\forall\, t$, *then the abnormal return estimator is consistent:*

$$\tau^{AR}(s,t) - \tau(s,t)^{ATT} \to_p 0. \tag{31}$$

The most complex part of this proof, proof of asymptotic unbiasedness of the synthetic control estimator, follows directly from Ferman (2021), who show that the synthetic control estimator is asymptotically unbiased under the assumption of exact matching. The other two results follow from the assumptions and the definition of the estimators.

Remark 2. Both the misspecified abnormal return estimator and the difference-in-means estimator in a given time period are inconsistent. Both converge to a random variable that is a linear combination of the two factors, but the linear combination varies depends on the factor loadings (and factor correlation).[4] In contrast, the synthetic control estimator is consistent and converges to the true effect. If the abnormal return estimator is correctly specified, then it is also consistent.

[4]These inconsistencies are similar to the inconsistencies highlighted in Theorem 1 of Andrews (2005).

We next consider how these results change if there are multiple event periods.

Theorem 1 (Bias in multiple event periods)**.** *Let Assumptions 1, 2 and 5 hold. Then,*

1. *In finite samples, the estimators differ from their target estimands as a function of how closely the constructed counterfactual portfolio matches the treated portfolio:*

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} = \sum_{s \in \mathcal{S}} w_s \left((\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_{s+\kappa} - \hat{\beta}_s \boldsymbol{F}^{o}_{s+\kappa}) \right) \tag{32}$$

$$\hat{\theta}^{cont}_{\kappa} - \theta^{ATT}_{\kappa} = \sum_{s \in \mathcal{S}} w_s ((\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) F_t) \tag{33}$$

$$\hat{\theta}^{synth}_{\kappa} - \theta^{ATT}_{\kappa} = \sum_{s \in \mathcal{S}} w_s \left((\alpha_s - \hat{\alpha}^{synth}_s) + (\beta_s - \hat{\beta}^{synth}_s) F_t \right). \tag{34}$$

2. *Then, if $n_s, n_c, T \to \infty$ and $p_t(\boldsymbol{X}_i, \boldsymbol{F}) > \epsilon_i$ for a non-trival share of $t \in \mathcal{S}$, then asymptotically, the synthetic control estimator is unbiased, but the other two estimators are biased and converge to a weighted combination of conditional expected risk premia:*

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} = E\left((\alpha_s - \tilde{\alpha}_s) + (\beta_s \boldsymbol{F}_{s+\kappa} - \tilde{\beta}_s \boldsymbol{F}^{o}_{s+\kappa}) \mid T_i \in \mathcal{S} \right) \tag{35}$$

$$\hat{\theta}^{cont}_{\kappa} - \theta^{ATT}_{\kappa} = E\left((\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) F_t \mid T_i \in \mathcal{S} \right) \tag{36}$$

$$\hat{\theta}^{synth}_{\kappa} - \theta^{ATT}_{\kappa} \to_p 0. \tag{37}$$

3. *If in addition, Assumption 3 holds, then the difference-in-means estimator is asymptotically unbiased, even with T fixed.*

4. *If instead Assumption 4 holds, then the abnormal return and difference-in-means estimators can be written as:*

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} = E\left(\alpha_s - \tilde{\alpha}_s \mid T_i \in \mathcal{S} \right) \tag{38}$$

$$+ E\left(\beta_i \mid T_i \in \mathcal{S} \right) E\left(\boldsymbol{F}_t \right) \tag{39}$$

$$- E\left(\tilde{\beta}_i \mid T_i \in \mathcal{S} \right) E\left(\boldsymbol{F}^{o}_{s+\kappa} \right) \tag{40}$$

$$\hat{\theta}^{cont}_{\kappa} - \theta^{ATT}_{\kappa} = E\left(\alpha_s - \alpha_\infty \mid T_i \in \mathcal{S} \right) + E\left(\beta_s - \beta_\infty \mid T_i \in \mathcal{S} \right) E\left(\mathbf{F}_t \right) \tag{41}$$

4 Discussion of estimator performance in different settings

4.1 Negligible bias in short-run (high-frequency) event studies

A common phrase described in event studies is that the structure of the model used in τ^{AR} does not have significant impacts on the estimated effects. For example, in footnote 5, Shleifer

(1986) states "The [index inclusion] results were not materially different when returns were not corrected for market movements. Similarly, combining the before and after estimation periods did not make much difference." Or in Edmans (2012) "I use the standard short event-study window so that the calculation of abnormal returns is relatively insensitive to the benchmark asset pricing model used."

We can show why short-run abnormal return models work quite well, regardless of modeling choices, using our results from the previous section. First, consider the estimated ATT for a stock market event on announcement day on a given day, $\tau^{AR}(s,s)$. The bias in this estimate relative to $\tau(s,t)^{ATT}$ is:

$$\tau^{AR}(s,s) - \tau^{ATT}(s,s) = (\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_s - \hat{\beta}_s \boldsymbol{F}^o_s). \tag{42}$$

First, note $(\alpha_s - \hat{\alpha}_s)$ is typically quite small. One reason is because the average return in excess of aggregate risk exposure for the portfolio of assets experiencing the event in period s, α_s, is typically quite small (e.g. under many models of returns, this should be zero for a full portfolio (Chamberlain and Rothschild, 1983)). Hence, α is ruled out as a source of error for much of our discussion.

Second, the error component associated with the factors, $(\beta_s \boldsymbol{F}_t - \hat{\beta}_s \boldsymbol{F}^o_t)$ is likely very small in the short run. To see why, consider the extreme case where $\boldsymbol{F}^o_t = 0$, and hence no factors are considered. Then, it is a simple exercise to consider how large $\beta_s \boldsymbol{F}_t$ can be in a daily return. Note that this effect is typically quite small *on average* – for example, the average daily return for the market factor is 0.05%. However, there is substantial variation in the size of these factors, with an interquartile range of 1% and very large fat tails. Hence, the correlation of the factors with the timing of the event is very important. This will be apparent in our first empirical example of Acemoglu et al., 2016. Formally,

Proposition 2. *Hold fixed* $\alpha, \beta, \hat{\alpha}, \hat{\beta}$ *and* $\mathcal{F}^o_t$*. Then,* $|\tau^{AR}(s,s) - \tau^{ATT}(s,s)|$ *is increasing in* $|F_t|$ *for entries where* β_s *is non-zero.*

Remark 3. Let $|\beta_s F_t|/|\tau^{ATT}(s,t)| = a(s,t)$ be the relative size of the aggregate factor impact compared to the treatment effect size. If $a(s,t)$ is small, then the percentage bias for $\tau^{AR}(s,t)$ will be small (so long as the estimated model does not make the estimate significantly worse).

Remark 4. The volatility of $|\beta_s F_t|$ is time varying (e.g. the market is much more volatile in some time periods), then there some periods where $a(s,t)$ is large, and the percentage bias may be large.

Remark 5. The percentage bias may be large simply if $|\tau^{ATT}(s,t)|$ is small. Hence, these biases become much more significant when the effects are smaller in size.

Now consider the aggregated ATT θ_{κ}^{ATT}:

$$\hat{\theta}_{\kappa}^{ar} - \theta_{\kappa}^{ATT} = \sum_{s \in \mathcal{S}} w_s \left((\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_{s+\kappa} - \hat{\beta}_s \boldsymbol{F}^o_{s+\kappa}) \right) \tag{43}$$

Here, $\sum_{s \in \mathcal{S}} w_s (\alpha_s - \hat{\alpha}_s)$ is even more likely to be zero, as the portfolio across all events likely has small alpha. The second term, $\sum_{s \in \mathcal{S}} w_s \left(\beta_s \boldsymbol{F}_{s+\kappa} - \hat{\beta}_s \boldsymbol{F}^o_{s+\kappa} \right)$, can be written as

$$\sum_{s \in \mathcal{S}} w_s \left(\beta_s \boldsymbol{F}_{s+\kappa} - \hat{\beta}_s \boldsymbol{F}^o_{s+\kappa} \right) = E \left(\beta_s \boldsymbol{F}_{s+\kappa} - \hat{\beta}_s \boldsymbol{F}^o_{s+\kappa} \mid s \in \mathcal{S} \right), \tag{44}$$

integrating over the different event time periods. Assumption 4 assumes that factors are randomly drawn and the factor loadings of the firms who are treated are not correlated with the factor draw on a given day. Thus, if event timings are *random*, such that $E(\boldsymbol{F}_{s+\kappa} \mid s \in \mathcal{S}) = E(\boldsymbol{F}_{s+\kappa})$, then we can effectively assume $E(\boldsymbol{F}_{s+\kappa} \mid s \in \mathcal{S}) \approx 0$ (in the case of the market, average daily return is four basis points).

Hence, for θ_k^{AR}, so long as the event timings are random across time, the choice of the model is irrelevant for point identification.

Corollary 1. *If the timing of $\mathcal{S}$ is random, then θ_k^{AR} is approximately unbiased, regardless of the choice of $\boldsymbol{F}^o_t$.*

4.2 Increasing cumulative bias in long-run event studies

Researchers are often interested in the trends or cumulative impact of events on returns. These are sometimes referred to as cumulative abnormal returns or buy-and-hold abnormal returns (referring to additive vs. geometric return differences). This gets mapped to different economic and behavioral theories about how the market processes information (e.g. Daniel et al. (1998) is a theory to explain these effects from a behavioral perspective; Kwon and Tang (2022) consider 90 day post-announcement effects relative to announcement day effects).

Some papers have pointed to flaws in studying these types of long-run perspectives – for example, Mitchell and Stafford (2000) highlight the flaws in the inference around long-run abnormal return studies of firm activity. We use our results to highlight exactly how these problems become amplified as we focus on the long-run.

Fix our aggregate average treatment effect estimand for a single event timing s to be

$$\tau^{AATT}(s) = \sum_{\kappa=0}^{K_0} \tau(s, s+\kappa) \tag{45}$$

and over all timings

$$\tau^{AATT} = \sum_{s \in \mathcal{S}} w_s \sum_{\kappa=0}^{K_0} \tau(s, s+\kappa). \tag{46}$$

Our estimators are analogous, summing up over the different estimators. For example,

$$\hat{\tau}^{AR,AATT}(s) = \sum_{\kappa=0}^{K_0} \hat{\tau}^{AR}(s, s+\kappa). \tag{47}$$

Now consider the impact of cumulating the bias over time:

$$\hat{\tau}^{AR,AATT}(s) - \tau^{AATT}(s) = K_0(\alpha_s - \hat{\alpha}_s) + \beta_s \sum_{\kappa=0}^{K^0} F_{s+\kappa} - \hat{\beta}_s \sum_{\kappa=0}^{K^0} F^o_{s+\kappa} \tag{48}$$

$$\hat{\tau}^{cont,AATT}(s) - \tau^{AATT}(s) = K_0(\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) \sum_{\kappa=0}^{K^0} F_{s+\kappa} \tag{49}$$

$$\hat{\tau}^{synth,AATT}(s) - \tau^{AATT}(s) = K_0(\alpha_s - \hat{\alpha}^{synth}_s) + (\beta_s - \hat{\beta}^{synth}_s) \sum_{\kappa=0}^{K^0} F_{s+\kappa}. \tag{50}$$

Note that the bias in these estimators is intimately related to the properties of $\sum_{\kappa=0}^{K^0} F_{s+\kappa}$ – the cumulative sum of daily returns for the different factors. Unlike in the short-run, factors are expected to have a positive drift associated with (as the risk of the factors leads to positive expected return). Hence, we can rewrite $\sum_{\kappa=0}^{K^0} F_{s+\kappa} \approx K^0 E(F_t | s+\kappa \geq t \geq T_i)$. Now we can see that it's possible to generate drift proportional to the expected value of the factors during the time period, scaled by the relative estimation error in β_s.

Now consider estimating the long-run impact of a merger on stock market prices. Raghavendra Rau and Vermaelen (1998) find a three-year long run effect of -4% for all mergers, while Savor and Lu (2009) find a three-year long-run effect of -13.1% for stock-financed mergers and 1.6% for cash financed mergers. These results are well-motivated by Shleifer and Vishny (2003), but their magnitude may reflect bias due to the errors in $\hat{\beta}$. To give an example, assume that $E(F_t \mid s+\kappa \geq t \geq T_i)$ is strictly positive (reflecting risk premia). Then, the bias will be negative if the synthetic portfolio is more exposed to the risk factors with larger premia:

$$(\beta_s - \hat{\beta}_s)E(F_t) = \sum_{k=1}^{K} (\beta_{sk} - \hat{\beta}_{sk})E(F_{tk}). \tag{51}$$

If $K = 1$, for example, and was equal to the market, then our expected excess return is 6%. If $\beta_{sk} - \hat{\beta}_{sk})$ was -0.1, then at the three year level, we might expect a bias of -1.8%. This is of course an empirical question of which way the biases would go; is the constructed portfolio of firms too heavily loaded on risk factors?

Note that these issues do *not* disappear as we move to integrate over all event timings. This bias in factors cannot average out to zero, and so the only source by which we can achieve zero bias is through mean zero differences in the loadings.

It is also worth remarking how the results from Mitchell and Stafford (2000) can be seen analytically in our statistical terms. While the misspecification term $(\beta_s - \hat{\beta}_s)\sum_{\kappa=0}^{K^0} F_{s+\kappa}$ creates bias, it also creates cross-correlation in errors for every event-timing.[5]

4.3 Individual estimates are noisy, but not necessarily biased

Consider the case of a single firm being treated. To analyze this case, we need to allow for slightly more flexibility in our notation.

Assumption 7. *Let* $R_{it}(\infty) = \alpha_i + \beta_i \boldsymbol{F}_t + \varepsilon_{it}$*, where* ε_{it} *is i.i.d. across firms, and i.n.i.d. across time, and mean zero.*

Remark 6. This assumption implies we can write $R_{it}(T_i) = R_{it}(\infty) + \tau_i(s,t) = \alpha_i + \beta_i \boldsymbol{F}_t + \tau_i(s,t) + \varepsilon_{it}$.

Then, consider the case of a single firm estimated in each estimator:

$$\tau_i^{AR}(s,s) - \tau_i(s,s) = (\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_s - \hat{\beta}_s \boldsymbol{F}_s^o) + \varepsilon_{it}. \tag{52}$$

Statistically, there are now three objects with randomness to worry about: the estimated parameters, the aggregate factors, and the idiosyncratic variance for the individual firm. Note that with several treated units, this last term disappears, but with a single unit, we have insurmountable noise. This is a common problem flagged in the event studies literature looking at securities litatigation (Baker, Gelbach, et al., 2020).

However, consider an approach that estimates many individual treatment effects in this manner (such as Kogan et al. (2017)). On, average, these estimates will be subject to the same results outlined above, but each one is quite noisy. This is equivalent to problems associated with estimating many treatment effects. One approach is to consider shrinkage estimators. Another would be to pool the firms based on characteristics of interest, and construct portfolios this way. This would remove ε.

4.4 Key takeaways re: randomness

Key takeaways for practitioners are four-fold:

[5] As they state: "[M]ajor corporate events cluster through time by industry. This leads to positive cross-correlation of abnormal returns, making test statistics that assume independence severely overstated."

1. If treatment is randomly assigned across firms, then comparing returns to the average of the market is as good as any other approach.

2. If treatment is randomly assigned across periods, and there are multiple event timings, then the model used to estimate effects does not matter in the short-run.

3. If treatment is randomly assigned across periods, but the model used to estimate effects is misspecified, then the estimates will be biased, even with many event timings.

4. These results are identical whether there is a single treated firm or many treated firms.

5 Simulations

We highlight how the non-random timing and assignment, together with a misspecified factor model, could affect the bias with different estimators of treatment effects, using a simple simulation exercise. In the simulation, the returns follow a two-factor structure, with the second factor omitted in the estimation of abnormal returns. We compare the expected bias, root mean square error, and coverage with random vs. nonrandom assignment and timing.

5.1 Simulation Design with 2 Factors and Selection

We simulate a panel of stock returns with a linear factor structure:

$$r_{it} = r_{f,t} + \beta_{i,mkt}(r_{mkt,t} - r_{f,t}) + \beta_{i,smb} r_{smb,t} + \varepsilon_{i,t}, \tag{53}$$

where the return for each stock equals to the risk-free rate, plus the exposure times risk premium of a market factor and a size factor (small-minus-big), and a stock-level idiosyncratic component.

We assume that both factor loadings follow independent normal distributions: $\beta_{i,mkt}, \beta_{i,smb} \sim \mathcal{N}(1, 0.3^2)$. We further assume that the idiosyncratic component of each stock is drawn i.i.d. from a Normal distribution: $\varepsilon_{i,t} \sim \mathcal{N}(0, 0.1^2)$. We choose a standard deviation of around 0.1 so that the residual variance constitutes approximately half of the total variance.

We simulate returns for 500 firms, with pre-treatment period of 239 days, 1 event day, and 10 post-treatment periods. Roughly 10% of firms are treated, following one of two treatment assignment processes, discussed below. Treated firms get a true effect of 3% on the treatment day, and nothing afterwards. The factor returns and the risk-free rate are randomly sampled from daily Fama-French returns from July 1926 to 2022 with block sampling to preserve the correlation structure between factors.

Treatment assignment process We compare expected bias with different treatment assignment selection and timing selection. For firm assignment, we either completely randomly assign the treatment to 10% of firms (consistent with Assumption 3), or to instead relax this assumption, we model that the probability of a firm getting treated follows a logit function of the beta on the SMB factor

$$p(treated)_i = \frac{\exp(\delta\beta_{i,smb})}{1+\exp(\delta\beta_{i,smb})}, \tag{54}$$

where $\delta = \frac{\log(0.1)}{E(\beta_{i,smb})} < 0$ to achieve an average probability of 10%. The lower the simulated SMB factor loading of the firm, the more likely to be treated.

For treatment period selection, we similarly use two different assignment mechanisms. The first is to randomly sample the 250 data periods, and always set the treatment period equal to $t = 240$. This effectively makes the treatment period's factor draw uncorrelated with the treated firms' factor loadings. The secon approach with timing selection works as follows. First, we rank the SMB factor in 250 candidate treatment periods. We then use the rank of SMB returns as inputs to the selection function.[6] The probability of any one of the candidate period being the treatment period is

$$p(selected)_t = \frac{\exp(\delta Rank_{2t})}{1+\exp(\delta Rank_{2t})}, \tag{55}$$

where $\delta = \frac{\log(1/250)}{E(Rank_t)}$. We then draw indicator variables for each candidate period from binomial distributions with respective treatment probability in each period. If multiple periods are drawn to be the event period, we use the one with the highest factor realization. Thus, if a period has a high factor realization of the omitted factor, it is more likely to become the treatment period.

5.2 Simulation Results with 2 Factors and Selection

In Table 1, we compare the performance of four different estimators across 50 simulations: mean difference between treated and control firms, average abnormal returns using the market factor (estimating the factor loading for each treated firm in the pre-period), average abnormal returns using the both factors (estimating the factor loadings for each treated firm in the pre-period), and average treatment effects from the generalized synthetic control method (Gsynth). Estimated bias is reported in percentage points. We also report the root mean square error (RMSE) and coverage of 95% confidence intervals.

[6] Raw factors returns have positive and negative values with mean close to 0, which will make the logit function highly sensitive.

Table 1: **Treatment Effect Bias and Coverage in Simulations: Two-Factor Structure** This table presents the bias and coverage of different estimators of treatment effects in financial returns. We simulate 500 firms with 10% treated. The estimation period is 239 days and post-event period is 11 days. More details on the simulations is in Section 5.1. Panel A reports simulation results with no selections, Panel B with only assignment selection, Panel C with only timing selection, and Panel D with both. We consider several estimators: difference in simple average, CAPM and 2-factor abnormal returns, and generalized synthetic methods. The expected biases and coverage are from 50 simulations.

Panel A: Random Assignment + Random Timing

	All Periods			Treated Periods			Untreated Periods		
Model	E(Bias)	MAD	RMSE	E(Bias)	MAD	Coverage	E(Bias)	MAD	Coverage
Simple Means	0.00	0.04	0.58	0.01	0.17	1	0.00	0.04	0.03
CAPM	-0.06	0.16	2.11	-0.07	0.43	1	-0.06	0.17	0.44
Correct Factor Structure	-0.01	0.04	0.54	0.00	0.16	1	-0.01	0.04	0.04
Gsynth (PCA)	0.00	0.04	0.56	0.02	0.17	1	0.00	0.04	0.03

Panel B: Assignment Selection + Random Timing

	All Periods			Treated Periods			Untreated Periods		
Model	E(Bias)	MAD	RMSE	E(Bias)	MAD	Coverage	E(Bias)	MAD	Coverage
Simple Means	0.02	0.05	0.71	0.04	0.18	1.00	0.02	0.05	0.11
CAPM	-0.05	0.13	1.78	-0.04	0.35	0.98	-0.05	0.14	0.40
Correct Factor Structure	-0.01	0.03	0.54	0.02	0.14	1.00	-0.01	0.04	0.04
Gsynth (PCA)	0.00	0.04	0.57	0.03	0.15	1.00	0.00	0.04	0.05

Panel C: Random Assignment + Timing Selection

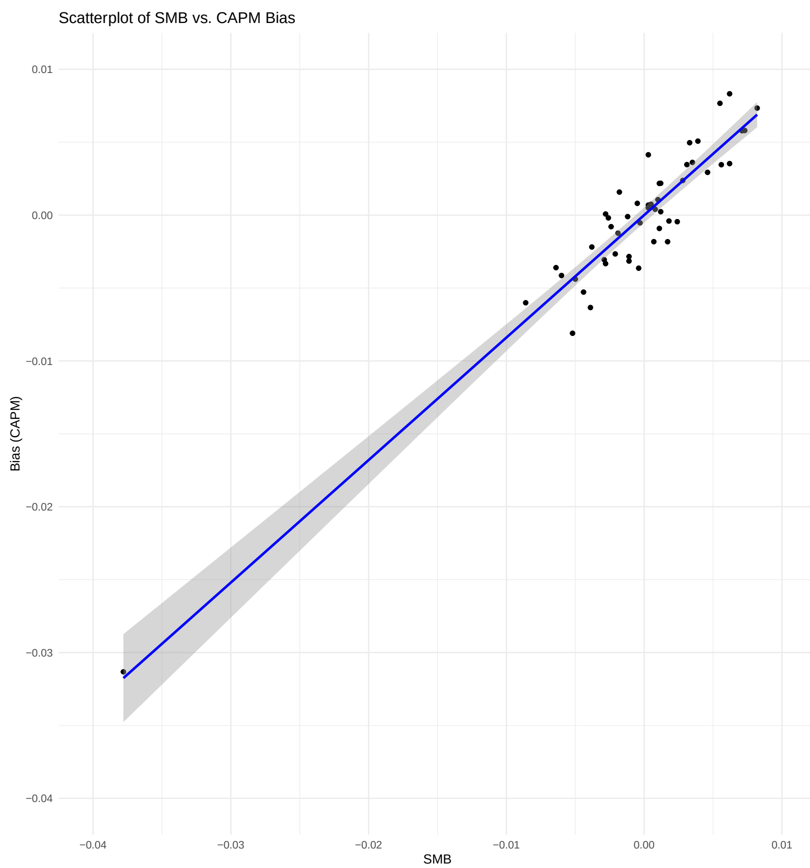
	All Periods			Treated Periods			Untreated Periods		
Model	E(Bias)	MAD	RMSE	E(Bias)	MAD	Coverage	E(Bias)	MAD	Coverage
Simple Means	-0.01	0.05	0.63	0.00	0.21	1	-0.01	0.05	0.05
CAPM	0.25	0.27	3.49	2.71	2.71	1	0.00	0.16	0.46
Correct Factor Structure	-0.02	0.04	0.54	0.00	0.12	1	-0.02	0.04	0.04
Gsynth (PCA)	-0.01	0.04	0.57	0.01	0.13	1	-0.01	0.04	0.04

Panel D: Assignment Selection + Timing Selection

	All Periods			Treated Periods			Untreated Periods		
Model	E(Bias)	MAD	RMSE	E(Bias)	MAD	Coverage	E(Bias)	MAD	Coverage
Simple Means	-0.05	0.07	0.88	-0.52	0.52	1	-0.01	0.05	0.08
CAPM	0.21	0.23	2.92	2.26	2.26	1	0.00	0.13	0.40
Correct Factor Structure	-0.02	0.04	0.52	0.01	0.12	1	-0.02	0.04	0.05
Gsynth (PCA)	-0.01	0.04	0.56	-0.01	0.14	1	-0.01	0.04	0.04

First, in Panel A, we see that the average bias is small even with the wrong factor structure, if the treatment is randomly assigned. Similarly, in Panel B, if we only have non-random assignment selection, the expected bias is also insignificant on average. However, this masks the variation across simulations - if a time period has a larger factor draw on the treatment day, that leads to much larger bias (Section 5.2).

Figure 1: **Bias from CAPM Model on SMB Returns with Assignment Selection** This figure plots the biases from a CAPM estimator on the treatment period over realizations of the second factor across 50 simulations. We simulate 500 firms with 10% of them getting treated. The estimation period is 239 days and post-event period is 11 days. More details on the simulations is in Section 5.1. Panel A reports simulation results with no selections, Panel B with only assignment selection, Panel C with only timing selection, and Panel D with both. We consider several estimators: difference in simple average, CAPM and 2-factor abnormal returns, and generalized synthetic methods. The expected biases and coverage are from 50 simulations.



In Panel C, we consider random assignment of treatment to units, but non-random event timing. As in Panel A, the difference in means is unbiased thanks to the results in Proposition 1. Since treatment is uncorrelated with factor loadings, there is no endogeneity and the simple means estimator is an unbiased estimator of the treatment effect. However, with non-random timing, the CAPM model is biased, because the abnormal return (as discussed in Section 2.2) will be the average β for the omitted factor multiplied by the largest possible factor draw. In contrast, the difference in means is unbiased because while both treated and untreated firms are exposed to the high factor draw, they have identical factor exposures, which cancels out. For the correctly specified model, the estimated model correctly specifies the counterfactual, and so there is no bias. Finally, the Gsynth estimator is able to identify the correct underlying factor structure, and has limited bias as well.

Once we have both types of selection in treatment in Panel D, we see that the simple difference in means is now biased. However, it is still less biased in absolute value than the misspecified CAPM model. This is because the *gap* in the treatment and control factor loadings for the simple mean difference is still smaller than the level misspecification in the factor loadings in the CAPM estimation. Again, the Gsynth approach does quite well, with similar performance to the correctly specified factor model.

6 Applications

6.1 Empirical Example 1: Geithner as Treasury Secretary

We now turn to our first empirical example, and study the period when the annoucement of Timothy Geithner as Treasury Secretary was leaked, as in the setup of Acemoglu et al. (2016). This first example allows us to highlight the results of Proposition 1 in a simultaneous treatment setting. We first show that in this setting, the bias from an incorrect factor structure could be huge, and using synthetic control methods helps alleviate the bias. Second, we argue that the bias comes from two factors. One, in the event window, we see turbulent market and factor returns with large daily realizations. And counterfactual returns come from control firms with very different factor exposures. We show that synthetic methods, which greatly alleviate biases, match the beta of treated firms well.

Empirical setup We look at the announcement of Timothy Geithner as nominee for Treasury Secretary in November 21, 2008. Same as Acemoglu et al. (2016), we consider the average treatment effects in the 11 days on and after the announcement date, from November 21, 2008 as day 0, 24 as day 1 to December 8, 2008 as day 10. For treated and bank controls, we use the returns provided by the authors who collect daily returns from Datastream.[7] In all trading days before and after the events, the returns are full trading day returns in the regular trading hour, while for returns on the event day, the returns are from 3 p.m. until the market close at 4 p.m. We consider two types of control firms. First, we use the same set of financial firms listed on the NYSE or NASDAQ that are not connected with Geithner, as in Acemoglu et al. (2016). Second, we use all NYSE, AMEX, and NASDAQ (exchange code 1-3) common stocks (share code 10 or 11).

[7]We thank Amir Kermani for providing the replication code and data on his website.

6.1.1 Post-Event ATT

In this section, we study how different counterfactuals would lead to different results for the stock price reaction for firms with a connection to Timothy Geithner after the announcement of his nomination as U.S. Treasury Secretary. We study the 11-day (day 0 to day 10) average treatment effect after the announcement date on November 21, 2008.

Non-connected banks as controls We first use public financial institutions that are not connected to Geithner, as control firms. We report the average treatment effects of the 11-day post-event window in Panel A of Table 2. In column 1, we report the estimated difference in average returns between treated and control firms. This approach imposes the counterfactual returns as the simple average of returns from firms without a connection with Geithner, which is the same method as in Table 2 of Acemoglu et al. (2016). In column 2, we report the estimate from a difference-in-differences estimator, in column 3, we report the estimate using a standard synthetic control method (Abadie et al., 2010), in column 4, we use the synthetic difference-in-differences estimator from Arkhangelsky et al., 2021, and in column 5, we use a generalized synthetic control method (Gsynth) from Xu (2017). For columns 2 to 4, the pre-event periods for estimation is day -256 to -31, which is similar to day -280 to -31 used in the original paper. We keep a shorter estimation period to maintain a balanced panel for synthetic methods.

We see that on average, within 10 days after the event, schedule connection treated firms have 2.6% higher daily returns, personal connection treated firms have 2.9% higher daily returns, and New York connection treated firms have 1.9% higher daily returns. Using difference-in-differences does not significantly change the estimate. However, using synthetic control methods significantly decreases the estimated effects.[8] The standard synthetic control methods decreases effects by 33%, possibly due to a closer match of control firms with treated firms. The synthetic DinD has a similar decrease in average treatment effects (25%), and Gsynth has a slightly larger decrease of 46%. However, these effects are still significant over this period.

All public firms as controls We then expand the set of control firms to all common shares in NYSE, AMEX, and NASDAQ and report the results in Panel B of Table 2. The reason for this is quite simple: given that equities markets are integrated, the underlying factors can be well-constructed using all the stocks traded on U.S. equities. When trying to control for an

[8]Our results stand in contrast to Acemoglu et al. (2016), who use the synthetic control method to attempt to provide robustness results to their design. However, much of their approach is somewhat *ad hoc* due to a lack of extant literature at the time on how to deal with many treated units in a synthetic control approach appropriately. For example, they are forced to use placebo designs for confidence intervals, rather than asymptotic normality, leading to likely quite conservative confidence intervals.

Table 2: **ATT of Treasury Secretary Announcement** This table presents average treatment effects after the announcement of Timothy Geithner as Treasury Secretary. Event day 0 is November 21, 2008 from 3pm (when the news leaked) to market closing, consistent with Acemoglu et al. (2016). The average treatment effect is estimated using post periods from trading day 0 to day 10. We consider two control samples: banks or financial services firms trading on the NYSE or Nasdaq (Panel A), and all NYSE, AMEX, and NASDAQ common stocks (Panel B). We consider several estimators: difference in simple average, difference-in-differences, synthetic control, synthetic DinD, and generalized synthetic methods. Standard errors of simple average is from a two-sample t-test. Standard errors of DID, synthetic control, and synthetic DID are calculated using placebo inference following Arkhangelsky et al. (2021) with 100 repetitions. Standard errors of Gsynth is computed using parametric bootstrap with 1,000 samples. Standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Panel A: Bank Controls

	(1) Average	(2) DID	(3) SC	(4) SDID	(5) Gsynth
Schedule connections	0.026***	0.027***	0.016***	0.018***	0.012**
	(0.007)	(0.005)	(0.005)	(0.005)	(0.006)
Personal connections	0.029***	0.030***	0.004	0.009**	0.008
	(0.010)	(0.006)	(0.003)	(0.005)	(0.007)
New York connections	0.019***	0.020***	0.009***	0.012***	0.009**
	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)
Observations	5,995	129,165	129,165	129,165	129,625

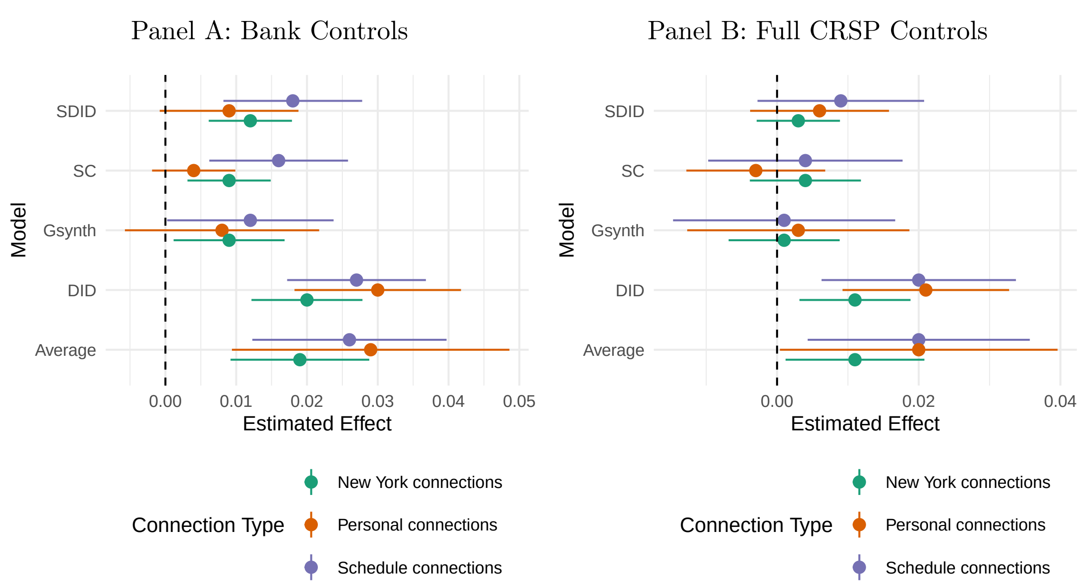
Panel B: All Firm Controls

	(1) Average	(2) DID	(3) SC	(4) SDID	(5) Gsynth
Schedule connections	0.020**	0.020***	0.004	0.009	0.001
	(0.008)	(0.007)	(0.007)	(0.006)	(0.008)
Personal connections	0.020*	0.021***	-0.003	0.006	0.003
	(0.010)	(0.006)	(0.005)	(0.005)	(0.008)
New York connections	0.011**	0.011***	0.004	0.003	0.001
	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)
Observations	45,045	966,420	966,420	966,420	916,388

underlying latent factor process, there are limited reasons to focus on just banks as a control, unless there is an important omitted variable that correlates with the banking industry that cannot be spanned by other stocks.

With a larger set of controls, synthetic control methods do a much better job constructing counterfactual returns. In Panel B, we see that with a larger set of control firms, while the average difference (and difference-in-difference) methods continue to give significant treatment effects, the synthetic methods estimate small and insignificant average treatment effects in the post-event window with standard SC, synthetic DinD, and generalized SC. In fact, the Gsynth approach estimates almost an exact zero. In contrast, the simple average in means between treated and all public control firms yields similar positive significant treatment effects of 2%, possibly due to a mismatch of factor loadings. We now consider why these estimated effects differ so much between estimators and the two panels.

Figure 2: **Connections to Geithner and Returns after Treasury Secretary News.** This figure plots the average treatment effects on the treated from Table 2 after the announcement of Timothy Geithner as Treasury Secretary. Event day 0 is November 21, 2008 from 3pm (when the news leaked) to market closing, consistent with Acemoglu et al., 2016. The average treatment effect is estimated using returns from trading day 0 to day 10. We consider two control samples: banks or financial services firms trading on the NYSE or Nasdaq (Panel A), and all NYSE, AMEX, and NASDAQ common stocks (Panel B). We consider several estimators: difference in average, difference-in-differences, synthetic control, synthetic DinD, and generalized synthetic methods. Standard errors of difference in average is from a two-sample t-test. Standard errors of DID, synthetic control, and synthetic DID are calculated using placebo inference following Arkhangelsky et al., 2021 with 100 repetitions. Standard errors of Gsynth is computed using parametric bootstrap with 1,000 samples.



6.1.2 Market Returns around Event

We now consider the reasons driving the biased estimates in the original analysis. First, we investigate where the realization of daily returns on the treatment dates are distributed on the daily return distribution. In Figure 3, we first plot the kernel density of the daily returns of S&P 500 index, and overlay it with the daily returns of 10 days after the announcement. We observe that the market return realizations are very volatile during the event period, swinging to the tails of the return distributions. The market returns on the onset of the event, November 21 and 24, 2008, are 6.9% and 6.5%. The most negative market realization happened on event day 5 (December 1, 2008). With a simultaneous single event treatment, large swings in underlying market factor returns can result in large biases if the beta on treated and counterfactual returns are not matched well. As show in Proposition 1, the estimator that uses a single period to estimate treatment effects is highly influenced by the relative size of the potential omitted factors. As a result, this could explain why we see a large decrease in average treatment effects using synthetic methods instead of comparing the mean returns between treated and control firm, since Proposition 1 suggests that synthetic control methods are much less sensitive to this problem.

6.1.3 Treated and Synthetic Betas

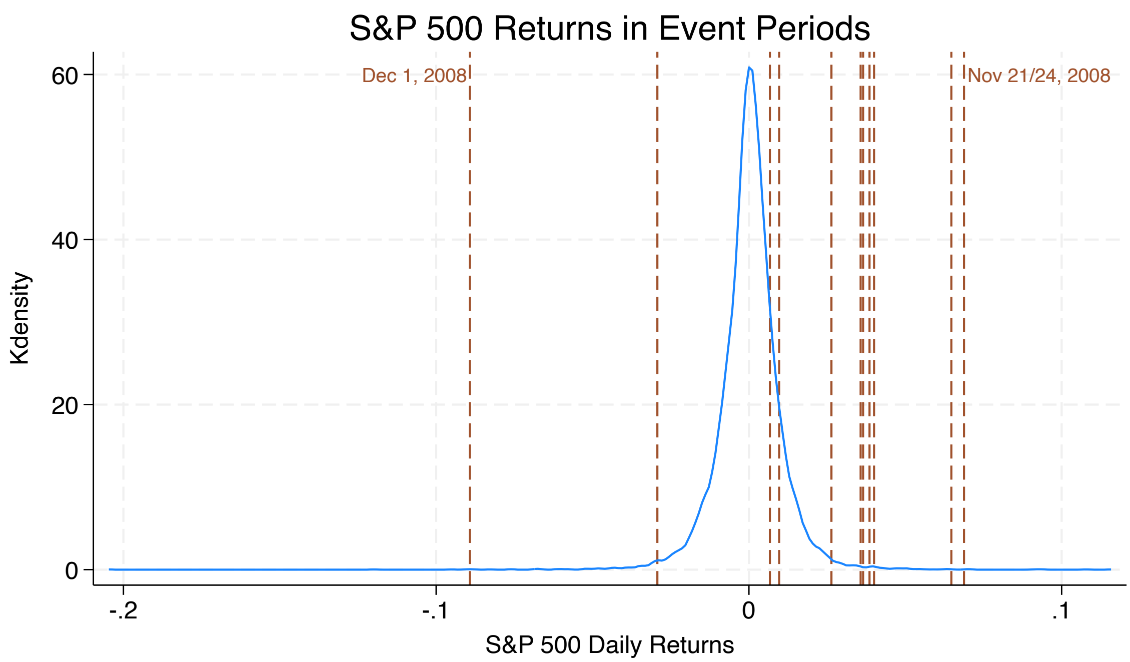
In this section, we show direct evidence on the betas of treated firms and the betas of portfolios of control firms formed using different synthetic methods. We estimate the betas using daily returns from day -280 to day -31 in the pre-event periods by running firm-by-firm time-series regressions of firms' daily returns on market returns (S&P 500 index returns) for CAPM betas, and on Fama-French three-factor returns.[9]

We report the (weighted) average of betas of treated and control firms in Table 3. First, in Panel A, we first show the average CAPM and Fama-French three-factor betas of the treated firms and equal-weighted averages of financial firm controls and all public firm controls. The average CAPM beta of the treated firms is 1.43, much higher than 0.83 from the control firms. Expanding to a three-factor model, we still see a higher market beta in treated firms. Given these mismatches of treated and control betas, together with turmoil market returns, as shown in Section 6.1.2, could lead to large biases in average treatment effects by comparing treated versus control firms.

In Panel B, we compute the weighted average betas of control firms using synthetic control weights, with both standard synthetic control and synthetic difference in differences. First, we see that synthetic methods match the beta in the treated firms well. For example, the

[9]Fama-French factor returns are downloaded from Ken French's website https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Figure 3: **S&P 500 Returns around Treasury Secretary Announcement** This figure plots the daily returns of S&P 500 index around the announcement of Timothy Geithner as Treasury Secretary. Event day 0 is November 21, 2008 from 3pm (when the news leaked) to market closing, consistent with Acemoglu et al., 2016. The blue solid line plots the kernel density function of daily S&P 500 returns from 1962 to 2023, and the sienna dashed vertical lines are the realization of daily returns in the post periods from trading day 0 to day 10. We label the dates with the largest outliers. The most positive realization is on event days November 21 and 24.



synthetic control gives a weighted average beta of 1.33, much closer to the treated beta of 1.43 than the equal-weighted average. Fama-French three-factor betas of the treated firms are 1.28 on the market, 0.23 on SMB, and 0.61 on HML, and synthetic control weights give a market beta of 1.15, SMB beta of 0.48, 0.75 (closer than 0.66, 0.75, and 0.72 with simple average). Second, if we extend the set of possible control firms from financial firms in Acemoglu et al. (2016) to all public firms in CRSP, we obtain better matches across all synthetic methods. For synthetic control specifically, controlled firms give an average beta of 1.38, closer to 1.43 in the treated firm. There is also a significant improvement in matching the Fama-French three-factor betas, synthetic control betas are 1.22, 0.38, and 0.67 (compared to treated betas of 1.28, 0.23, and 0.61). Finally, standard synthetic control methods give slightly better weights than synthetic difference-in-differences, who is more directly related to a mimicking portfolio approach.

Overall, synthetic methods matches the beta of treated firms well, which results in a lower bias in the average treatment effects.

Table 3: **Treated and Control Betas in Geithner as Treasury Secretary** This table presents the average CAPM and Fama-French three-factor betas for the treated and control firms. We first estimate firm-level betas using daily stock returns from 280 to 30 days before the announcement of Timothy Geithner as Treasury secretary on Nov 21, 2008. We then average the betas within the treated firms and two control samples: banks or financial services firms trading on the NYSE or Nasdaq, and all NYSE, AMEX, and NASDAQ common stocks. In Panel A, we show the simple average of treated firms and two control firms, and in Panel B, we calculate weighted average beta using weights from various synthetic methods: synthetic control and synthetic DinD.

Panel A: Simple Averages

	Treated	Control	Control (All CRSP)
CAPM Beta	1.427	0.825	0.832
FF3F Market Beta	1.275	0.659	0.857
FF3F Size Beta	0.233	0.748	0.553
FF3F Value Beta	0.607	0.720	0.144

Panel B: Weighted Averages with Synthetic Methods

	Bank Controls		All CRSP Controls	
	SC	SDID	SC	SDID
CAPM Beta	1.331	1.111	1.383	1.281
FF3F Market Beta	1.148	0.905	1.220	1.165
FF3F Size Beta	0.480	0.819	0.377	0.627
FF3F Value Beta	0.750	0.872	0.674	0.593

6.2 Empirical Example 2: Index Inclusion

In this section, we study S&P 500 index inclusion announcements, looking at both the immediate announcement returns and the pre-announcement drifts to highlight our theoretical predictions on identification in a staggered event setting.

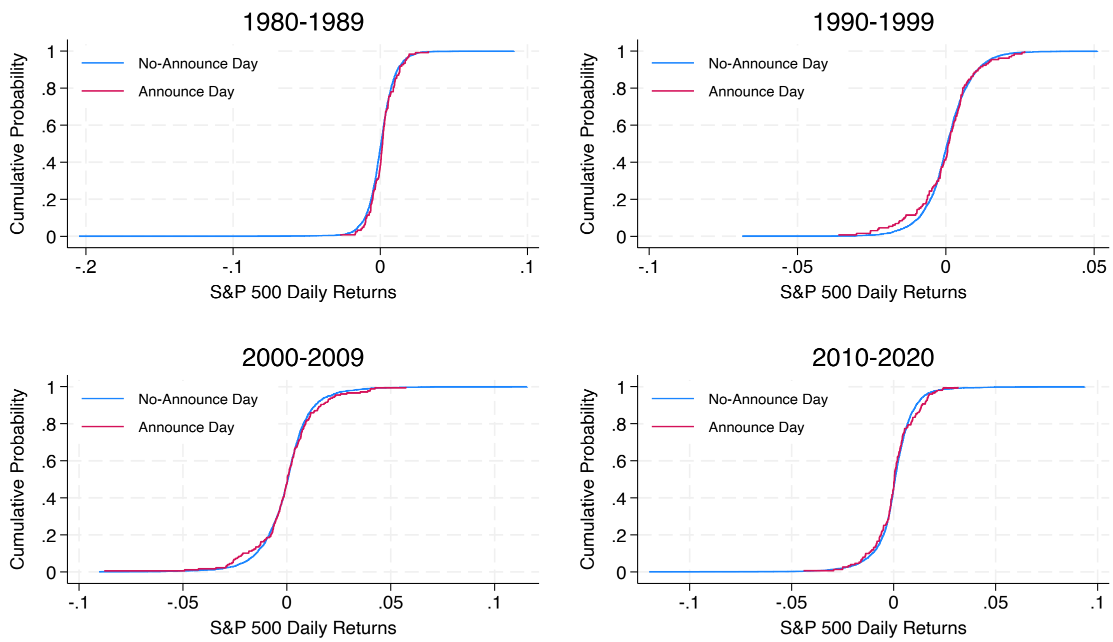
Empirical Setting We follow Greenwood and Sammon (2025) and use data from Siblis Research to obtain index inclusion dates. We match tickers from Siblis to the PERMNO code in CRSP using the CRSP header information. Siblis collects announcement dates for S&P 500 index inclusion. For those where the announcement dates are missing, from Sep 1976 to Sep 1989, index changes were announced after the close of market on Wednesdays, and the change in the index became effective the next day, so we can use the previous day of the effective date as the announcement date. We use the return on the announcement date if the inclusion is on a trading day. If not, we use the most recent trading day before the announcement date.

6.2.1 Factor Returns on Announcement Days

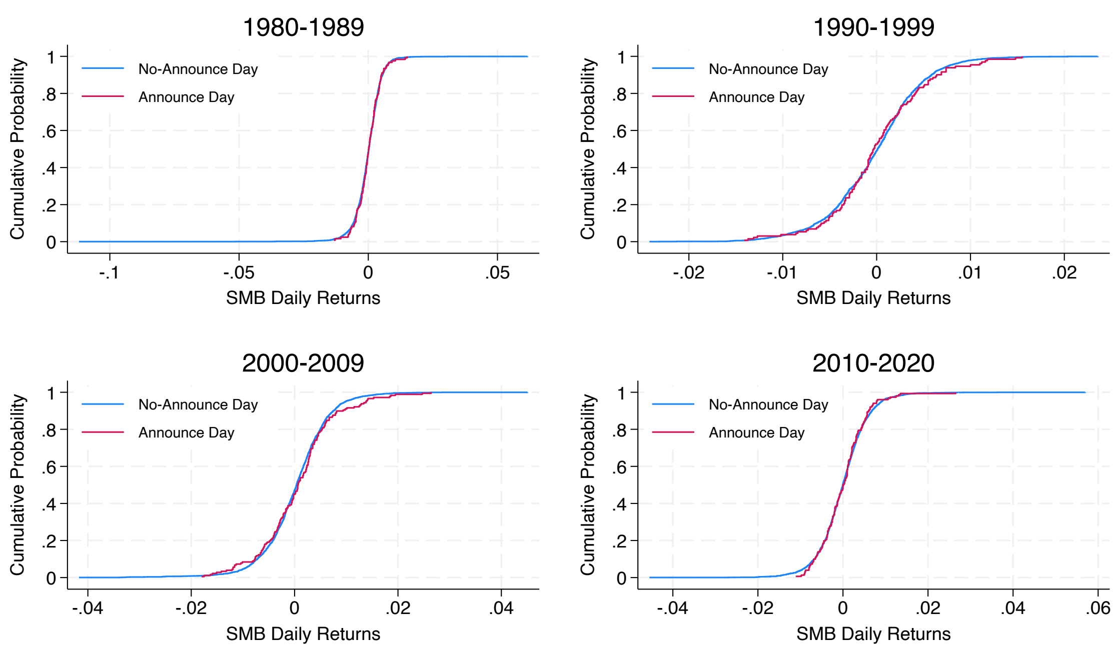
First, in Section 6.2.1, we show that the distributions of daily market returns on S&P 500 index inclusion announcement days are very similar to the ones on no-event days. This is true from 1980-1989 when the data start, to the recent decade 2010-2020. This is similarly true for the small-minus-big factor (Figure 4). Hence, the average timing of the events is defensible as random.

Figure 4: **Cumulative Distributions of Factor Returns by Announcement Status** This figure plots the daily returns of the S&P 500 index and Small-minus-Big (SMB) factor on the dates when there are index inclusion announcements versus the dates without. The blue line plots the overall cumulative distribution function from 1962 to 2023, and the red lines plot the cumulative distribution function of daily returns on the days when there is an index inclusion event.

Panel A: S&P 500 Daily Returns



Panel B: SMB Factor Daily Returns



6.2.2 Betas of Included Firms

In Table 4, we estimate the CAPM and Fama-French 3-factor betas for firms added to the S&P500 index using the daily returns from -250 to -100 days to the announcement date. We report the beta distributions separately for each decade from 1980 to 2020.

Table 4: **Beta Distributions of Included Firms across Decades** This table presents the average CAPM and Fama-French three-factor betas for firms included in S&P 500, compared with a random set of control firms of the same sample size. For each treated firm and inclusion date, we randomly pick a non-treat firm in CRSP sample with common share in NYSE, NASDAQ, or AMEX, which at least 250 trading days of returns before the announcement date. We then estimate firm-level betas using daily stock returns from 250 to 100 days before the announcement of inclusions into S&P 500 index. We provide the summary statistics for the distribution of betas of included firms, separately for each decade.

	Treated		Random Control	
	Mean	Std	Mean	Std
Panel A: 1980-1989				
CAPM Beta	0.961	0.523	0.582	0.551
FF3F Mkt Beta	1.108	0.539	0.854	0.784
FF3F SMB Beta	0.558	0.604	0.815	1.044
FF3F HML Beta	-0.148	0.987	0.021	1.188
Panel B: 1990-1999				
CAPM Beta	1.025	0.660	0.651	0.754
FF3F Mkt Beta	1.171	0.660	0.873	0.911
FF3F SMB Beta	0.489	0.661	0.805	1.215
FF3F HML Beta	-0.015	1.242	0.022	1.475
Panel B: 2000-2009				
CAPM Beta	1.087	0.697	0.824	0.985
FF3F Mkt Beta	1.079	0.560	0.820	0.688
FF3F SMB Beta	0.271	0.674	0.667	0.929
FF3F HML Beta	-0.002	1.227	0.075	1.482
Panel D: 2010-2020				
CAPM Beta	1.060	0.388	0.973	0.997
FF3F Mkt Beta	1.026	0.343	0.872	0.614
FF3F SMB Beta	0.225	0.520	0.628	1.201
FF3F HML Beta	-0.273	0.590	0.311	1.272

We see that in every decade, the average beta on the market portfolio is close to 1. Therefore, using a market model as the abnormal return model may not lead to material biases in the short run, as on average, the betas of treated firms are close to 1. This is consistent with our theoretical prediction in Section 3. If the average beta on the factor of treated firms is close to 1, using a market model that ignores beta estimation is approximately similar to estimating the CAPM model directly.

Both of these features – random timing of the many events, combined with seemingly limited selection – suggest that there should be limited bias in simple abnormal return estimates in the short-run according to Proposition 1. In Table 5, we find limited differences in treatment effects across estimators, consistent with this hypothesis.

Table 5: **Announcement-Day Treatment Effects of Index Inclusion** This table presents average treatment effects on the announcement days of index inclusion, averaged across inclusions for each decade. We consider several estimators: difference in simple average, CAPM, Fama-French 3-factor, and generalized synthetic methods. The estimation window of factor loadings are from -250 to -101 before the announcement dates. Standard errors of simple average is from a two-sample t-test. Standard errors of DID, synthetic control, and synthetic DID are calculated using placebo inference following Arkhangelsky et al., 2021 with 100 repetitions. Standard errors of Gsynth is computed using parametric bootstrap with 1,000 samples.

	Diff-in-Means	Market	CAPM	FF3F	Gsynth
1980-1989	3.27%	3.25%	3.15%	3.05%	3.06%
1990-1999	4.61%	4.62%	4.69%	4.71%	4.79%
2000-2009	3.42%	3.43%	3.33%	3.22%	3.41%
2010-2020	1.14%	0.94%	0.85%	0.85%	0.93%

6.3 Pre-inclusion Drift

Proposition 1 implies that there should be limited short-run bias (as we document above), but potentially large long-run bias (unless the omitted factors are exactly accounted for. We now examine the "pre-announcement drift," highlighted in Greenwood and Sammon, 2025, broken out by decade, as a form of potential longer-run bias.

However, studying the pre-announcement differences requires us to simultaneously assess Assumption 5, the limited anticipation condition. In financial markets, front-running of index inclusion is plausible, and as discussed in Greenwood and Sammon (2025), stocks with larger market cap that are close in size to the market cap of included stocks are much more likely to be included. We consider this possibility first, and how it affects the pre-announcment drift.

6.3.1 Anticipation and synthetic methods

Predictability in this context is complex. There is both the predictability of *when* index inclusion occurs (e.g. what day of the year) and then *which* firm. Ideally, we would capture both dynamics. Currently, we capture just the latter, by estimating a propensity score measure for the probability of whether a firm is added to the index using lagged characteristics.

Specifically, we estimate a logistic regression each year:

$$1(\text{Stock Added})_{i,y,m} = \alpha_y + \beta_y MktCapRank_{i,y,m-1} + \varepsilon_{i,y} \tag{56}$$

where we use the lagged market capitalization in end of last month before announcement to predict the probability of getting added to S&P. Consistent with Greenwood and Sammon (2025), we also find that in recent years, the addition to S&P is becoming more predictable using lagged size.

Then, for index inclusion events in each month, we match included stocks to a firm that was not included using the nearest neighbor in estimated propensity scores. This gives us two sets of returns: a portfolio of "included" stocks and "pseudo-included" stocks. If we believe that it was near-random which stock was included, and that markets anticipate this before the annoucement, then they should behave more similarly.

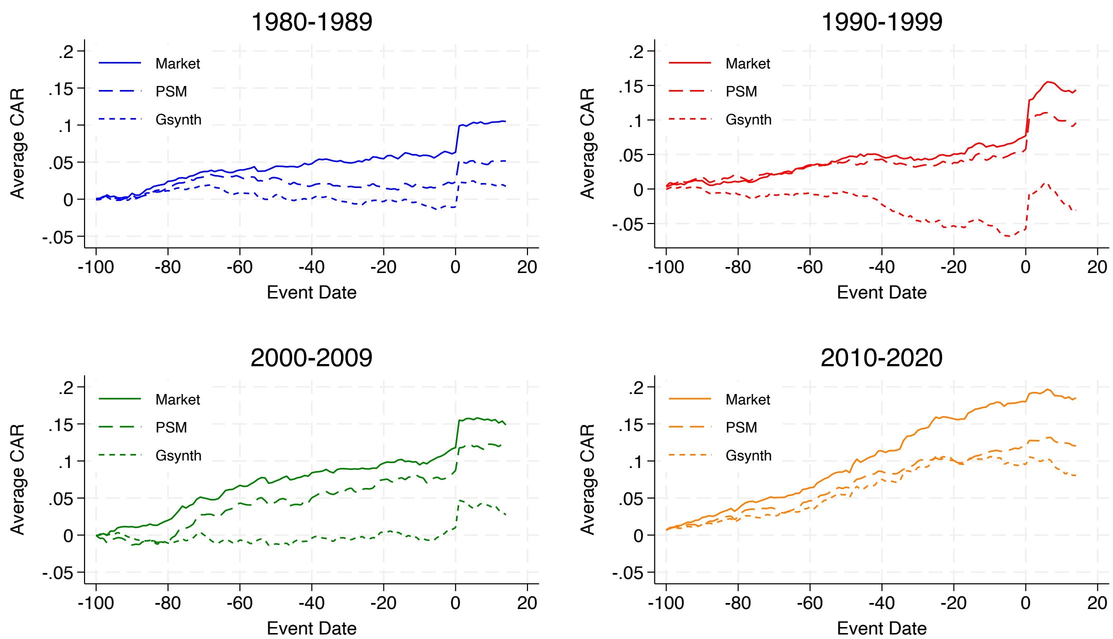
Next, for each announcement date, we estimate a Gsynth model (Xu & Liu, 2022) using 250 to 101 days prior to the event as the estimation period. This allows us to construct an average portfolio return for the *included* stocks, and also a synthetic control portfolio over the periods 100 to 15 days after.

Combined with both sets of counterfactual returns ("pseudo-included" and synthetic controls), we can compare the cumulative return of the included firms, before and after the event, starting 100 days priors to the inclusion event. We consider three measures: the market-adjusted abnormal return measure (as used in Greenwood and Sammon (2025)), the cumulative gap between the included stocks and the "pseudo-included" stocks, and the cumulative gap between the included stocks and the synthetic portfolio constructed by the Gsynth method.

First, we find that across all four decade, the pre-announcement drift as estimated by either the propensity score matched difference, or by the Gsynth approach drops significantly when compared to the market adjusted method. In fact, in several decades, there is almost *no* evidence of a pre-inclusion effect (1980s and 2000s). However, in the most recent decade, this adjustment only reduces the pre-announcment returns by half.

The effectiveness of Gsynth is quite striking in this setting, and suggests that longer-run cumulative effects can be substantially biased. What can explain the differences identified between these estimated methods? In Figure 6, we plot the average cumulative return for the market and SMB factor over this period, for both the event timing and random non-event dates in this period. On average, there is a substantial drift across most decades. Considering the positive loadings in Table 4, this suggests that the counterfactual return needs to sufficiently account for any and all potential unobserved factors driving expected returns to avoid this bias highlighted in Proposition 1. In future work, we turn to other settings where the long-run

Figure 5: **Cumulative abnormal pre-addition returns (market model, PSM, and Gsynth)** This figure plots the average cumulative abnormal returns following index inclusion announcements in event time, averaged across inclusions for each decade. We use several definitions of abnormal returns with different counterfactual returns. Solid lines plot abnormal returns with S&P 500 market returns, dashed lines plot abnormal returns with a propensity-score-matched counterfactual firm on lagged market cap rank, and dotted lines plot abnormal returns with synthetic portfolios from the generalized synthetic method (Xu & Liu, 2022). The returns are normalized to start at zero, 100-trading days before the announcement.

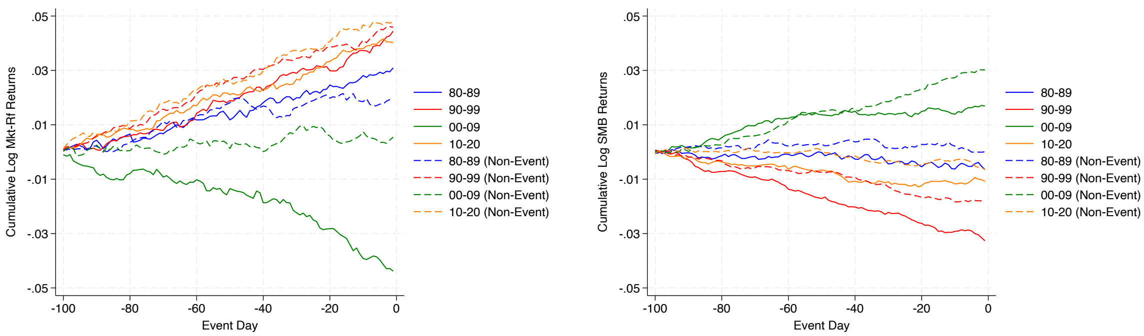


effects after the event matter economically.

Figure 6: **Pre-addition Cumulative Market Factor Returns (Inclusion vs. Randomized No-Inclusion Days)** This figure plots the average cumulative returns on the market and the SMB factor following index inclusion announcements in event time, averaged across inclusions for each decade. We also plot the average cumulative returns on the market following randomized no-inclusion days. For each inclusion date, we pick a random date on no-inclusion dates. The returns are normalized to start at zero, 100-trading days before the announcement.

Panel A: S&P 500 Cumulative Returns

Panel B: SMB Cumulative Returns



7 Conclusion

This paper brings modern causal inference techniques to financial event studies, highlighting important limitations in standard approaches while providing constructive solutions. We demonstrate that traditional abnormal return estimators face inconsistency problems due to factor model misspecification—a concern that becomes particularly severe in long-horizon analyses where small daily biases accumulate substantially over time.

While staggered event timing helps mitigate these issues in short-horizon studies by averaging out factor realizations, this solution proves inadequate for long-horizon analyses. The key insight is that misspecification bias compounds over longer horizons, regardless of how events are distributed across time.

Synthetic control methods offer a promising alternative by directly modeling counterfactual security paths without requiring correct specification of the underlying factor structure. Our empirical applications to political connections during market turbulence and S&P 500 index inclusions convincingly demonstrate the practical value of these methods.

Our findings suggest that many influential results based on long-horizon event studies may reflect factor model misspecification rather than genuine causal effects. We recommend that researchers employ synthetic control methods as a robust complement to traditional approaches, particularly when studying extended price responses or when events occur during periods of high market volatility.

In future versions of this work, we plan to examine inference procedures as well, as the synthetic control methods provide an alternative approach to estimating standard errors. We also plan to extend our analysis to other event studies, such as mergers and acquisitions, where the long-horizon bias is particularly pronounced.

References

Abadie, A., & Cattaneo, M. D. (2021). Introduction to the special section on synthetic control methods.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, *105*(490), 493–505.

Abadie, A., & L'hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, *116*(536), 1817–1834.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., & Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, *121*(2), 368–391.

Andrews, D. W. (2005). Cross-section regression with common shocks. *Econometrica*, *73*(5), 1551–1585.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118.

Baker, A., Gelbach, J. B., et al. (2020). Machine learning and predicted returns for event studies in securities litigation. *Journal of Law, Finance, and Accounting*, *5*(2), 231–272.

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of financial economics*, *75*(2), 283–317.

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789–1803.

Ben-Michael, E., Feller, A., & Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *84*(2), 351–381.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting research*, *27*, 1–36.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*(1), 3–31. https://doi.org/https://doi.org/10.1016/0304-405X(85)90042-X

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Campbell, J., Lo, A., & MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton University Press. https://books.google.com/books?id=lkeKhnqUHx8C

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis on large asset markets. *Econometrica*, *51*(5).

Connor, G. (1984). A unified beta pricing theory. *Journal of Economic Theory*, *34*(1), 13–31. https://doi.org/https://doi.org/10.1016/0022-0531(84)90159-5

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, *53*(6), 1839–1885.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review*, *110*(9), 2964–2996.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International economic review*, *10*(1), 1–21.

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association*, *116*(536), 1764–1772.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, *225*(2), 254–277.

Greenwood, R., & Sammon, M. (2025). The disappearing index effect. *The Journal of Finance*, *80*(2), 657–698. https://doi.org/https://doi.org/10.1111/jofi.13410

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Kothari, S. P., & Warner, J. B. (2007). Econometrics of event studies. In *Handbook of empirical corporate finance* (pp. 3–36). Elsevier.

Kwon, S. Y., & Tang, J. (2022). Extreme events and overreaction to news.

Prabhala, N. R. (1997). Conditional methods in event studies and an equilibrium justification for standard event-study procedures. *The Review of Financial Studies*, *10*(1), 1–38.

Raghavendra Rau, P., & Vermaelen, T. (1998). Glamour, value and the post-acquisition performance of acquiring firms. *Journal of Financial Economics*, *49*(2), 223–253. https://doi.org/https://doi.org/10.1016/S0304-405X(98)00023-3

Savor, P. G., & Lu, Q. (2009). Do stock mergers create value for acquirers? *The journal of finance*, *64*(3), 1061–1097.

Schwert, G. W. (1996). Markup pricing in mergers and acquisitions. *Journal of Financial economics*, *41*(2), 153–192.

Shleifer, A., & Vishny, R. W. (2003). Stock market driven acquisitions. *Journal of financial Economics*, *70*(3), 295–311.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, *225*(2), 175–199.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*(1), 57–76.

Xu, Y., & Liu, L. (2022). *Gsynth: Generalized synthetic control method* [R package version 1.2.1]. https://yiqingxu.org/packages/gsynth/

A Additional Simulation Results

For the simulation sample where treatment is selected based on loading to the second factor and random timing, We plot the bias from difference in mean, CAPM, and Gsynth estimators, across simulation samples.

Figure A.1: Bias from Difference-in-Mean Model on SMB Returns with Assignment Selection

This figure plots the biases from a difference-in-mean estimator on the treatment period over realizations of the second factor across 50 simulations. We simulate 500 firms with 10% of them getting treated. The estimation period is 239 days and post-event period is 11 days. More details on the simulations is in Section 5.1. Panel A reports simulation results with no selections, Panel B with only assignment selection, Panel C with only timing selection, and Panel D with both. We consider several estimators: difference in simple average, CAPM and 2-factor abnormal returns, and generalized synthetic methods. The expected biases and coverage are from 50 simulations.

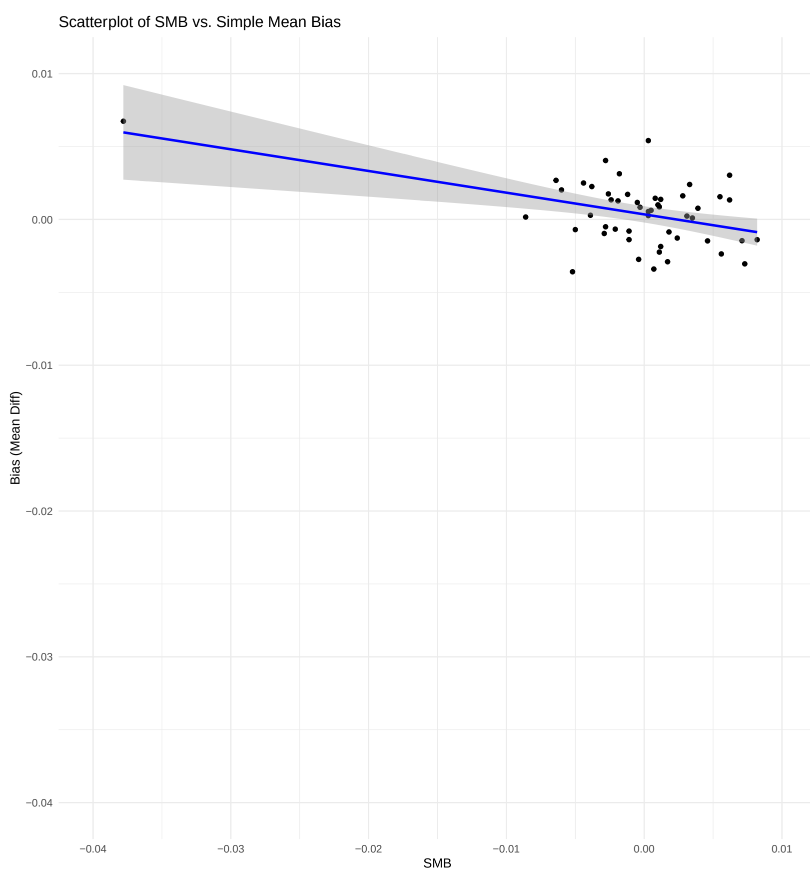
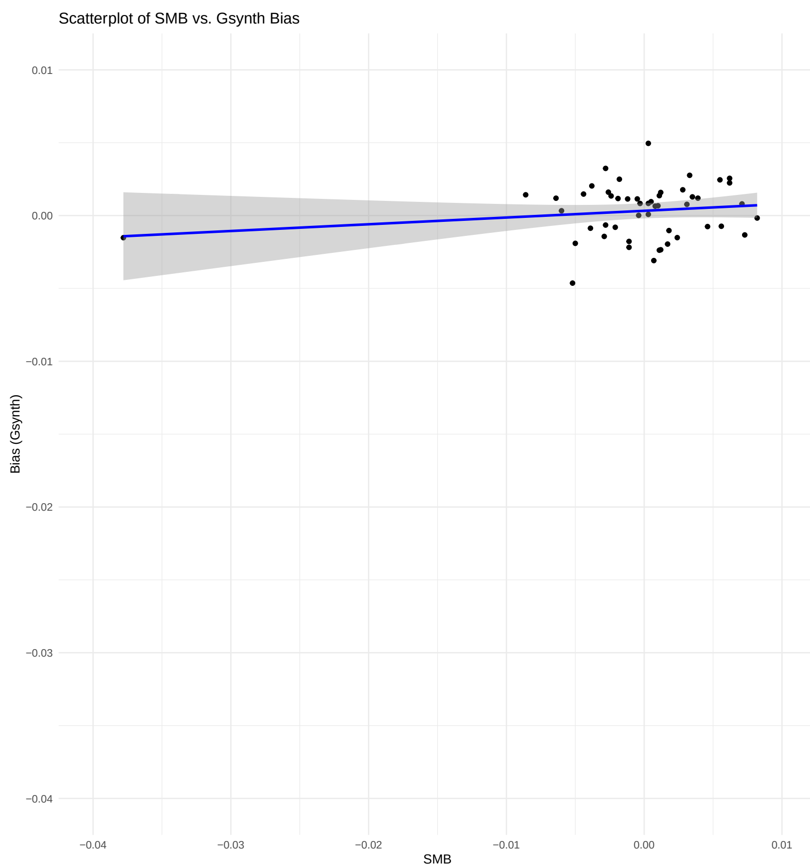


Figure A.2: Bias from Gsynth Model on SMB Returns with Assignment Selection

This figure plots the biases from a Gsynth estimator on the treatment period over realizations of the second factor across 50 simulations. We simulate 500 firms with 10% of them getting treated. The estimation period is 239 days and post-event period is 11 days. More details on the simulations is in Section 5.1. Panel A reports simulation results with no selections, Panel B with only assignment selection, Panel C with only timing selection, and Panel D with both. We consider several estimators: difference in simple average, CAPM and 2-factor abnormal returns, and generalized synthetic methods. The expected biases and coverage are from 50 simulations.



B Additional Results for Geithner

B.1 Period-by-Period ATT

In this section, we compare how different counterfactual affects the daily ATT in the post-event period. The 'Average' column computes the difference in the simple mean of treated versus control firms, as reported in Panel A of Table 2 in the original paper. The 'Synthetic

Control' column computes the weighted average daily return with synthetic control weights, and the 'Synthetic Diff-in-Diff' column uses the synthetic diff-in-diff weights instead.

For standard errors, in the 'Average' column, we use the same approach as the original paper and adjust the standard errors for pre-event correlation between firms. In the 'Synthetic Control' column, we report bootstrap standard errors estimated separately for each period. Since synthetic control weights will not change with the post-period, this method gives the correct standard errors period-by-period. We cannot use the same methodology for synthetic difference-in-differences because the estimated unit weight also depends on the data from the post-period.

We see that with synthetic control weights, the estimated ATT is much smaller compared to the simple mean.

Table B.1: Period-by-Period ATT to Geithner Announcement (Schedule connections)

		Average			Synthetic Control			Synthetic Diff-in-Diff		
Event day	Date	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference
0	11/21/08	0.086	0.042	0.043***	0.086	0.066	0.019*	0.086	0.058	0.028
1	11/24/08	0.130	0.046	0.084***	0.130	0.080	0.050**	0.130	0.063	0.067
2	11/25/08	0.026	0.015	0.011	0.026	0.045	-0.019	0.026	0.018	0.008
3	11/26/08	0.112	0.041	0.071***	0.112	0.070	0.042	0.112	0.055	0.057
4	11/28/08	0.056	0.018	0.038**	0.056	0.028	0.027	0.056	0.025	0.030
5	12/1/08	-0.131	-0.076	-0.056***	-0.131	-0.119	-0.013	-0.131	-0.102	-0.030
6	12/2/08	0.046	0.043	0.003	0.046	0.039	0.007	0.046	0.056	-0.010
7	12/3/08	0.034	0.018	0.016	0.034	0.035	-0.001	0.034	0.024	0.011
8	12/4/08	-0.009	-0.013	0.005	-0.009	-0.028	0.019	-0.009	-0.016	0.008
9	12/5/08	0.063	0.024	0.038**	0.063	0.034	0.028**	0.063	0.031	0.031
10	12/8/08	0.064	0.027	0.037**	0.064	0.047	0.017	0.064	0.033	0.031

Table B.2: Period-by-Period ATT to Geithner Announcement (Personal connections)

		Average			Synthetic Control			Synthetic Diff-in-Diff		
Event day	Date	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference
0	11/21/08	0.075	0.043	0.033	0.075	0.073	0.003	0.075	0.069	0.007
1	11/24/08	0.143	0.047	0.096***	0.143	0.106	0.037	0.143	0.074	0.069
2	11/25/08	0.057	0.014	0.043*	0.057	0.059	-0.002	0.057	0.023	0.034
3	11/26/08	0.112	0.042	0.071***	0.112	0.113	0.000	0.112	0.070	0.042
4	11/28/08	0.085	0.018	0.067***	0.085	0.077	0.008	0.085	0.031	0.054
5	12/1/08	-0.144	-0.076	-0.067***	-0.144	-0.140	-0.004	-0.144	-0.121	-0.023
6	12/2/08	0.044	0.043	0.001	0.044	0.063	-0.019	0.044	0.066	-0.022
7	12/3/08	0.043	0.018	0.024	0.043	0.033	0.010	0.043	0.025	0.017
8	12/4/08	0.005	-0.014	0.019	0.005	-0.024	0.029	0.005	-0.015	0.020
9	12/5/08	0.042	0.025	0.017	0.042	0.046	-0.004	0.042	0.039	0.003
10	12/8/08	0.043	0.028	0.015	0.043	0.055	-0.012	0.043	0.042	0.002

Table B.3: Period-by-Period ATT to Geithner Announcement (New York connections)

		Average			Synthetic Control			Synthetic Diff-in-Diff		
Event day	Date	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference	Conn.	Non-conn.	Difference
0	11/21/08	0.085	0.040	0.044***	0.085	0.069	0.016*	0.085	0.051	0.033
1	11/24/08	0.078	0.046	0.031***	0.078	0.082	-0.004	0.078	0.058	0.020
2	11/25/08	0.032	0.014	0.018	0.032	0.011	0.021*	0.032	0.016	0.016
3	11/26/08	0.087	0.040	0.048***	0.087	0.065	0.022	0.087	0.048	0.040
4	11/28/08	0.016	0.019	-0.003	0.016	0.023	-0.006	0.016	0.022	-0.005
5	12/1/08	-0.105	-0.075	-0.030***	-0.105	-0.106	0.001	-0.105	-0.093	-0.012
6	12/2/08	0.090	0.040	0.050***	0.090	0.052	0.037***	0.090	0.050	0.039
7	12/3/08	0.031	0.018	0.013	0.031	0.025	0.005	0.031	0.021	0.009
8	12/4/08	-0.020	-0.013	-0.008	-0.020	-0.031	0.010	-0.020	-0.014	-0.006
9	12/5/08	0.050	0.024	0.026**	0.050	0.046	0.004	0.050	0.029	0.021
10	12/8/08	0.050	0.027	0.023**	0.050	0.055	-0.006	0.050	0.031	0.018

B.2 Placebo Period ATT

Table B.4: Placebo Period ATT to Geithner Announcement (Schedule connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.006*	-0.006**	-0.004	-0.003
	(0.004)	(0.003)	(0.003)	(0.003)
Observations	16,350	139,520	139,520	139,520

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table B.5: Placebo Period ATT to Geithner Announcement (Personal connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.007	-0.006**	0.001	-0.002
	(0.005)	(0.002)	(0.003)	(0.003)
Observations	16,350	139,520	139,520	139,520

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table B.6: Placebo Period ATT to Geithner Announcement (New York connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.003	-0.002	-0.000	-0.000
	(0.002)	(0.001)	(0.001)	(0.001)
Observations	16,350	139,520	139,520	139,520

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

B.3 Placebo Period

In this section, we test how synthetic methods perform in a placebo period before the event. The placebo period is day -30 to day -1, which is not used in estimation but also the event is not yet happening. If we assume that synthetic methods perform well in capturing the underlying factor structure and the factor loadings stay stable before the event, we would expect that the ATT in the placebo period is close to 0.

Figure B.1 plots the average treatment effect of raw returns on the left and the average treatment effect of abnormal returns (relative to a CAPM model with beta estimated using daily returns from day -280 to -31). In Figure B.2, we plot all the ATT on one graph for better comparison.

We see that synthetic control does the best job in the placebo period, but also has the least treatment effect post-period. By comparing the treatment effect of raw returns using synthetic controls with the treatment effect of abnormal returns with a simple average, we see that they are relatively close, which suggests that synthetic control does a good job matching the underlying market beta exposure of treatment firms.

Figure B.1: Period-by-Period ATT in Placebo and Post Period (Schedule connections)

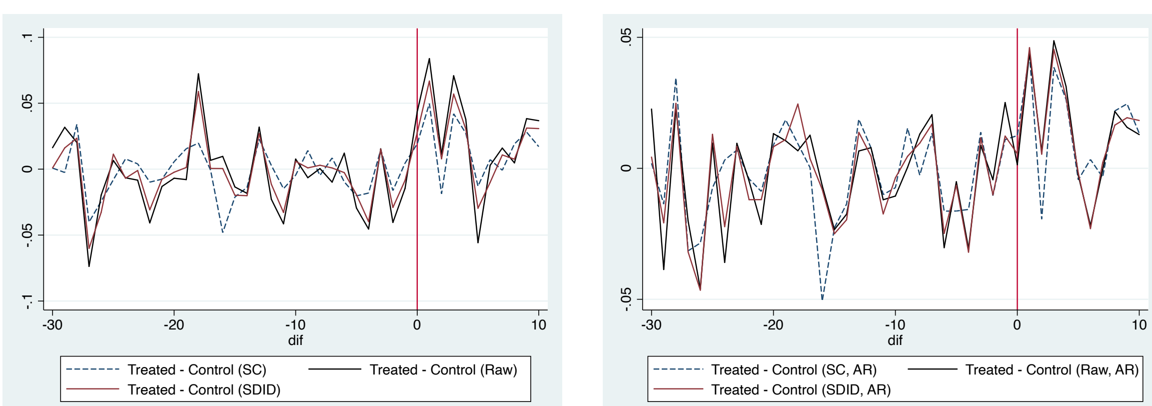
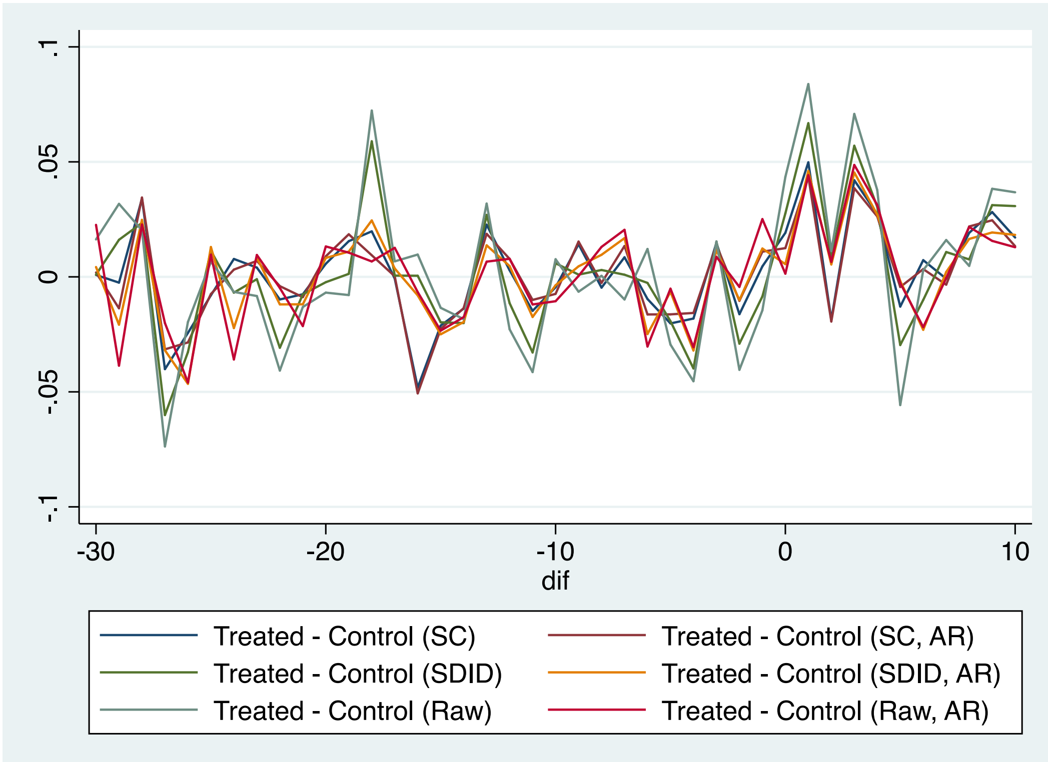


Figure B.2: Period-by-Period ATT in Placebo and Post Period (Schedule connections)



B.4 Pre- versus Post-Event Beta and Weights

In this section, we investigate how control beta is compared to treatment beta with different control weights. We also compare beta estimated pre-event with beta estimated post-event to see if the event also has a treatment effect on beta loadings. The pre-event beta is estimated over daily returns from day -280 to day -31, and the post-event beta is estimated over daily returns from day 31 to day 65. We exclude the immediate post-period because the returns can be confounded by the event effect. We also compare the synthetic weights estimated with pre- and post-period by comparing the treatment effect with pre- and post-weights.

First, we see that indeed synthetic control weights match control beta to treatment beta the best, compared to a simple average and synthetic diff-in-diff weights. For the pre-event, we see a control beta of 1.33 with synthetic control, compared to a treatment beta of 1.43. For the post-event, we have a control beta of 1.71, which is very close to a treatment beta of 1.73. The same conclusion can be drawn with Fama-French three-factor betas. Synthetic control weights give the closest control betas to treatment betas for market, size, and value factors.

Second, we see that post-betas are on average higher than pre-betas, suggesting that the event does have an effect on the underlying factor loadings of treatment firms. CAPM market beta increases from 1.43 to 1.73, a 21% increase. In the three-factor model, we see the largest increase in size and value betas. Size beta increases from 0.23 to 0.41 (78%), and value beta increases from 0.61 to 1.00 (64%).

Third, Figure B.3 show the daily ATT with synthetic control weights for the placebo period (day -30 to -1), post-event period (day 0 to 30), and post-event-estimation period (day 31 to 65). We see that using post-event synthetic control weights gives us a larger event treatment effect, but it also gives a more positive ATT in the placebo period.

Table B.7: Pre-/Post-Event Market Beta from CAPM

Panel A: Pre Beta, Pre Weights		
	Market	
	Treated	Control
Average	1.4269	0.8251
SDID	1.4269	1.1111
SC	1.4269	1.3309

Panel B: Post Beta, Post Weights		
	Market	
	Treated	Control
Average	1.7304	0.9377
SDID	1.7304	1.4076
SC	1.7304	1.7083

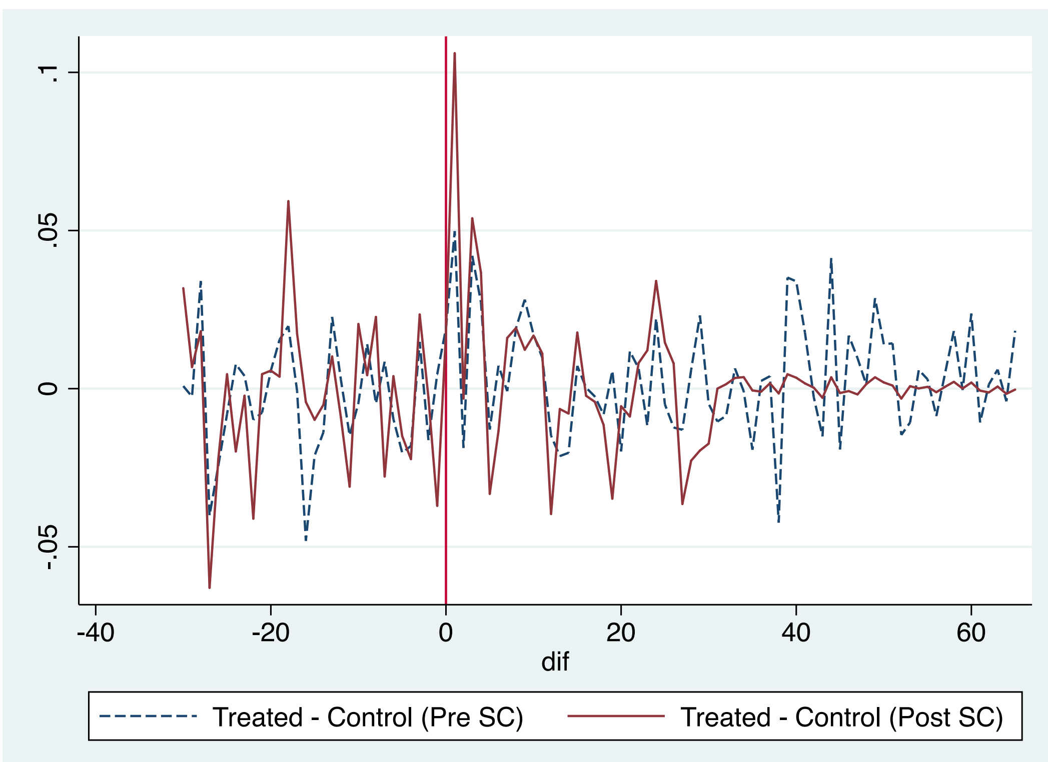
Panel C: Pre Beta, Post Weights		
	Market	
	Treated	Control
Average	1.4269	0.8251
SDID	1.4269	1.0954
SC	1.4269	1.0751

Panel D: Post Beta, Pre Weights		
	Market	
	Treated	Control
Average	1.7304	0.9377
SDID	1.7304	1.2105
SC	1.7304	1.3664

Table B.8: Pre-/Post-Event Beta from Fama-French Three Factors

Panel A: Pre Beta, Pre Weights						
	Market		SMB		HML	
	Treated	Control	Treated	Control	Treated	Control
Original	1.2748	0.6592	0.2330	0.7484	0.6068	0.7196
SDID	1.2748	0.9051	0.2330	0.8187	0.6068	0.8724
SC	1.2748	1.1477	0.2330	0.4796	0.6068	0.7495
Panel B: Post Beta, Post Weights						
	Market		SMB		HML	
	Treated	Control	Treated	Control	Treated	Control
Original	1.2454	0.6265	0.4139	0.5633	0.9991	0.6898
SDID	1.2454	0.9544	0.4139	0.6791	0.9991	0.9785
SC	1.2454	1.2130	0.4139	0.4697	0.9991	1.0273

Figure B.3: Period-by-Period ATT with Pre & Post SC Weights



B.5 Beta: All Public Firms as Control

Table B.9: Pre-Event Market Beta from CAPM

Panel A: Pre Beta, Pre Weights		
	Market	
	Treated	Control
Average	1.4269	0.8324
SDID	1.4269	1.2814
SC	1.4269	1.3830

Table B.10: Pre-Event Beta from Fama-French Three Factors

Panel A: Pre Beta, Pre Weights						
	Market		SMB		HML	
	Treated	Control	Treated	Control	Treated	Control
Original	1.2748	0.8569	0.2330	0.5526	0.6068	0.1436
SDID	1.2748	1.1654	0.2330	0.6273	0.6068	0.5934
SC	1.2748	1.2201	0.2330	0.3774	0.6068	0.6743

B.6 Placebo Period ATT: All Public Firms as Control

Table B.11: Placebo Period ATT to Geithner Announcement (Schedule connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.003	-0.003	-0.004	-0.002
	(0.004)	(0.002)	(0.003)	(0.002)
Observations	122,850	1,044,225	1,044,225	1,044,225

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table B.12: Placebo Period ATT to Geithner Announcement (Personal connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.004	-0.003	-0.001	-0.002
	(0.006)	(0.004)	(0.003)	(0.003)
Observations	122,850	1,044,225	1,044,225	1,044,225

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table B.13: Placebo Period ATT to Geithner Announcement (New York connections)

	(1)	(2)	(3)	(4)
	Average	DID	SC	SDID
Treated	-0.000	0.000	-0.002	0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Observations	122,850	1,044,225	1,044,225	1,044,225

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

C Additional Results for Index Inclusion

Figure C.1: Cumulative pre-addition market-adjusted returns (Treated vs. propensity score matched)

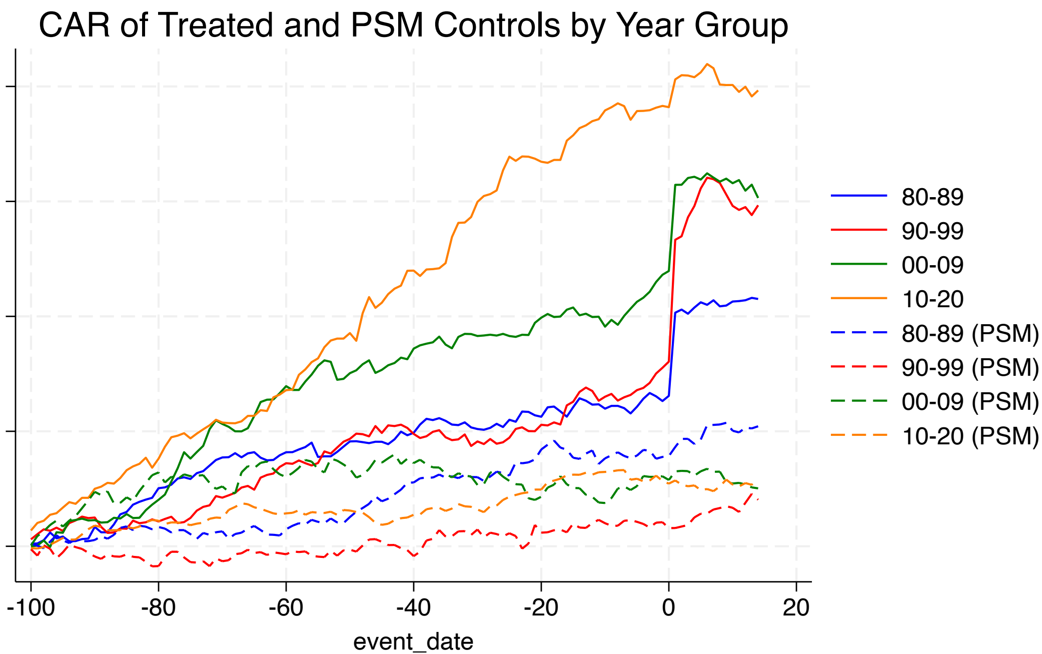


Figure C.2: Cumulative pre-addition market-adjusted returns (Treated vs. synthetic method)