Causal Inference in Financial Event Studies*

Paul Goldsmith-Pinkham
Yale University & NBER

Tianshu Lyu
Yale University

May 6, 2026

Abstract

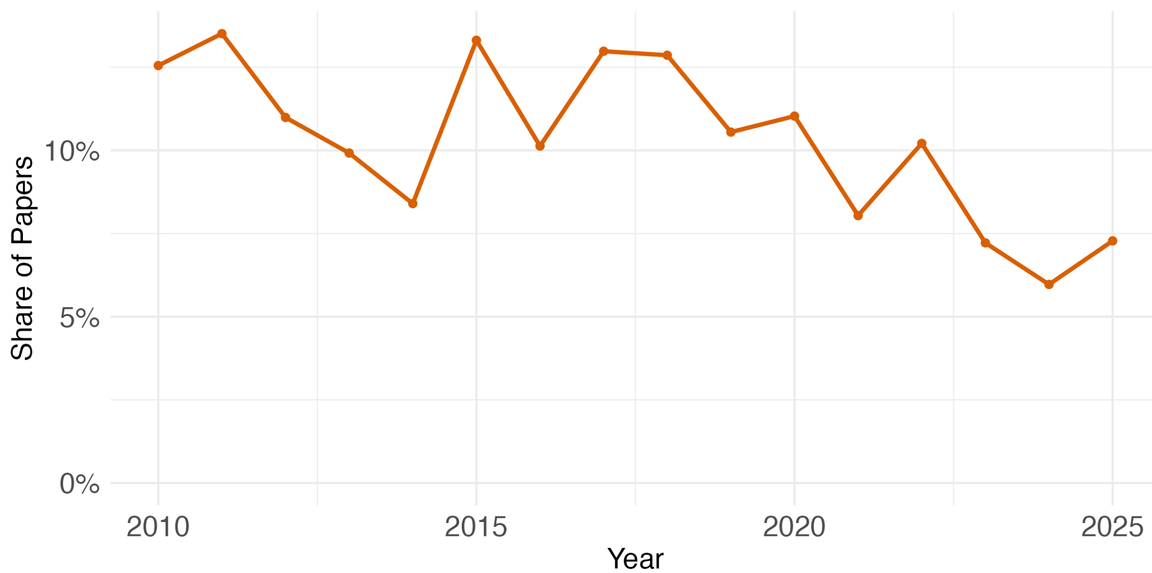
Financial event studies estimate causal effects by constructing counterfactual returns using asset pricing factor models. By design, these factor models are designed to capture *priced* risk—factors that carry risk premia and explain the cross-section of expected returns. But constructing a valid counterfactual return requires accounting for *all* systematic variation in returns, including factors that carry no risk premium. We show that the gap between these two objects is inconsequential under conditions common in classic applications: many randomly-timed events and stationary factor distributions, so that unpriced factor realizations average out and the distribution of priced factor realizations is representative. When these conditions fail—as with a single event date, event timing that coincides with unusual market conditions, or long horizons with shifting factor distributions—traditional estimators can produce substantial bias. We derive precise identification conditions and analytic bias expressions, and propose synthetic control methods that match on realized pre-event return paths, implicitly capturing exposure to both priced and unpriced factors. Revisiting four empirical applications, we show that some established findings—the Geithner Treasury Secretary announcement effect on banks' stock prices (Acemoglu et al., 2016), pre-inclusion drift for index inclusion, and M&A acquirer effects—may reflect unmeasured systematic risk rather than true treatment effects.

*Contact: paul.goldsmith-pinkham@yale.edu We thank Nick Barberis, Stefano Giglio and Will Goetzmann for helpful discussions, and audiences at the NBER Summer Institute Forecasting & Empirical Methods session, SEA, Dallas Fed, Kellogg Finance department, and SMU Statistics department.

1 Introduction

Financial economists were practicing causal inference well before the credibility revolution (Angrist & Pischke, 2010). By examining how asset prices respond to information events—such as merger announcements, earnings releases, or regulatory changes—financial event studies compare the returns of treated assets to benchmark comparison asset returns. The approach remains central: between 2011 and 2025, 577 articles in the Journal of Finance, Journal of Financial Economics, and the Review of Financial Studies reference event-study methods (Figure 1).

Figure 1: **Prevalence of financial event studies in finance journals:** This figure plots the share of articles by year that mention the word "cumulative abnormal returns" OR "announcement returns" in the Journal of Finance or Review of Financial Studies. Source: https://paulgp.com/econlit-pipeline/search.html



While financial event studies target causal effects as their estimands, the suite of estimators used in financial event studies are antiquated relative to the many tools available. The textbook approach, starting as early as Fama et al. (1969) and canonized in the Campbell et al. (1997) textbook, relies on linear factor models with known factors to construct counterfactual returns, i.e. what a security's return would have been absent the event. Researchers typically estimate a security's exposure to market factors during a pre-event window, then use these estimated loadings to predict what returns during the event window. The difference between actual and predicted returns constitutes the abnormal return.

This paper identifies a fundamental tension between what causal event studies require and what asset pricing factor models provide. Factor models used in event studies—from the market model to multi-factor specifications—are designed to capture *priced* risk: systematic factors that earn risk premia and explain why some securities have higher expected returns than others. A correctly specified expected return model may omit factors that carry no risk premium, because such factors are irrelevant for pricing the cross-section. But constructing a counterfactual return on a specific event date is a different task. It requires accounting for *all* systematic factors that drive return variation on that date—whether or not those factors carry risk premia. An unpriced factor (say, an industry shock) contributes nothing to expected returns but can generate large return variation on any given day. If treated firms load differently on such a factor than the model-implied counterfactual, the result is bias, even though the expected return model is "correct" in the asset pricing sense.

This problem is compounded by a second channel: even for priced factors that *are* included in the model, the standard approach works by estimating an intercept and factor loadings in a pre-event window and projecting forward. The intercept absorbs the average risk premium—the *unconditional* mean of the factor. But if events systematically occur during unusual market conditions—crises, booms, periods of elevated volatility—then the factor realizations during event windows are drawn from a *conditional* distribution that may look very different from the unconditional one.

The key insight of this paper is that these two sources of bias—unpriced systematic risk and non-representative event timing—both vanish under conditions that hold in many classic applications. When many events occur at random times and the factor distribution is stationary, two things happen simultaneously. First, unpriced factors, which have zero mean by definition, average out across event dates via the law of large numbers. Second, random event timing ensures that the distribution of priced factor realizations during event windows is representative of the unconditional distribution, so the intercept estimated in the pre-event window correctly absorbs the average risk premium. Together, these imply that the choice of factor model is largely irrelevant—*any* expected return model, even one that ignores factors entirely, will produce approximately unbiased estimates. This is precisely why the folk wisdom that "the model doesn't matter" has held up so well empirically. For example, footnote 5 of Shleifer (1986) states "The [index inclusion] results were not materially different when returns were not corrected for market movements." This irrelevance follows from two key features: (1) random timing of many events over time (in the Shleifer (1986) case, index inclusions) and (2) very short-run estimates such that the treatment effect dominates any shifts in the distribution of the risk premium.

However, the coincidence breaks down in three empirically relevant scenarios—each activating one or both channels. First, when there is a single event date (or a small number of clustered events), there is no averaging: both unpriced and priced factor realizations on that date enter the bias directly. This is especially severe during periods of extreme market volatility, when unpriced factors can be enormous, and even correctly modeled priced factors take extreme values far from their unconditional means. Second, when event timing correlates with market conditions—for instance, if mergers cluster during booms or crisis-related events occur during high-volatility regimes—the factor realizations during event windows are systematically non-representative. Even unpriced factors, which have zero *unconditional* mean, may have nonzero *conditional* mean when conditioned on the market state that triggers events. For priced factors, the average risk premium observed during event windows may differ substantially from the unconditional premium absorbed by the OLS intercept. Third, in long-horizon event studies, the factor distribution may shift between the pre-event estimation window and the post-event window, so that the intercept estimated in one regime produces biased counterfactuals in another—a form of non-representative timing that accumulates over longer horizons.

We provide precise conditions under which traditional event study methods identify causal effects and derive analytic expressions for the asymptotic bias when these conditions fail. Identification of the average treatment effect on the treated requires either correct specification of the factor model (including all systematic factors, priced or not), random assignment of treatment across securities, or random assignment of many events across time with a stationary factor distribution. Our results clarify *why* the folk wisdom holds when it does and characterize precisely when and how it breaks down.

We contrast three types of estimators that can be used for financial event studies, and compare their properties: (1) classic abnormal return estimators, based on specified factors, (2) difference-in-mean estimators, which construct control groups through decisions of the econometrician, and (3) synthetic estimators (Abadie & Gardeazabal, 2003; Abadie et al., 2010), which use historical prices from control assets to construct either a replicating portfolio (synthetic control) or to construct a set of factors using PCA (Xu, 2017). The key advantage of synthetic methods is that by matching on the *realized* pre-event return path of treated securities, they implicitly match on exposure to *all* systematic factors—both priced and unpriced—without requiring the researcher to specify which factors matter or whether they carry risk premia. A portfolio that tracks the treated group's actual returns day-by-day must share similar loadings on whatever common factors drove those returns, regardless of whether those factors appear in any asset pricing model. If such a replicating portfolio exists, it should provide valid counterfactual returns in the post-event period without requiring correct specification of the underlying factor model.

Our theoretical results hinge on the assumption that the expected return for a cohort of treated stocks follows an unknown time-invariant linear factor model. This assumption is not innocuous, and likely not true for all time periods. But, it is also a weaker assumption than the traditional abnormal return estimators. Any approach that uses a model to infer the counterfactual outcomes for the treated stocks will require some kind of model stability assumption (without additional structure like Kelly et al. (2019)). We view it as valuable future work to see if other more robust asset pricing models can be used to generate counterfactual returns, such as Giglio et al. (2025), Kelly et al. (2019), and Daniel et al. (2020).

Our work focuses on the case of having multiple treated firms. In many settings, especially in law, single treated firms are the focus on study. However, this estimand is notoriously challenging to estimate and raises further inference issues (see Baker, Gelbach, et al. (2020) for a discussion of this setting and a simulation assessment of some of the synthetic control method we propose here). As we show in this paper, the use of many events and many firms allows for estimands that can be estimated under much weaker parametric assumptions on the idiosyncratic error of the return.

One key benefit of focusing carefully on the estimand of interest is that we are able to show that buy-and-hold abnormal return estimates are particularly challenging to estimate because they require the matching portfolio to not just match on expected returns, but also on volatility. If the control group's returns have different variance, then the differing volatility drag will lead to very different results. To make this concrete: imagine that there is *no* treatment effect, but a diversified portfolio is used as a control group for a stock, both with equal expected returns. The lower variance for the diversified portfolio will lead to a *negative* treatment effect from a buy-and-hold perspective, despite no actual treatment effect. This implies that doing buy-and-hold abnormal returns with an index can be seriously flawed.

We also assess theoretically the calendar-time portfolio approach advocated by Fama (1998) and others as an alternative to abnormal return estimation. We show that the usual estimator can be written as a reweighted version of the standard cumulative abnormal return estimator, where the difference in weightings reflects the differences in the event-timings across periods, and the path of treatment effects. Under relatively balanced distribution of events, the two estimators coincide.

Our theoretical results connect to a substantial econometric literature concerned with inference in event studies, but reveal that these inference problems are symptoms of a deeper identification failure.[1] For example, Boehmer et al. (1991) show that event-induced variance increases cause standard tests to over-reject, while Kolari and Pynnönen (2010) demonstrate that even low cross-correlation among abnormal returns leads to severe over-rejection when event dates cluster. Cohn et al. (2025) find that cross-sectional regressions of returns on firm

[1] An non-exhaustive list of papers considering alternative inference approaches include Corrado (1989), Corrado and Zivney (1992), Kramer (2001), Barber and Lyon (1997b), Bernard (1987), Kolari and Pynnönen (2010), Brown and Warner (1985b), and Cohn et al. (2025).

characteristics reject at the 1% level on over 20% of non-event days, suggesting the bar for rejection is far too low.

Our framework clarifies why these problems arise: they are all manifestations of unmeasured systematic risk on event dates. With a single event date, the estimator converges not to the true treatment effect but to a quantity that depends on the particular realization of *all* systematic factors—priced and unpriced—on that day. No expected-return model can account for the unpriced component, and even correctly modeled priced factors take a single draw that may be far from their mean. The cross-sectional correlation documented by Kolari and Pynnönen (2010) reflects this common factor exposure—all firms on the same event day share the same factor draws (priced and unpriced alike), so their estimation errors are mechanically correlated. The event-induced variance of Boehmer et al. (1991) arises because volatile periods feature large realizations of both priced and unpriced factors, amplifying the unmodeled systematic component. And the spurious cross-sectional patterns in Cohn et al. (2025) occur because factor realizations vary across days, creating characteristic-return relationships even absent any true effect—firms that load heavily on an unpriced factor will appear to have abnormal returns on days when that factor takes large values.

The inference adjustments proposed in this literature—modified test statistics, clustered standard errors, or benchmarking against non-event days—appropriately widen confidence intervals to reflect uncertainty about the factor realization. But they cannot eliminate the bias itself; they simply acknowledge that a single event date provides essentially one noisy observation. Our results propose a different tack: either multiple random event timings that allow factor terms to average out, or synthetic control methods that do not require correct factor specification.[2] While this literature has primarily documented inference problems through simulation studies, we provide analytic expressions for the asymptotic bias under misspecification, characterizing precisely when and why standard estimators fail to identify causal effects.

We revisit four empirical settings that span the range of typical applications. First, we reexamine the Acemoglu et al. (2016) study of political connections during the 2008 financial crisis, where the Treasury Secretary announcement coincided with extreme market volatility—daily returns exceeded 6% on multiple event days. The original estimates using simple averaging suggest economically large effects of political connections. Even abnormal return models using the Fama-French 3 factor model suggest economically meaningful effects. However, the estimates disappear when using our proposed synthetic methods, suggesting that model misspecification with a single event can create spurious results when events coincide

[2]The GLS approach prescribed in Cohn et al. (2025) has similarities to the synthetic control methods advocated in this paper, but are used primarily to improve power in cross-sectional tests, rather than improving point estimate estimation.

with volatile market conditions.

Next, we analyze S&P 500 index inclusions, and show that since the index inclusion events appear random across time, the effect of short-run model misspecification is non-existent, echoing the folk wisdom above. However, we show that the substantial pre-announcement drift, often pointed to as a source of possible index inclusion front-running, disappears once we properly account for the unobserved factor exposures of included firms. This finding suggests that what appears to be anticipation or momentum may actually reflect model misspecification.

Third, we examine the effect of acquistions in merger deals on acquiring firms with some studies finding large negative abnormal returns over several years.(Loughran & Vijh, 1997; Rau & Vermaelen, 1998) We demonstrate that these long-run patterns are highly sensitive to model specification, consistent with our theoretical prediction that misspecification bias accumulates over longer horizons.

Our last empirical result applies a version of LaLonde (1986) to our analysis by using quasi-experimental variation to provide a benchmark for our model-based approaches. The treatment and control groups for the baseline are found in close merger contests where multiple firms bid for the same target. Following Malmendier et al. (2018), contest losers provide a natural counterfactual for winners since they are ex ante similar firms competing for identical targets. The results are not supportive of abnormal return models at all, but only weakly support synthetic methods. These results suggest that for long-run analyses, it is far better to construct counterfactuals based on quasi-experimental variation than using model-based approaches.

These empirical findings have important implications for the interpretation of the vast event study literature in finance. Many influential results—particularly those involving long horizons, volatile periods, or systematic event timing—may reflect unpriced systematic risk or non-representative factor realizations rather than true treatment effects. However, we emphasize that our results do not invalidate the entire enterprise. For studies with plausibly random event timing and stationary factor distributions, unpriced factors average out and priced factors are well-captured by the OLS intercept, so traditional methods remain reliable—our empirical work confirms they produce similar estimates to more sophisticated approaches. The key insight is recognizing that event study counterfactuals require matching on *all* systematic risk, not just priced risk, and that this distinction matters most when event timing is non-random or when few events prevent averaging over factor realizations.

Our work connects several distinct literatures. Methodologically, we build on the econometrics of event studies in finance (MacKinlay, 1997; Kothari & Warner, 2007) while incorporating insights from the modern causal inference literature (Imbens & Rubin, 2015; Abadie & Cattaneo, 2021). We also contribute to the older debate about long-run event studies (Mitchell

& Stafford, 2000; Barber & Lyon, 1997a) by providing a formal framework for understanding when and why these studies are problematic.

2 Estimands and estimators in financial event studies

This section formalizes the setup of financial events on stock market returns in the language of potential outcomes. We begin by introducing the basic notation (Section 2.1), defining potential returns and treatment indicators for each security over time. We then specify the causal estimands of interest, 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.

2.1 Setup and notation

We study the causal effects of corporate events on security returns using a potential outcomes framework. Consider a panel of N securities indexed by $i = 1, 2, \ldots, N$ observed over T time periods indexed by $t = 1, 2, \ldots, T$.

2.1.1 Event Timing and Treatment Status

For each security i, let T_i denote the time when an event occurs:

$$T_i = \begin{cases} s & \text{if security } i \text{ experiences the event at time } s \\ \infty & \text{if security } i \text{ never experiences the event} \end{cases} \tag{1}$$

We denote the set of event times as $\mathcal{S} \subseteq \{1, \ldots, T\}$ and the set of never-treated (control) securities as $\mathcal{C} = \{i : T_i = \infty\}$. Following standard practice in event studies, we assume events are irreversible—once an event occurs (e.g., a merger announcement or earnings release), it cannot be undone.

2.1.2 Potential Outcomes Framework

Now, we define 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):

$$R_{i,t} = R_{i,t}(\infty) + \sum_{s \in \mathcal{S}} (R_{i,t}(s) - R_{i,t}(\infty)) 1(T_i = s). \tag{2}$$

2.1.3 Treatment Effects

We postulate that financial event studies are focused on identifying the difference between the *realized* returns for a treated firm ($R_{it}(s)$) versus the returns in the *absence* of the event. We define the difference in returns due to the event in period s for firm i in period t as the *individual treatment* or equivalently, the *abnormal firm return*:

Definition 1 (*Individual treatment effect / abnormal firm return*).

$$\tau_i(s,t) = \underbrace{R_{i,t}(s)}_{\substack{\textit{observed for}\\ \textit{treated firm}}} - \underbrace{R_{i,t}(\infty)}_{\substack{\textit{unobserved}\\ \textit{counterfactual}}} . \tag{3}$$

For a firm that has the event occur in period s, $R_{it}(s)$ is observed, and hence is identified. But, $R_{it}(\infty)$ is not. Indeed, in asset pricing, the challenge of modeling the exact return for an individual asset is viewed as a near-impossible task, even with a structural model. Instead, a large number of asset pricing papers focus on the challenge of estimating the *average* return for firms given a set of characteristics and/or risk factors (E.g. Chamberlain & Rothschild, 1983; Connor, 1984; Fama & French, 1993; Ross, 2013; Kelly et al., 2019; Bryzgalova et al., 2025).

This focus on expected returns makes causal inference and asset pricing models natural companions—but also reveals a subtle tension. The inability to know the exact counterfactual return is known as the *fundamental problem of causal inference* and leads to a focus on alternative estimators, often constructing *average* counterfactual returns for a group of treated units. However, there is a gap between what asset pricing models are designed to deliver and what event study counterfactuals require. Expected return models focus on *priced* risk factors—those that carry risk premia—because these are the factors that explain cross-sectional differences in average returns. Factors that carry no premium (e.g., certain liquidity or volatility shocks) are correctly excluded from such models. But a valid event study counterfactual must account for *all* systematic factors that move returns on a given date, whether or not they are priced. An unpriced factor contributes nothing to expected returns, yet can generate large return variation on any specific day. Moreover, even for priced factors, the counterfactual is only correct if the event-window factor realizations are representative of those in the estimation window—a condition that fails when event timing correlates with market conditions. As we show below, both sources of bias vanish when many randomly-timed events allow unpriced factors to average out (since they have zero mean) and ensure a representative draw of priced factor realizations. But the bias can be substantial when events are few, clustered, or coincide with extreme market conditions.

A significant body of empirical work and legal scholarship focuses on identifying the effect of events on single firms' valuations, since these valuations are used in litigation to estimate damages (Baker, Gelbach, et al., 2020). But our view is that in academic research studying financial event studies, a much more natural estimand to target is the *average* treatment effect on the treated (ATT), using many treated firms to estimate an overall average effect, rather than the effect on a single firm. We view the estimated abnormal returns for single firm events as case studies of a much more stable design that focuses on the *average* effect.

Definition 2 (*Cohort-Period Average Treatment Effect on the Treated (ATT)*). *Let the average treatment effect on returns in period t for firms treated in period s be*

$$\tau(s,t)^{ATT} = E(\tau_i(s,t) \mid T_i = s) = E(R_{i,t}(s) \ - \ R_{i,t}(\infty) \mid T_i = s). \tag{4}$$

This cohort-period ATT describes the effect of a treatment happening in period s during period t *for those firms who are experience the period s event.* If these firms are special in some way, then this may not be the same effect for other firms (for example, if these firms are riskier, and the effect differs by risk profile).

2.1.4 Event-Time Analysis

These cohort-period ATTs can be combined in a number of ways. Most crucially for our results, combining event cohorts to study effects relative to an event time will average across different event timings. The average treatment effect κ periods after an event is:

$$\theta_\kappa^{ATT} = \sum_{s \in \mathcal{S}} w_s \cdot \tau^{ATT}(s, s+\kappa) \tag{5}$$

where w_s represents the weight on event cohort s. A natural choice is $w_s = N_s / \sum_{s'} N_{s'}$, where N_s is the number of securities with $T_i = s$.

Many empirical papers studying these announcements are interested in cumulating the effects. The Cumulative Average Treatment Effect (CATT), analogous to cumulative abnormal returns (CAR), from event time 0 to H is:

$$\theta_H^{CATT} = \sum_{\kappa=0}^{H} \theta_\kappa^{ATT} \tag{6}$$

In this paper, we focus on these linear transformations of the ATT because they are well-behaved econometrically. However, an alternative approach to cumulative arithmetic returns is the buy-and-hold abnormal return, which we discuss briefly here to highlight its econometric challenges.

2.1.5 Geometric Returns and Buy-and-Hold Abnormal Returns

Announcement effects are often cumulated using *buy-and-hold* returns, which correspond to geometric returns. The usual approach for defining abnormal buy-and-holds returns in the literature differences out the buy and hold return of a counterfactual portfolio or stock (Savor & Lu, 2009; Barber & Lyon, 1997a) from a stock's buy-and-hold return. In our setting, this is analogous to

$$\prod_{\kappa=0}^{H}(1+R_{i,s+\kappa}(s)) - \prod_{\kappa=0}^{H}(1+R_{i,s+\kappa}(\infty)). \tag{7}$$

This object is challenge to analyze analytically, and has many challenging statistical properties (Barber & Lyon, 1997a; Mitchell & Stafford, 2000).

In our notation, this corresponds to the following geometric estimands. Let the cohort-horizon geometric ATT for cohort s at horizon H as

$$\begin{aligned}\tau^{geo,ATT}(s,H) &= E(\log(\prod_{\kappa=0}^{H}(1+R_{i,s+\kappa}(s)) - E(\log(\prod_{\kappa=0}^{H}(1+R_{i,s+\kappa}(\infty)))) \\ &= \sum_{\kappa=0}^{H} E(\log(1+R_{i,s+\kappa}(s))) - E(\log(1+R_{i,s+\kappa}(\infty))).\end{aligned}$$

As with the arithmetic ATT, this can be averaged over the event timings:[3]

$$\theta_H^{geo,ATT} = \sum_s w_s \tau^{geo,ATT}(s,H)$$

Note that this estimand effectively studies the *percentage* difference in gross cumulative returns, rather than *level* difference in gross cumulative returns. For researchers interested in sign tests (e.g. positive or negative long-run returns), both objects work equally well.

We now present a result tying the arithmetic (abnormal return) and geometric (buy and hold) ATT together.

Lemma 1. *The following holds for $\tau^{geo,ATT}(s,t)$ under all models of $R_{it}(\infty)$:*

$$\theta_H^{geo,ATT} = \theta_H^{ATT} - \sum_s w_s \sum_{\kappa=0}^{H}\left[E(R_{i,s+\kappa}(\infty)\tau_i(s,s+\kappa) + \frac{1}{2}\tau_i(s,s+\kappa)^2 \mid T_i = s)\right]. \tag{8}$$

If treatment effects and control return are independent across cohort s, such that we can

[3] Note that in the case of the geometric cumulative return, we first cumulate over the holding period, and then average across periods, since the non-linear structure makes the two non-interchangeable.

write $\mu = E(R_{i,s+\kappa}(\infty)|T_i = s)$ for all κ and s, then this can be simplified to

$$\theta_H^{geo,ATT} = (1-\mu)\theta_H^{ATT} - \sum_{\kappa=0}^{H} \frac{1}{2} Var(\theta_\kappa^{ATT}) - \frac{1}{2}(\theta_\kappa^{ATT})^2 \tag{9}$$

This result shows that buy-and-hold returns incorporate both volatility drag and the interaction between base returns and treatment effects, making them more complex to analyze than arithmetic returns. An important implication of this is if the counterfactual return $\hat{R}$ chosen for $R_{it}(s)$ identifies $E(R_{it}(\infty)|T_i = s)$, it may be a *bad* counterfactual for buy-and-hold returns because *it does not match on volatility.* For example, a portfolio with identical returns to $E(R_{it}(\infty)|T_i = s)$ may have much lower variance (due to diversification). As a result, the volatility drag from the *treated* observed units will bring down the geometric returns, even in the absence of any true effect!

As a result of Lemma 1, we focus on estimating the arithmetic ATT, rather than approximating the buy-and-hold return.[4] Geometric returns would require a counterfactual return portfolio that matches on *both* level and variance, and since the variance of a portfolio does not have the same theoretical guidance for a model as expected returns, finding this counterfactual portfolio is quite hard. It also suggests that papers that use buy-and-hold abnormal returns may contaminate their results as a function of how many firms are included in the counterfactual return portfolio due to diversification differences.

2.1.6 Factor Model Structure

We now operationalize our model for $E(R_{it}(\infty)|T_i = s,)$, based on a long literature in asset pricing (Chamberlain & Rothschild, 1983; Connor, 1984).

Assumption 1 (*Linear Factor Model*). *In the absence of the event, the average return of the portfolio of assets exposed to the event in period s follows a linear factor model with intercept α_i, K time varying factors $\mathbf{F}_t$ and factor weights β_i, such that*

$$E\left(R_{it}(\infty) \mid T_i = s\right) = \alpha_s + \beta_s \mathbf{F}_t, \tag{10}$$

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

Note that the linear factor 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

[4]The issues raised here are analogous to problems in difference-in-difference for log vs. level outcomes. If the parallel trends assumption holds for a level outcome, than it almost surely cannot equivalently hold for a log outcome, unless the treatment is randomly assigned. (Roth & Sant'Anna, 2023)

anticipate an event (rationally) in the future if the event does not eventually occur.[5] However, it nests generally almost all financial event study methods, such as using the market model, CAPM, or Fama-French factors to construct the counterfactual return (Campbell et al., 1997). It is also possible that this model could only hold for a short period of time, allowing for varying loadings over a longer period of time (as in Kelly et al., 2019).

2.1.7 Identification Assumptions

Assumption 2 (*Limited Anticipation*). *For some known $\delta \geq 0$,*

$$R_{i,t}(s) = R_{i,t}(\infty) \quad \text{for all } t < s - \delta \tag{11}$$

Remark 1. Since we can write

$$\begin{aligned} E(R_{i,t}(T_i) \mid T_i = s, \boldsymbol{F}_t) &= E(R_{i,t}(\infty) \mid T_i = s, \boldsymbol{F}_t) + \tau(t,s)^{ATT} && (12) \\ &= \alpha_s + \beta_s \boldsymbol{F}_t + \tau(s,t)^{ATT}, && (13) \end{aligned}$$

Assumption 2 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 δ periods before the event. Setting $\delta > 0$ allows for some pre-event information leakage, while $\delta = 0$ assumes no anticipation.

Assumption 3 (*Event Assignment*). *The probability that security i experiences an event at time t is:*

$$p_t(\mathbf{X}_i, \mathbf{F}) = \Pr(T_i = t \mid \mathbf{X}_i, \mathbf{F}) \tag{14}$$

where $\mathbf{X}_i = (\alpha_i, \boldsymbol{\beta}_i)$ represents security characteristics and $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_T)$ represents all factor realizations.

Two important special cases are:

- **Random assignment**: $p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{F})$ (event assignment independent of security characteristics)
- **Random timing**: $p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{X}_i)$ (event timing independent of factor realizations)

These assumptions formalize when simple estimators will be unbiased and when more sophisticated methods are needed.

[5] This issue is considered in a series of papers in the finance literature, e.g. Prabhala (1997), that consider conditional events.

This assumption implies that the treated group cannot have an impact from the announcement for a sufficient window 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 5.2 – 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.

2.2 Estimators

We now present four sets of estimators and characterize the conditions under which they identify the ATT. In all cases, we assume returns are already adjusted for the risk-free rate.

2.2.1 The Abnormal Returns Approach

Consider first the canonical abnormal returns model used in finance research (Campbell et al., 1997; Brown & Warner, 1985a). The researcher begins by selecting a set of observable factors F_t^o and estimates factor loadings $(\hat{\alpha}_i, \hat{\beta}_i)$ using ordinary least squares on data prior to $T_i - \delta$:

$$R_{it} = \alpha_i + \beta_i F_t^o + \varepsilon_{it}, \quad t < T_i - \delta \tag{15}$$

These estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$ minimize squared prediction errors for stock i's returns using the observed factors. The factors F_t^o may include no factors, a single factor (the market return), or multiple factors (e.g., Fama-French factors).

Definition 3 (*Abnormal Returns Estimator*). *Define the predicted return for stock i at time t as $\hat{R}_{it} = \hat{\alpha}_i + \hat{\beta}_i F_t^o$. The abnormal return is:*

$$AR_{it} = R_{it} - \hat{R}_{it} \tag{16}$$

The cohort-period abnormal return estimator is:

$$\tau^{AR}(s,t) = \mathbb{E}(AR_{i,t}|T_i = s) = \mathbb{E}(R_{i,t}|T_i = s) - \mathbb{E}(\hat{R}_{it}|T_i = s) \tag{17}$$

This approach attempts to remove the component of returns attributable to systematic factor exposure, leaving only the "abnormal" component. Under correct specification of the factor structure ($F_t^o = F_t$ for all relevant factors), this abnormal component should isolate the treatment effect. However, when factors are omitted or mismeasured, the estimated loadings $\hat{\beta}_i$ may fail to capture the true exposures β_i, leading to bias.

2.2.2 Alternative Approaches

We compare the abnormal returns approach to three alternatives estimation approaches.

Definition 4 (*Difference-in-Means Estimator*). *The difference-in-means estimator compares average returns of treated securities to a control group:*

$$\hat{\tau}^{cont}(s,t) = \mathbb{E}(R_{i,t}|T_i = s) - \mathbb{E}(R_{i,t}|i \in C) \tag{18}$$

$$\hat{\theta}_{\kappa}^{cont} = \sum_{s \in S} w_s \hat{\tau}^{cont}(s, s+\kappa) \tag{19}$$

When the control group consists of all securities weighted by market capitalization, this estimator corresponds to the "market-adjusted-return model" of Campbell et al. (1997) and Brown and Warner (1985a). Alternatively, the control group might consist of matched firms selected based on observable characteristics, as in Barber and Lyon (1997a) and Loughran and Vijh (1997).

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

Definition 5 (*Synthetic Control Estimator*). *Let $R_{s,t} = \mathbb{E}(R_{it}|T_i = s)$ denote the average return of securities treated at time s. The synthetic control estimator constructs a weighted portfolio of control securities to match the pre-event return path of the treated portfolio:*

$$\hat{\tau}^{synth}(s,t) = R_{s,t} - \sum_{j \in C} \hat{\omega}_j R_{j,t} \tag{20}$$

$$\hat{\theta}_{\kappa}^{synth} = \sum_{s \in S} w_s \hat{\tau}^{synth}(s, s+\kappa) \tag{21}$$

where the weights $\hat{\omega}_j$ solve:

$$\hat{\omega} = \arg\min_{\omega} \sum_{t<s-\delta} \left(R_{s,t} - \sum_{j \in C} \omega_j R_{j,t} \right)^2 \tag{22}$$

and are subject to a non-negativity constraint: $\omega_j \geq 0$.

The synthetic control method originated in Abadie and Gardeazabal (2003) and Abadie et al. (2010) and has expanded and grown as a method over the last decade. Synthetic control directly constructs a counterfactual by matching the pre-event dynamics of treated securities using a portfolio of controls. This synthetic control is then used as a counterfactual return following the event.

The key distinction from abnormal returns is that synthetic control does not require the researcher to specify or estimate the underlying factor structure. Instead, it searches for portfolio weights that replicate the treated group's returns in the pre-period, effectively letting the data determine the appropriate factor exposures. If such a replicating portfolio exists, it should continue to provide valid counterfactual returns in the post-period (absent the treatment).

We could depart from the original synthetic control applications by allowing negative weights. Traditionally, synthetic control methods restrict $\omega_j \geq 0$ to ensure the counterfactual represents a convex combination of control units. However, this restriction is unnecessarily limiting in financial applications. Allowing negative weights permits short positions and significantly expands the set of achievable factor loadings, making it more likely that a replicating portfolio exists. This flexibility is natural in financial markets and consistent with standard long-short portfolio construction. However, absent this restriction, we are not able to prove our results on unbiasedness using results from Ferman (2021).

We focus on constructing a single synthetic control for the portfolio of treated securities ($R_{s,t}$) rather than constructing separate synthetic controls for each individual security. This choice reflects both practical and theoretical considerations. Empirically, individual stock returns contain substantial idiosyncratic noise that would make firm-by-firm matching challenging. Theoretically, our estimands target average treatment effects for groups of securities, not individual effects, making portfolio-level analysis natural. This approach follows very naturally the approach advocated in Ben-Michael et al. (2022) for staggered synthetic control.

In practice, perfect pre-period fit may not be achievable. Extensions by Abadie and L'hour (2021) and Ben-Michael et al. (2021, 2022) allow for approximate rather than exact matching, trading off pre-period fit against overfitting concerns. However, most importantly for our analysis in financial event studies, Ferman (2021) shows that if the data follows a linear factor structure, then with sufficient pre-event time periods and control units, the estimator is consistent. This is consistent with a wide-range of asset pricing work highlighting the importance of having assets that span risk factors (Giglio & Xiu, 2021; Giglio et al., 2025).There is also a close connection to the mimicking-portfolio approach (E.g. Huberman et al., 1987).

We also consider a third estimator, following Xu (2017), which uses PCA regression with cross-validation to estimate a factor structure with unknown factors:

Definition 6. *The Gsynth approach assumes that non-treated stocks follow an interactive fixed effects model:*

$$R_{it}(\infty) = \alpha_i + \boldsymbol{\lambda}_i'\mathbf{F}_t + \varepsilon_{it} \tag{23}$$

where $\mathbf{F}_t$ are r unobserved common factors and $\boldsymbol{\lambda}_i$ are unit-specific factor loadings.

The estimation proceeds in three steps:

Step 1: Initial Factor Estimation *Using only control units, estimate factors via principal components:*

$$(\hat{\mathbf{F}}, \hat{\boldsymbol{\Lambda}}) = \arg\min_{\mathbf{F},\boldsymbol{\Lambda}} \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} (R_{it} - \alpha_i - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2 \tag{24}$$

subject to normalization constraints $\mathbf{F}'\mathbf{F}/T = \mathbf{I}_r$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ diagonal.

Step 2: Cross-Validation for Model Selection *Select the number of factors r via cross-validation:*

$$\hat{r} = \arg\min_{r\in\{1,\dots,r_{max}\}} CV(r) \tag{25}$$

where $CV(r)$ is the cross-validated mean squared prediction error using pre-treatment periods for the treatment group.

Step 3: Counterfactual Construction *For each treated unit i with $T_i = s$:*

1. *Estimate unit-specific loadings using pre-treatment data:*

$$\hat{\boldsymbol{\lambda}}_i = \arg\min_{\boldsymbol{\lambda}} \sum_{t<s-\delta} (R_{it} - \alpha_i - \boldsymbol{\lambda}'\hat{\mathbf{F}}_t)^2 \tag{26}$$

2. *Construct counterfactual for post-treatment periods:*

$$\hat{R}_{it}^{GSC}(\infty) = \hat{\alpha}_i + \hat{\boldsymbol{\lambda}}_i'\hat{\mathbf{F}}_t \tag{27}$$

The treatment effect estimate is:

$$\hat{\tau}^{GS}(s,t) = \frac{1}{N_s} \sum_{i:T_i=s} \left(R_{it} - \hat{R}_{it}^{GS}(\infty)\right) \tag{28}$$

The Gsynth estimator more directly leans on the linear factor structure, but does not require knowing the true factors, and uses the set of control firms to construct the set of counterfactual returns.

We focus on these two alternative estimators, but other alternative methods, such as IPCA Kelly et al. (2019) or the three-pass method in Giglio and Xiu (2021) may work as well or better. We leave it to future work to consider what approaches may work best.

2.3 Theoretical Results

We now establish conditions under which these estimators identify the ATT. For this proposition, it is convenient to see how these estimators differ from the target single event-period estimand:

$$\tau^{AR}(s,t) - \tau^{ATT}(s,t) = (\alpha_s - \hat{\alpha}_s) + (\beta_s F_t - \hat{\beta}_s F_t^o) + \varepsilon_{st} \tag{29}$$

$$\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) = (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) F_t + (\varepsilon_{st} - \varepsilon_{\infty,t}) \tag{30}$$

$$\hat{\tau}^{alt}(s,t) - \tau^{ATT}(s,t) = (\alpha_s - \hat{\alpha}_s^{alt}) + (\beta_s - \hat{\beta}_s^{alt}) F_t + \varepsilon_{st} \tag{31}$$

where $\alpha_s = \mathbb{E}(\alpha_i | T_i = s)$, $\beta_s = \mathbb{E}(\beta_i | T_i = s)$ are the average intercept and factor loadings for treated securities, α_∞ and β_∞ are corresponding quantities for the control group, $\hat{\alpha}_s$ and $\hat{\beta}_s$ are the estimated loadings from the abnormal returns approach, and $\hat{\alpha}_s^{alt}$ and $\hat{\beta}_s^{alt}$ are the implied loadings from either the synthetic control or gsynth estimator. $\varepsilon_{st} = n_s^{-1} \sum \varepsilon_{it}$ is the average idiosyncratic noise for the i cohort, and $\varepsilon_{\infty,t} = n_{i\in\mathcal{C}}^{-1} \sum varepsilon_{it}$ is the average noise for the control group.

Proposition 1 (*Single Event Finite Sample and Asymptotic Bias*). *Let Assumptions 1, and 2 hold. Then:*

1. Asymptotic properties. As $n_s, n_c, T_{pre} \to \infty$:

$$\tau^{AR}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} (\alpha_s - \tilde{\alpha}_s) + (\beta_s F_t - \tilde{\beta}_s F_t^o) \tag{32}$$

$$\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) F_t \tag{33}$$

$$\hat{\tau}^{synth}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0 \tag{34}$$

   where $\tilde{\alpha}_s$ and $\tilde{\beta}_s$ are probability limits of the estimated parameters.

2. Under random assignment ($p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{F})$),as $n_s, n_c \to \infty$, the difference-in-

means estimator is consistent even with fixed T_pre:

$$\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0 \tag{35}$$

3. Under correct specification, ($F_t^o = F_t$ for all t), the abnormal returns estimator is consistent as $n_s, T_{pre} \to \infty$:

$$\tau^{AR}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0 \tag{36}$$

All proofs are in Appendix A.

Remark 2 (*Priced Risk, Unpriced Risk, and Event Timing*). The bias expressions in Equations (32)–(33) can be decomposed into two distinct channels. First, the true return-generating process may include systematic factors F_t^u that carry no risk premium ($E[F_t^u] = 0$) and are therefore absent from standard asset pricing models. These unpriced factors contribute nothing to expected returns but generate return variation on any given date; the bias from omitting them is $\beta_s^u F_t^u$, which can be large in a single event but averages to zero across many randomly-timed events by the law of large numbers. Second, even for priced factors F_t^o that *are* included in the model, the OLS intercept absorbs the *unconditional* average premium $E[F_t^o]$. If event timing is non-random—so that the conditional distribution of factor realizations during event windows, $E[F_t^o \mid \text{event at } t]$, differs from the unconditional mean—the intercept absorbs the wrong premium, generating bias proportional to $\beta_s^o(E[F_t^o \mid \text{event at } t] - E[F_t^o])$. Both channels vanish under random timing with stationary factors: unpriced factors average out because $E[F_t^u] = 0$, and priced factors are well-captured because random timing ensures $E[F_t^o \mid \text{event at } t] = E[F_t^o]$. This is why the folk wisdom that "the model doesn't matter" holds under these conditions but fails when events are few, clustered, or coincide with unusual market states.

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 an unknown linear factor model and many control units. The results for Gsynth also follow directly from Xu (2017). The other two results follow from the assumptions and the definition of the estimators.

Remark 3. 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). These inconsistencies are similar to the inconsistencies highlighted in Theorem 1 of Andrews (2005). 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. If the treatment is randomly assigned, then the difference-in-means estimator is also consistent, since $\beta_s = \beta_\infty$ in the limit. This implies that these estimators are highly susceptible to coincident shocks at the same time, and the inference will be highly suspect (hence the need to cluster on event-timings in many financial event studies).

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 $\tau^{ATT}(s,t)$. 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. 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. As a result, this bias can be quite large. Formally, we can write the following:

Corollary 1. *Consider the single-event bias from Proposition 1. Suppose $F_t^o = F_t$ (correct factors with misestimated loadings). Let $a = \alpha_s - \tilde{\alpha}_s$ denote the intercept bias and $b = \beta_s - \tilde{\beta}_s$ denote the loading bias. Then:*

$$|\tau^{AR}(s,s) - \tau^{ATT}(s,s)| = |a + b \cdot F_t| \tag{37}$$

which is minimized at $F_t = -a/b$ (when $b \neq 0$) and increases in $|F_t|$ for $|F_t| > |a/b|$. In particular:

1. *If the intercept bias is zero ($a = 0$), the absolute bias is monotonically increasing in $|F_t|$.*
2. *If $|F_t|$ is sufficiently large relative to $|a/b|$, the absolute bias is approximately $|b| \cdot |F_t|$ and increases with factor magnitude.*

Hence, during periods of extreme factor realizations, even small loading misspecification ($b \neq 0$) can generate substantial bias.

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

Theorem 2.1 (*Bias with multiple events*). *Let Assumption 1 and 2 hold.*

1. *If $n_s, n_c, T_{pre} \to \infty$, then asymptotically, the synthetic control and gsynth estimators are unbiased,*

$$\hat{\theta}_\kappa^{alt} - \theta_\kappa^{ATT} \to_p 0. \tag{38}$$

where $alt \in \{synth, gs\}$.

2. *If $|\mathcal{S}| > 0$ and $1 > p_t(\mathbf{X}_i, \mathbf{F}) > \epsilon > 0$, then if $n_s, n_c, T_{pre} \to \infty$, the other two estimators are biased and converge to a weighted combination of conditional expected risk premia across the event periods:*

$$\hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} = E\left((\alpha_s - \tilde{\alpha}_s) + (\beta_s \boldsymbol{F}_{s+\kappa} - \tilde{\beta}_s \boldsymbol{F}^o_{s+\kappa}) \mid T_i \in \mathcal{S}\right) \tag{39}$$
$$\hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT} = E\left((\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty) F_t \mid T_i \in \mathcal{S}\right) \tag{40}$$

3. *If random assignment across firms holds, then for $n_s, n_c \to \infty$,*

$$\hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT} \to_p 0. \tag{41}$$

4. *If random timing holds and factors are stationary, then for $n_s, T_{pre} \to \infty$,*

$$\hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} \xrightarrow{p} 0. \tag{42}$$

The asymptotic unbiasedness follows because the OLS intercept absorbs the mean factor premium: the population identity $\alpha_i - \tilde{\alpha}_i = -\beta_i E[F_t] + \tilde{\beta}_i E[F_t^o]$ exactly cancels the factor terms in the bias expression.

5. *If random timing holds but factors are non-stationary such that $E[F_t \mid t < s - \delta] \neq E[F_t]$, then the bias is:*

$$\hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} = E(\beta_i \mid T_i \in \mathcal{S}) \left(E[F_t] - E[F_t \mid t \in \textit{pre-period}]\right) \tag{43}$$
$$- E(\tilde{\beta}_i \mid T_i \in \mathcal{S}) \left(E[F_t^o] - E[F_t^o \mid t \in \textit{pre-period}]\right) \tag{44}$$

Remark 4. The zero-bias result under random timing may seem counterintuitive given that the factor model is misspecified. The key insight is that the OLS intercept estimation creates an automatic adjustment: any persistent bias from omitted factors is absorbed into

$\tilde{\alpha}_i$ during the pre-event window. This adjustment is correct on average when timing is random because the same factor means apply to both estimation and event windows. The bias reappears when (i) timing correlates with factor realizations, or (ii) factor distributions shift between windows.

A implication of this Theorem is that under random timing with stationary factors, the abnormal returns estimator is unbiased, even when the factor model is misspecified. The OLS intercept absorbs the mean factor premium, yielding exact cancellation. This provides a theoretical foundation for the simulation evidence in Brown and Warner (1985a) showing that the form of the abnormal return estimator has limited effects on the estimates, and for the common view that "the structure of the model does not matter." However, this zero-bias result requires both random timing *and* factor stationarity. When either condition fails, the bias expressions in parts 2 and 4b apply.

An implication of this is that the abnormal returns estimator is can be 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 factors, consistent with the simulation evidence in Brown and Warner (1985a) that the form of the abnormal return estimator has limited effects on the estimates.[6] In fact, 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."

2.4 When does bias accumulate in long-run event studies?

Researchers are often interested in the trends or cumulative impact of events on returns, as measured by cumulative abnormal returns or buy-and-hold abnormal returns. 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

[6]The simulations in Brown and Warner (1985a) 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."

abnormal return studies of firm activity. As we show in Lemma 1, the buy-and-hold abnormal return has additional challenges caused by variance considerations in the counterfactual portfolio. We now use our results in Proposition 1 and Theorem 2.1 to characterize when misspecification bias accumulates over longer horizons.

A perhaps surprising implication of Theorem 2.1 part 4 is that under random timing with stationary factors, the asymptotic bias of the abnormal returns estimator is *exactly zero*, even when the factor model is misspecified. This occurs because the OLS intercept $\tilde{\alpha}_i$ absorbs the mean factor premium during estimation. Formally, the population identity $\alpha_i - \tilde{\alpha}_i = -\beta_i E[F_t] + \tilde{\beta}_i E[F_t^o]$ ensures exact cancellation with the factor terms in the bias expression. This provides a theoretical foundation for the folk wisdom that "the structure of the model does not matter"—but only under specific conditions.

The cancellation breaks down in two empirically relevant scenarios. First, when event timing is not random (Theorem 2.1 part 2), the bias reflects the covariance between factor realizations and treatment timing. If mergers cluster during market booms or crisis-related events occur during high-volatility regimes, this bias can be substantial. Second, even under random timing, bias emerges if factors are non-stationary—that is, if the mean factor return during the estimation window differs from the unconditional mean.

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 if either condition fails. Under non-stationary factors, the bias takes the form:

$$E[\beta_i \mid T_i \in \mathcal{S}]\,(E[F_t \mid \text{post}] - E[F_t \mid \text{pre}]) - E[\tilde{\beta}_i \mid T_i \in \mathcal{S}]\,(E[F_t^o \mid \text{post}] - E[F_t^o \mid \text{pre}]) \quad (45)$$

which accumulates linearly with horizon H. The first term reflects drift in the true factors weighted by true loadings; the second reflects drift in the observed factors weighted by estimated loadings. When the observed factors F_t^o are stationary but the true factors F_t are not, only the first term contributes to bias.

To illustrate the potential magnitude, suppose the observed factors F_t^o are stationary, so the second term vanishes. The bias then depends entirely on omitted factors. If the true model includes an unobserved factor—say, a liquidity or volatility factor—that averages 2% annually during the pre-event estimation window but 5% during the post-event window, the drift is 3% per year. If treated firms have an average loading of 0.2 on this omitted factor, the bias would be approximately 1.8% at the three-year horizon. This bias reflects the combination of misspecification (such that the factor is not accounted for) combined with a different conditional expected factor realizations following the treatment.

Note that these issues are not solved by using multiple event timings when timing correlates with market conditions. In such cases, the bias in factors cannot average out to zero. It is also worth remarking how the results from Mitchell and Stafford (2000) can be seen analytically in our statistical framework. While the misspecification term creates bias under non-random timing, it also creates cross-correlation in errors for every event-timing, as they note: "[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."

3 Extensions

3.1 Testing for over- and underreaction

A natural question in event studies is whether the market correctly incorporates information at announcement, or whether there is subsequent drift indicating initial mispricing. The behavioral finance literature has proposed various theories of over- and underreaction (Daniel et al., 1998; Barberis et al., 1998; Hong & Stein, 1999), and one empirical approach examines whether short-run announcement effects predict subsequent long-run returns (Kwon & Tang, 2025; Lyu & Xu, 2024).

We can formalize this as a regression of long-run effects on short-run effects. Define the short-run cumulative treatment effect for firm i treated at time s as

$$\tau_i^{short} = \sum_{\kappa=\kappa_0}^{\kappa_1} \tau_i(s, s+\kappa) \tag{46}$$

where $[\kappa_0, \kappa_1]$ is a short window around the announcement (e.g., $[-1, +2]$). Similarly, define the long-run cumulative effect as

$$\tau_i^{long} = \sum_{\kappa=\kappa_1+1}^{H} \tau_i(s, s+\kappa) \tag{47}$$

for some horizon H (e.g., 90 days). The regression of interest is

$$\tau_i^{long} = \alpha + \beta \cdot \tau_i^{short} + \varepsilon_i \tag{48}$$

where the slope coefficient captures the cross-sectional relationship between initial and subsequent responses:

$$\beta = \frac{Cov(\tau_i^{long}, \tau_i^{short})}{Var(\tau_i^{short})}. \tag{49}$$

The coefficient β has a natural interpretation in terms of market efficiency. Under efficient pricing, the announcement effect should fully incorporate the information content of the event, leaving long-run effects uncorrelated with short-run effects conditional on firm characteristics. This implies $\beta = 0$. Deviations from zero suggest mispricing: $\beta > 0$ indicates underreaction, where firms with larger initial responses continue to drift in the same direction; $\beta < 0$ indicates overreaction, where initial responses partially reverse.

In practice, researchers estimate this regression using abnormal returns:

$$\widehat{CAR}_i^{long} = \alpha + \beta \cdot \widehat{CAR}_i^{short} + \varepsilon_i \tag{50}$$

where $\widehat{CAR}_i^{short}$ and $\widehat{CAR}_i^{long}$ are cumulative abnormal returns over the respective windows. The identifying assumption is that estimated abnormal returns are unbiased proxies for the true treatment effects τ_i^{short} and τ_i^{long}.

Our earlier results highlight two concerns with this approach. First, measurement error in short-run effects biases $\hat{\beta}$ toward zero. If $\widehat{CAR}_i^{short} = \tau_i^{short} + \eta_i$ where η_i is noise uncorrelated with true effects, then

$$\hat{\beta} \xrightarrow{p} \beta \cdot \frac{Var(\tau_i^{short})}{Var(\tau_i^{short}) + Var(\eta_i)} \tag{51}$$

which attenuates the estimated coefficient. This makes detecting true underreaction more difficult.

Second, and more importantly for our purposes, factor model misspecification can generate spurious correlation between short-run and long-run abnormal returns even in the absence of any over- or underreaction. To see this, decompose the abnormal return estimation error. Under a misspecified model, we can write

$$\widehat{CAR}_i^{short} = \tau_i^{short} + (\beta_i - \tilde{\beta}_i) \sum_{\kappa=\kappa_0}^{\kappa_1} F_{s+\kappa} + \nu_i^{short} \tag{52}$$

$$\widehat{CAR}_i^{long} = \tau_i^{long} + (\beta_i - \tilde{\beta}_i) \sum_{\kappa=\kappa_1+1}^{H} F_{s+\kappa} + \nu_i^{long} \tag{53}$$

where $\beta_i - \tilde{\beta}_i$ is the loading estimation error on omitted factors. If firms with larger true loadings on omitted factors (and hence larger estimation errors) cluster together, both short-run and long-run abnormal returns will be contaminated in the same direction. This generates positive covariance between $\widehat{CAR}_i^{short}$ and $\widehat{CAR}_i^{long}$ even when true effects are uncorrelated.

The magnitude of this spurious correlation depends on the cross-sectional variance of factor loading errors and the serial correlation structure of factor realizations. Under random timing and stationary factors, the factor realizations $\sum_{\kappa=\kappa_0}^{\kappa_1} F_{s+\kappa}$ and $\sum_{\kappa=\kappa_1+1}^{H} F_{s+\kappa}$ are independent across the short and long windows for a given firm, but the loading errors $\beta_i - \tilde{\beta}_i$ are common. This induces correlation:

$$Cov(\widehat{CAR}_i^{short}, \widehat{CAR}_i^{long}) = Cov(\tau_i^{short}, \tau_i^{long}) + Var(\beta_i - \tilde{\beta}_i) \cdot E\left[\sum_{\kappa=\kappa_0}^{\kappa_1} F_{s+\kappa}\right] E\left[\sum_{\kappa=\kappa_1+1}^{H} F_{s+\kappa}\right] \tag{54}$$

The second term is positive when factor premia are positive, potentially generating the appearance of underreaction.

This concern is particularly acute for long-horizon tests. As the long-run window H increases, the cumulative factor premium $E[\sum_{\kappa=\kappa_1+1}^{H} F_{s+\kappa}]$ grows proportionally, amplifying the spurious correlation. A researcher finding $\hat{\beta} > 0$ over a three-year horizon cannot easily distinguish between true underreaction and correlated misspecification bias. The synthetic control and Gsynth estimators offer a partial solution. By matching on pre-event return dynamics rather than imposing a factor structure, these methods can reduce the correlation between loading errors and estimated effects.

Conversely, using buy-and-hold abnormal returns can generate spurious evidence of *overreaction*. As shown in Lemma 1, geometric returns incorporate volatility drag that does not affect arithmetic returns. If firms with larger announcement effects also have higher return volatility—a plausible relationship if larger news generates greater uncertainty—the long-run buy-and-hold return will be disproportionately attenuated for these firms. To see this, note that the volatility drag term $\frac{1}{2}Var(R_{it})$ cumulates over the long-run window but is negligible over the short-run window. Firms with large positive short-run effects that also have high volatility will appear to "give back" their gains over longer horizons, even if the true arithmetic treatment effects are uncorrelated across windows. This generates $\hat{\beta} < 0$ in a buy-and-hold framework, mimicking overreaction.

The problem is compounded when researchers use a diversified portfolio (such as the market index) as the benchmark for buy-and-hold returns. The benchmark's volatility drag is smaller than that of individual stocks due to diversification, creating a systematic downward bias in long-run buy-and-hold abnormal returns relative to short-run effects. Together, these forces suggest that findings of long-run reversal using buy-and-hold returns should be interpreted with caution, as they may reflect the mechanical properties of geometric returns rather than true overreaction to information.

3.2 Individual estimates are noisy, but not necessarily biased

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

Assumption 4. *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 5. 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:

$$\tau_i^{AR}(s,s) - \tau_i(s,s) = (\alpha_s - \hat{\alpha}_s) + (\beta_s \boldsymbol{F}_s - \hat{\beta}_s \boldsymbol{F}_s^o) + \varepsilon_{it}. \tag{55}$$

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 litigation (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.

3.3 Calendar-Time Portfolio Approaches

An alternative to the event-time approach is the calendar-time portfolio method advocated by Fama (1998) and Mitchell and Stafford (2000). Rather than aligning returns by event time and averaging across firms, this approach forms a portfolio at each calendar date consisting of all firms recently affected by the event, then estimates abnormal performance relative to observerd factors. We now characterize the calendar-time estimand in terms of our cohort-period ATTs and analyze its properties under misspecification.

3.3.1 The Calendar-Time Estimand

At each calendar time t, define the set of firms currently within their event window as

$$\mathcal{T}_t = \{i : t - H \leq T_i \leq t\} \tag{56}$$

where H is the length of the event window. The calendar-time portfolio return is

$$R_t^{cal} = \frac{1}{N_t} \sum_{i \in \mathcal{T}_t} R_{it} \tag{57}$$

where $N_t = |\mathcal{T}_t| = \sum_{s:t-H \leq s \leq t} n_s$ is the number of firms in the portfolio at time t.

The calendar-time approach estimates abnormal returns via the time-series regression

$$R_t^{cal} = \alpha^{cal} + \beta^{cal} F_t^o + \varepsilon_t^{cal} \tag{58}$$

where F_t^o are the observed factors. The intercept $\hat{\alpha}^{cal}$ is interpreted as the abnormal return.

To connect this to our event-time estimands, we decompose the portfolio return by cohort. At calendar time t, cohort s is at event-time $\kappa = t - s$, so we can write

$$R_t^{cal} = \sum_{s:t-H \leq s \leq t} \frac{n_s}{N_t} \cdot \bar{R}_{s,t} \tag{59}$$

where $\bar{R}_{s,t} = n_s^{-1} \sum_{i:T_i = s} R_{it}$ is the average return of cohort s at calendar time t.

Proposition 2 (*Calendar-Time Estimand Decomposition*). *Under Assumption 1, the calendar-time intercept identifies*

$$\alpha^{cal} = \sum_{s \in \mathcal{S}} \sum_{\kappa=0}^{H} \omega_{s,\kappa}^{cal} \cdot \tau^{ATT}(s, s+\kappa) + \bar{\alpha}^{cal} + bias^{cal} \tag{60}$$

where the calendar-time weights are

$$\omega_{s,\kappa}^{cal} = \frac{1}{T_{cal}} \cdot \frac{n_s}{N_{s+\kappa}}, \tag{61}$$

T_{cal} is the number of calendar periods in the sample, $N_{s+\kappa} = \sum_{s':(s+\kappa)-H \leq s' \leq (s+\kappa)} n_{s'}$ is the portfolio size at calendar time $s+\kappa$, $\bar{\alpha}^{cal} = \sum_{s,\kappa} \omega_{s,\kappa}^{cal} \alpha_s$ is the weighted average intercept of treated firms, and $bias^{cal}$ is defined in Proposition 3.

Proof. Switching the order of summation from calendar time to cohort-event time:

$$\frac{1}{T_{cal}} \sum_{t=1}^{T_{cal}} R_t^{cal} = \frac{1}{T_{cal}} \sum_t \sum_{s:t-H \leq s \leq t} \frac{n_s}{N_t} \cdot \bar{R}_{s,t} \tag{62}$$

$$= \frac{1}{T_{cal}} \sum_{s \in \mathcal{S}} \sum_{\kappa=0}^{H} \frac{n_s}{N_{s+\kappa}} \cdot \bar{R}_{s,s+\kappa} \tag{63}$$

where we use the fact that each (s, κ) pair maps to exactly one calendar time $t = s + \kappa$.

Substituting $E[\bar{R}_{s,s+\kappa} \mid F_{s+\kappa}] = \alpha_s + \beta_s F_{s+\kappa} + \tau^{ATT}(s, s+\kappa)$ and using the OLS population formula $\alpha^{cal} = E[R_t^{cal}] - \tilde{\beta}^{cal} E[F_t^o]$ yields the result after collecting terms. □

Remark 6 (*Comparison to Event-Time Weights*). The event-time estimand $\theta_\kappa^{ATT} = \sum_s \omega_s^{event} \cdot \tau^{ATT}(s, s+\kappa)$ uses weights $\omega_s^{event} = n_s / \sum_{s'} n_{s'}$ that depend only on cohort size. In contrast, the calendar-time weights $\omega_{s,\kappa}^{cal}$ depend on $N_{s+\kappa}$—the total number of events within H periods of calendar time $s+\kappa$. When events cluster in time, calendar-time downweights each event in the cluster. The two approaches target the same estimand if and only if event flow is uniform over calendar time.

Definition 7 (*Calendar-Time Average Treatment Effect*). *The calendar-time average treatment effect is*

$$\theta^{cal} = \sum_{s\in\mathcal{S}} \sum_{\kappa=0}^{H} \omega_{s,\kappa}^{cal} \cdot \tau^{ATT}(s, s+\kappa) \tag{64}$$

where $\omega_{s,\kappa}^{cal} = n_s/(T_{cal} \cdot N_{s+\kappa})$ *and* $\sum_{s,\kappa} \omega_{s,\kappa}^{cal} = 1$.

3.3.2 Bias Under Factor Model Misspecification

We now derive the bias of the calendar-time estimator under misspecification, paralleling our results for the event-time approach in Theorem 2.1.

Proposition 3 (*Calendar-Time Bias*). *Under Assumptions 1 and 2, the calendar-time bias is*

$$bias^{cal} = \sum_{s\in\mathcal{S}} \sum_{\kappa=0}^{H} \omega_{s,\kappa}^{cal} \left[\beta_s (F_{s+\kappa} - E[F_t]) - \tilde{\beta}^{cal} (F_{s+\kappa}^o - E[F_t^o])\right] \tag{65}$$

where $\tilde{\beta}^{cal} = \operatorname{plim} \hat{\beta}^{cal}$ *is the probability limit of the calendar-time beta estimate.*

In the special case where the observed factors equal the true factors ($F_t^o = F_t$), this simplifies to

$$bias^{cal} = \sum_{s\in\mathcal{S}} \sum_{\kappa=0}^{H} \omega_{s,\kappa}^{cal} \cdot (\beta_s - \tilde{\beta}^{cal}) \cdot (F_{s+\kappa} - E[F_t]). \tag{66}$$

Proof. The OLS intercept satisfies $\alpha^{cal} = E[R_t^{cal}] - \tilde{\beta}^{cal} E[F_t^o]$. Substituting the factor model:

$$E[R_t^{cal}] = \sum_{s,\kappa} \omega_{s,\kappa}^{cal} \left(\alpha_s + \beta_s E[F_{s+\kappa}] + \tau^{ATT}(s, s+\kappa)\right) \tag{67}$$

$$= \bar{\alpha}^{cal} + \theta^{cal} + \sum_{s,\kappa} \omega_{s,\kappa}^{cal} \beta_s E[F_{s+\kappa}] \tag{68}$$

where $E[F_{s+\kappa}]$ is the expected factor at calendar time $s+\kappa$, which may differ from $E[F_t]$ if event timing is non-random. Subtracting $\tilde{\beta}^{cal} E[F_t^o]$ and rearranging yields the result. $\square$

Corollary 2 (*Conditions for Unbiasedness*). *The calendar-time estimator is asymptotically unbiased for $\theta^{cal} + \bar{\alpha}^{cal}$ under any of the following conditions:*

1. *Correct factor specification with correctly estimated loadings: $F_t^o = F_t$ and $\tilde{\beta}^{cal} = \sum_{s,\kappa} \omega_{s,\kappa}^{cal} \beta_s$.*
2. *Random timing with stationary factors: $E[F_{s+\kappa}] = E[F_t]$ for all s, κ.*

Remark 7 (*Comparison to Event-Time Bias*). The calendar-time bias has a similar structure to the event-time bias in Theorem 2.1, but with two key differences. First, the weighting scheme differs as characterized above. Second, and more subtly, the probability limit $\tilde{\beta}^{cal}$ is estimated from *post-treatment* data, since the calendar-time regression uses returns during the event window. This matters when treatment affects factor loadings, which we address next.

3.3.3 Treatment Effects on Factor Loadings

A conceptual distinction between calendar-time and event-time approaches emerges when treatment affects systematic risk. Consider an acquisition that increases a firm's leverage, thereby raising its market beta. Should this change in risk compensation be counted as part of the treatment effect?

To formalize this, let the post-treatment return process be

$$R_{it}(s) = \alpha_i^{post}(s) + \beta_i^{post}(s) F_t + \tilde{\tau}_i(s,t) + \varepsilon_{it} \tag{69}$$

where $\beta_i^{post}(s)$ may differ from $\beta_i = \beta_i^{pre}$ and $\tilde{\tau}_i(s,t)$ is the "pure" abnormal return after accounting for loading changes. The total treatment effect decomposes as

$$\tau_i(s,t) = \underbrace{(\alpha_i^{post} - \alpha_i^{pre})}_{\text{intercept change}} + \underbrace{(\beta_i^{post} - \beta_i^{pre}) F_t}_{\text{risk compensation change}} + \underbrace{\tilde{\tau}_i(s,t)}_{\text{pure abnormal return}} . \tag{70}$$

Proposition 4 (*Estimand Under Loading Changes*). *When treatment affects factor loadings:*

1. *The event-time ATT captures the total effect:*

$$\tau^{ATT}(s,t) = E[\alpha_i^{post} - \alpha_i^{pre} \mid T_i = s] + E[(\beta_i^{post} - \beta_i^{pre}) \mid T_i = s] \cdot F_t + E[\tilde{\tau}_i(s,t) \mid T_i = s] \tag{71}$$

2. *The calendar-time intercept identifies (under correct specification):*

$$\alpha^{cal} \xrightarrow{p} \sum_{s,\kappa} \omega_{s,\kappa}^{cal} \Big(E[\alpha_i^{post} - \alpha_i^{pre} \mid T_i = s] + E[\tilde{\tau}_i(s, s+\kappa) \mid T_i = s] \Big) \tag{72}$$

which excludes the systematic risk compensation change $(\beta_i^{post} - \beta_i^{pre})F_t$.

Proof. For part 1, substitute the decomposition into Definition 1. For part 2, note that the calendar-time regression estimates β^{cal} using post-treatment returns, so $\tilde{\beta}^{cal} \xrightarrow{p} \sum_{s,\kappa} \omega_{s,\kappa}^{cal} E[\beta_i^{post} \mid T_i = s]$. The OLS intercept then nets out the factor exposure at the post-treatment loadings, leaving only the intercept change and pure abnormal return. □

Remark 8 (*Estimand Ambiguity*). When treatment affects factor loadings, the calendar-time and event-time approaches target economically different causal quantities. The event-time ATT answers: "What is the total difference in returns caused by treatment?" The calendar-time alpha answers: "What is the abnormal return after adjusting for the firm's new risk profile?" Neither is inherently correct—the choice depends on the research question. For studying whether shareholders gained or lost value, the total effect (event-time) may be appropriate. For studying whether returns are anomalous relative to risk, the risk-adjusted effect (calendar-time) may be preferred.

4 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.

4.1 Simulation Design with 2 Factors and Selection

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

$$r_{it} = r_{f,t} + \beta_{i,mkt}(r_{mkt,t} - r_{f,t}) + \beta_{i,smb} r_{smb,t} + \varepsilon_{i,t}, \tag{73}$$

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, 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

$$p(treated)_i = \frac{\exp(\delta \beta_{i,smb})}{1 + \exp(\delta \beta_{i,smb})}, \tag{74}$$

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.[7] The probability of any one of the candidate period being the treatment period is

$$p(selected)_t = \frac{\exp(\delta Rank_{2t})}{1 + \exp(\delta Rank_{2t})}, \tag{75}$$

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

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

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.

4.2 Simulation Results with 2 Factors and Selection

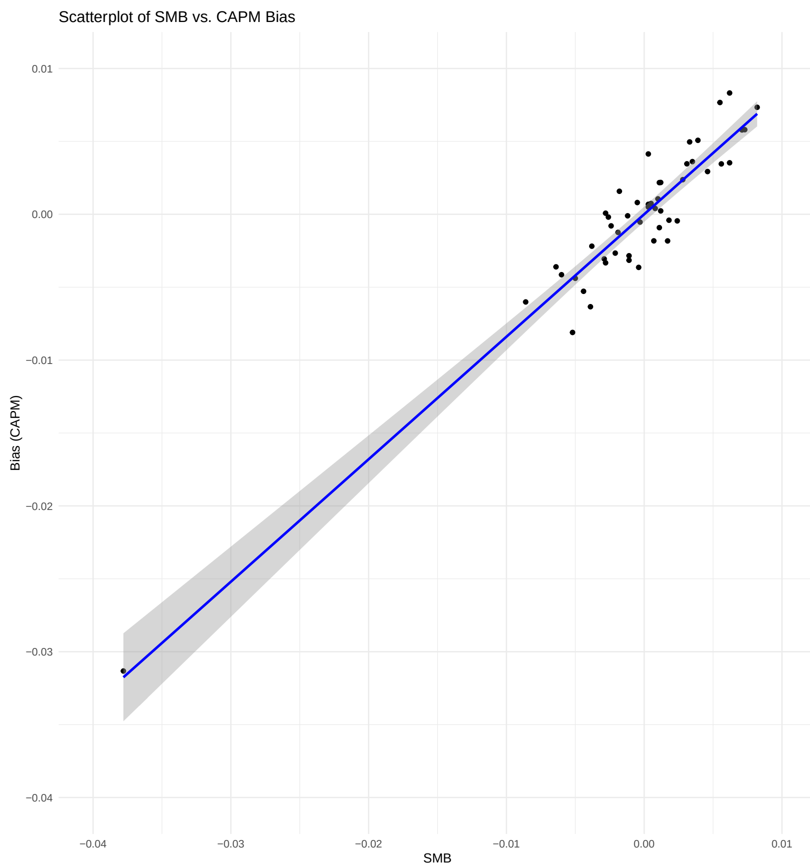
In Table 1, we compare the performance of four different estimators across 50 simulations: mean difference between treated and control firms, average abnormal returns using the market factor (estimating the factor loading for each treated firm in the pre-period), average abnormal returns using the both 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.

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 4.1. 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.

Figure 2: **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 4.1. 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 Theorem 2.1. 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.

5 Applications

5.1 Empirical Example 1: Geithner as Treasury Secretary

We now turn to our first empirical example, examining the period when the announcement of Timothy Geithner as Treasury Secretary was leaked, following the setup of Acemoglu et al. (2016). This example highlights the results of Proposition 1 in a simultaneous treatment setting. We demonstrate that the bias from an incorrect factor structure can be substantial in this setting, and that synthetic control methods help alleviate this bias. We argue that the bias arises from two sources: first, the event window coincides with turbulent market conditions characterized by large daily factor realizations; second, the counterfactual returns are constructed from control firms with substantially different factor exposures. We show that synthetic methods, which greatly reduce these biases, also match the factor loadings of treated firms for known factors such as size and value.

Empirical setup. We examine the announcement of Timothy Geithner as nominee for Treasury Secretary on November 21, 2008. Following Acemoglu et al. (2016), we estimate average treatment effects over the 11-day window encompassing and following the announcement date, from November 21, 2008 (day 0) through December 8, 2008 (day 10).[8] For treated and control bank returns, we use the data provided by the authors, who collected daily returns from Datastream.[9] For all trading days before and after the event, returns represent full trading day returns during regular trading hours. For the event day, returns are calculated from 3:00 p.m. (when the news leaked) until market close at 4:00 p.m.

We consider two sets of control firms. First, we use the same set of financial firms listed on the NYSE or NASDAQ that are not connected to Geithner, as in Acemoglu et al. (2016). Second, we expand the control group to include all NYSE, AMEX, and NASDAQ (exchange

[8]November 24, 2008 corresponds to day 1 due to the weekend.

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

codes 1–3) common stocks (share codes 10 or 11).

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

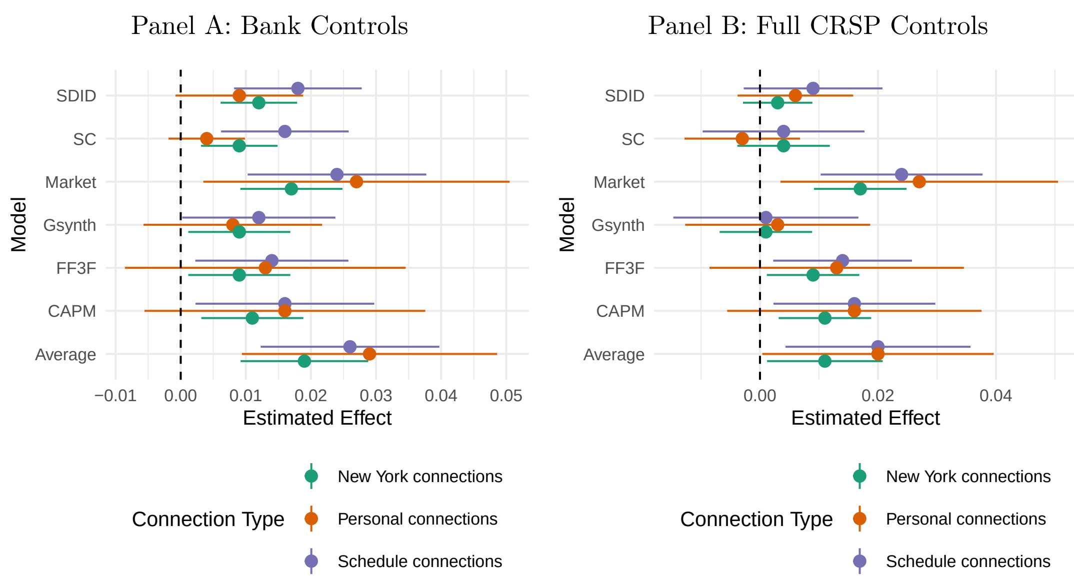
	(1) Average	(2) DID	(3) Market	(4) CAPM	(5) FF3F	(6) SC	(7) SDID	(8) Gsynth
Schedule connections	0.026***	0.027***	0.024***	0.016***	0.014***	0.016***	0.018***	0.012**
	(0.007)	(0.005)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)
Personal connections	0.029***	0.030***	0.027**	0.016	0.013	0.004	0.009**	0.008
	(0.010)	(0.006)	(0.012)	(0.011)	(0.011)	(0.003)	(0.005)	(0.007)
New York connections	0.019***	0.020***	0.017***	0.011***	0.009***	0.009***	0.012***	0.009**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Observations	5,995	129,165	129,165	129,165	129,165	129,165	129,165	129,625

Panel B: All Firm Controls

	Average	DID	Market	CAPM	FF3F	SC	SDID	Gsynth
Schedule connections	0.020**	0.020***	0.024***	0.016***	0.014***	0.004	0.009	0.001
	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)
Personal connections	0.020*	0.021***	0.027**	0.016	0.013	-0.003	0.006	0.003
	(0.010)	(0.006)	(0.012)	(0.011)	(0.011)	(0.005)	(0.005)	(0.008)
New York connections	0.011**	0.011***	0.017***	0.011***	0.009***	0.004	0.003	0.001
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
Observations	45,045	966,420	966,420	966,420	966,420	966,420	966,420	916,388

Non-connected banks as controls. We first use public financial institutions without connections to Geithner as control firms. Panel A of Table 2 reports the average treatment effects over the 11-day post-event window. Column 1 presents the difference in average returns between treated and control firms, implementing the counterfactual as a simple average of returns from non-connected firms—the same approach used in Table 2 of Acemoglu et al. (2016). Column 2 reports difference-in-differences estimates. Columns 3–5 present traditional factor model adjustments: the market model (Column 3), CAPM (Column 4), and Fama-French three-factor model (Column 5). Columns 6–8 employ synthetic control methods: standard synthetic control (Abadie et al., 2010), synthetic difference-in-differences (Arkhangelsky et al., 2021), and generalized synthetic control (Gsynth) from Xu (2017). For all models requiring pre-event estimation, we use days -256 to -31, slightly shorter than the -280 to -31 window in the original paper to maintain a balanced panel. We report a graphical version of Table 2 in Figure 3.

Figure 3: **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.



The results reveal a clear pattern. Simple averaging and difference-in-differences (Columns 1–2) show that firms with schedule connections experience 2.6–2.7% higher cumulative returns, those with personal connections show 2.9–3.0% higher returns, and firms with New York connections exhibit 1.9–2.0% higher returns. The market model adjustment (Column 3) produces minimal changes. However, risk-adjusted returns using CAPM and Fama-French models (Columns 4–5) reduce these estimates by approximately 40–50%, suggesting that connected firms have higher market betas. The synthetic control methods (Columns 6–8) produce even larger reductions, with standard synthetic control reducing schedule connection effects by 38% and personal connection effects becoming statistically insignificant.[10]

All public firms as controls. We next expand the control group to include all common shares traded on NYSE, AMEX, and NASDAQ, with results reported in Panel B of Table 2. This expansion is motivated by the integration of equity markets: systematic factors should be well-identified using the universe of traded stocks. Restricting controls to financial firms alone may be suboptimal unless banking-specific factors exist that cannot be spanned by the broader market.

The expanded control group dramatically changes the results from synthetic control methods while leaving traditional methods largely unaffected. Simple averaging and difference-in-differences (Columns 1–2) continue to show significant effects of approximately 2% for all connection types. The factor model adjustments (Columns 3–5) show a similar pattern to Panel A, with the market model producing minimal changes while CAPM and Fama-French adjustments reduce estimates by 30–50%.

Strikingly, the synthetic control methods now produce near-zero and statistically insignificant estimates. Standard synthetic control (Column 6) yields point estimates of 0.4% for schedule connections and −0.3% for personal connections. Gsynth (Column 8) estimates are particularly close to zero: 0.1% for schedule connections, 0.3% for personal connections, and 0.1% for New York connections—all statistically insignificant. This dramatic difference suggests that the broader control group allows synthetic methods to better match the factor exposures of treated firms, effectively eliminating the estimated treatment effects. The contrast between traditional factor adjustments (which still show significant effects) and synthetic methods (which do not) highlights the importance of allowing flexible, data-driven matching of factor exposures rather than imposing a specific factor structure.

5.1.1 Market Returns around Event

We now investigate the sources of bias in the original estimates. First, we examine the distribution of market returns during the event window. Figure 4 displays the kernel density of daily S&P 500 returns from 1962–2023, overlaid with the realized returns during the 11-day event window. The event period coincides with extraordinary market volatility, with returns falling in the extreme tails of the historical distribution. The market surged 6.6% on

[10] Our results contrast with Acemoglu et al. (2016), who employ synthetic control methods as robustness checks. Their approach was necessarily *ad hoc* given the limited literature at the time on handling multiple treated units in synthetic control settings.

November 21 (day 0) and 6.5% on November 24 (day 1), while the largest decline of −8.4% occurred on December 1 (day 5).

These extreme factor realizations have important implications for identification. As demonstrated in Proposition 1, when treatment occurs simultaneously for all units, abnormal return estimators are particularly sensitive to factor model misspecification. The bias is proportional to both the magnitude of factor realizations and the difference in factor loadings between treated and control firms: $(\beta_s - \hat{\beta}_s)F_t$. Large factor realizations during the event window amplify any misspecification bias arising from imperfect matching of factor exposures.

This mechanism explains the substantial reduction in estimated treatment effects when using synthetic control methods rather than simple averaging. Proposition 1 shows that both synthetic control and gsynth estimators are asymptotically unbiased even with omitted factors, as they construct control portfolios that match the pre-event factor structure of treated firms without requiring explicit factor model specification. The extreme market conditions during the Geithner announcement thus reveal the importance of proper counterfactual construction in volatile periods.

5.1.2 Factor loadings of treated units match synthetic control factor loadings

We now provide direct evidence on the factor exposure differences between treated and control firms. We estimate market betas using daily returns from day −280 to day −31 before the event, running firm-level time-series regressions on the S&P 500 index return for CAPM betas and on the Fama-French three factors for multifactor betas.[11]

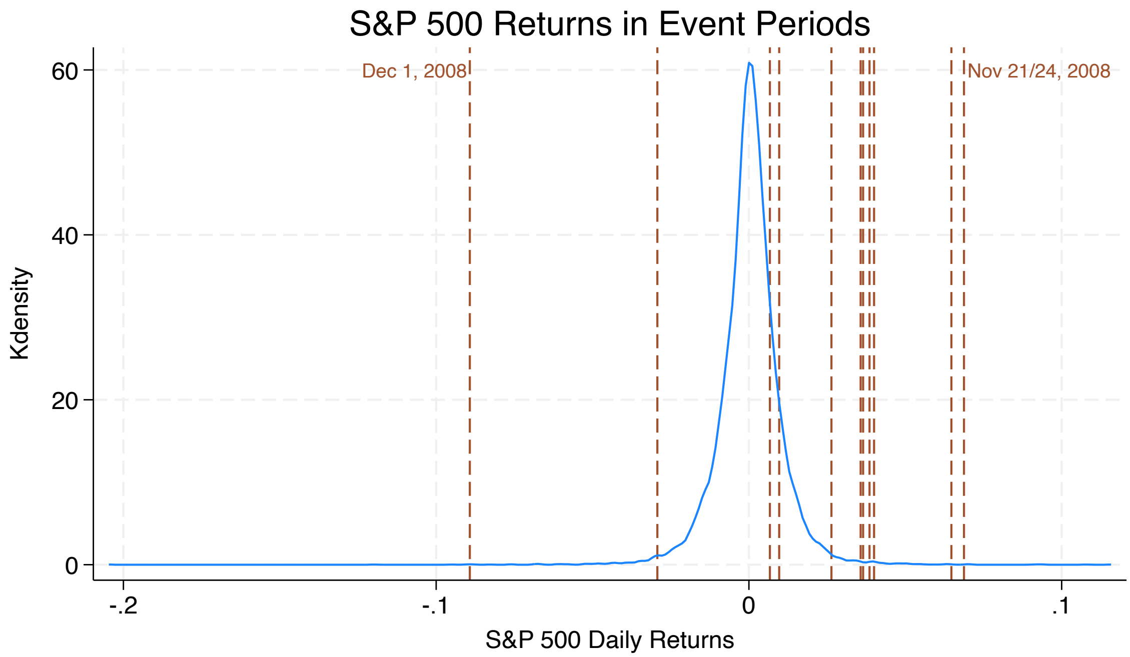
Table 3 reports the weighted average betas for treated and control portfolios. Panel A presents equal-weighted averages for treated firms and two control groups: financial institutions only and all public firms. The results reveal substantial factor exposure mismatches. Treated firms have an average CAPM beta of 1.43, compared to 0.83 for financial controls—a difference of 0.60. The Fama-French three-factor model confirms this pattern: treated firms exhibit a market beta of 1.28 versus 0.66 for controls, with similar disparities in SMB (0.23 vs. 0.75) and HML (0.61 vs. 0.72) exposures.

These factor loading differences, combined with the extreme market realizations documented in Section 5.1.1, generate substantial bias in simple difference estimators. During the event window, the 0.60 difference in market beta translates to a bias of approximately $0.60 \times 6.9\% = 4.1\%$ on November 21 alone. This mechanical bias explains much of the estimated effect found using naive averaging methods.

We report the (weighted) average of betas of treated and control firms in Table 3. First, in

[11]Fama-French factor returns are obtained from Kenneth French's data library: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Figure 4: **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.



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 5.1.1, 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 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).

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

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.

5.2 Empirical Example 2: Index Inclusion

We next examine S&P 500 index inclusion announcements, analyzing both immediate announcement returns and pre-announcement price dynamics to test our theoretical predictions regarding identification in staggered event settings.

We first demonstrate that in staggered event settings, announcement-day bias is negligible because factor returns on event days average close to zero, particularly when compared to the large treatment effects of 3–4%. However, consistent with Greenwood and Sammon (2025), we document substantial pre-announcement drift. Synthetic control methods that match on pre-event returns nearly eliminate this drift. This pattern is consistent with selection on unobserved factors: firms added to the index differ systematically from control firms along dimensions not captured by observable factors, generating apparent pre-event "drift" that actually reflects factor model misspecification.

Empirical setting. Following Greenwood and Sammon (2025), we obtain index inclusion dates from Siblis Research and match tickers to CRSP PERMNOs using header information. Siblis provides announcement dates for S&P 500 additions. For the period September 1976 through September 1989, when announcement dates are missing, we exploit the institutional detail that index changes were announced after Wednesday market close and became effective the following day, allowing us to infer announcement dates.[12] We measure returns on the announcement date when it falls on a trading day; otherwise, we use the most recent prior trading day.

To assess whether event timing can be treated as random, we examine the distribution of factor returns on announcement days. Appendix Figure D.1, Panel A shows that the distribution of daily market returns on S&P 500 index inclusion announcement days is virtually indistinguishable from the distribution on non-announcement days. This pattern holds consistently across our entire sample period, from 1980–1989 through 2010–2020. The small-minus-big (SMB) factor exhibits similar distributional stability (Panel B). These results support treating announcement timing as conditionally random with respect to factor realizations, satisfying a key identification assumption for our short-horizon analysis.

Table 4 reports CAPM and Fama-French three-factor betas for firms added to the S&P 500 index, estimated using daily returns from days -250 to -100 relative to announcement.[13] We present results separately by decade from 1980 through 2020 to examine temporal variation in the characteristics of included firms.

Across all decades, the average market beta of included firms is approximately one. When treated firms have market betas near unity, the simple market-adjusted return (which implicitly assumes $\beta = 1$) yields similar results to the more sophisticated CAPM adjustment that estimates firm-specific betas. This convergence occurs because the bias term $(1 - \beta_i) \times r_{m,t}$ approaches zero when $\beta_i \approx 1$, consistent with the theoretical predictions in Theorem 2.1.

The combination of two empirical regularities—random event timing with respect to factor realizations and limited selection on factor loadings—suggests that short-horizon abnor-

[12] During this period, S&P followed a predictable schedule of announcing changes after Wednesday close for Thursday implementation.

[13] We exclude the immediate pre-announcement period to avoid contamination from potential information leakage.

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

mal return estimates should exhibit minimal bias regardless of the specific factor model employed. This prediction from Theorem 2.1 finds strong empirical support in Table 5, where announcement-day treatment effects are remarkably stable across estimation methods. The difference between simple market adjustment and sophisticated synthetic control methods is less than 0.2 percentage points in most decades, confirming that model specification has negligible impact on short-horizon estimates when the conditions of Theorem 2.1 are satisfied.

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 gsynth. The estimation window of factor loadings are from -250 to -101 before the announcement dates.

	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%

5.2.1 Pre-inclusion Drift

While Theorem 2.1 predicts negligible bias in short-horizon studies, it also implies that long-horizon estimates may suffer from substantial bias unless factor exposures are correctly specified. We now examine the "pre-announcement drift" documented by Greenwood and Sammon (2025), analyzing it decade by decade as a manifestation of potential long-horizon bias.

Interpreting pre-announcement price movements requires careful consideration of Assumption 2, our limited anticipation assumption. This assumption is particularly tenuous in the index inclusion setting for two reasons. First, market participants have incentives to anticipate market index changes. Second, as Greenwood and Sammon (2025) document, inclusion is partially predictable: firms with market capitalizations just below the S&P 500 cutoff face substantially higher inclusion probabilities than other firms. This predictability complicates the identification of treatment effects, as observed pre-announcement returns may reflect either genuine anticipation (violating Assumption 2) or selection on unobserved characteristics that drive both inclusion probability and returns.

To disentangle these effects, we pursue a two-pronged empirical strategy. First, we implement propensity score matching based on observable firm characteristics to account for selection on observables. Second, we employ synthetic control methods that match on pre-event returns, effectively controlling for unobserved factors that drive both selection and returns. The difference between these two approaches helps identify whether pre-announcement drift reflects anticipation or factor model misspecification.

Index inclusion predictability operates along two dimensions: the timing of additions (when inclusions occur) and the cross-section of selections (which firms are added). While ideally we would model both, we focus on cross-sectional predictability by estimating inclusion propensities based on observable firm characteristics. Specifically, we estimate annual logistic regressions:

$$\mathbf{1}(\text{Added})_{i,y,m} = \alpha_y + \beta_y \cdot \text{MktCapRank}_{i,y,m-1} + \varepsilon_{i,y} \tag{76}$$

where $\text{MktCapRank}_{i,y,m-1}$ is firm i's market capitalization rank at the end of month $m-1$, and inclusion occurs in month m of year y. Consistent with Greenwood and Sammon (2025), we find increasing predictability over time, with recent decades showing stronger relationships between lagged size and inclusion probability.

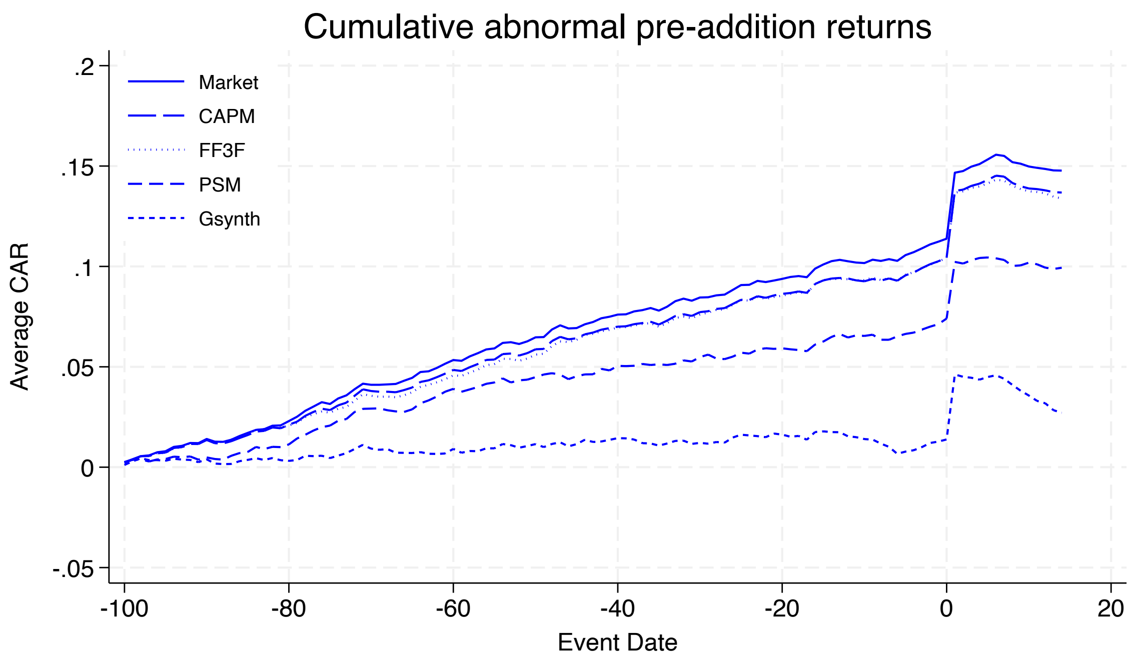
Using these propensity scores, we construct matched control groups via nearest-neighbor matching, creating portfolios of "pseudo-included" firms with similar inclusion probabilities but no actual inclusion. Under the assumption that selection between observationally equivalent firms is quasi-random, differences between included and pseudo-included firms should primarily reflect the causal effect of inclusion rather than selection bias.

To address selection on unobservables, we additionally implement the generalized synthetic control method of Xu and Liu (2022). For each announcement date, we estimate factor loadings using returns from days -250 to -101, deliberately excluding the immediate pre-announcement period where anticipation effects may contaminate estimation. We then construct synthetic control portfolios that match the pre-event return dynamics of included firms, examining the period from day -100 to -15.

This dual approach yields three distinct counterfactuals for cumulative abnormal returns (CARs): (i) simple market adjustment as in Greenwood and Sammon (2025) (we also do CAPM and FF3F adjustments for completeness, but do not subtract α for reasons that will be clear shortly) (ii) propensity score-matched pseudo-included firms that control for selection on observables, and (iii) synthetic controls that account for selection on unobserved factors. Comparing these counterfactuals from day -100 through the announcement date allows us to decompose pre-announcement drift into components attributable to observable characteristics versus unobserved factor exposures. If drift persists after propensity score matching but disappears with synthetic controls, this would suggest that unobserved factors—rather than anticipation based on observables—drive the pre-announcement returns.

First, we find 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. 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

Figure 5: **Cumulative abnormal pre-addition returns** 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.

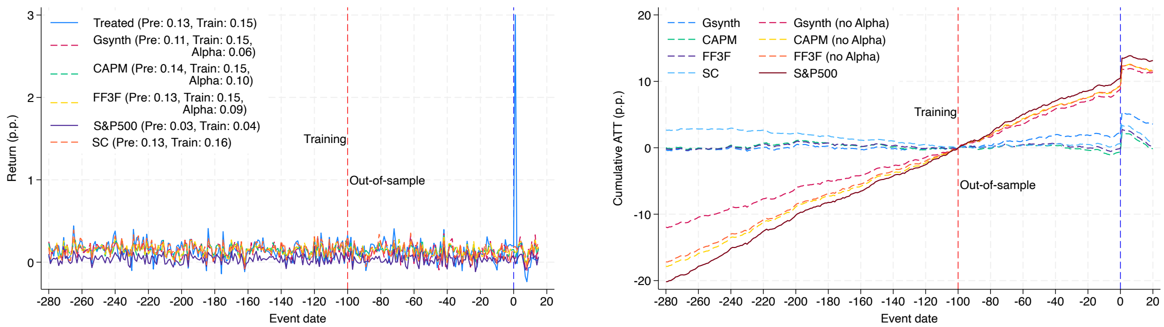


differences identified between these estimated methods? In Appendix Figure D.2, we show that there is a substantial drift in our known factors across most decades. Considering the positive loadings in Table 4, this suggests that the counterfactual return needs to sufficiently account for any and all potential unobserved factors driving the expected returns to avoid this bias highlighted in Theorem 2.1.

Figure 6: Per-Period and Cumulative ATT with factor models, gsynth, and synthetic methods

Panel A: Per-Period ATT

Panel B: Cumulative ATT



Capturing all potential unobserved factors is not easy, however. In Figure 6, Panel A, we plot the average return in event time for the treated group, and then our five counterfactuals. We report the average daily return for each group, and note a 0.13 p.p. daily return for the included firms (*prior* to inclusion), an unusually high daily return. In contrast, the S&P500 has a daily return of only 0.03 p.p. during this period. This suggests that the included firms are quite unusual. Recall that if our factor choice in a linear model sufficiently spans the risk factors, we should estimate an alpha of zero, even in the presence of positive average returns. Our counterfactual models do an excellent job of matching the average return in the training and pre-periods. However, for the CAPM and FF3F, more than half of the average predicted return comes from just the intercept, α. For gsynth, the estimated α is less than half, but still 0.06 p.p. Strikingly, the synthetic control counterfactual, which only takes a positive weighted average of control firms and does not include a constant, matches the pre-period return closely.

How should we interpret the estimated alpha in this linear factor models? There is presumably two components in an estimated factor model's alpha, true alpha, and model error:

$$\hat{\alpha} = \underbrace{\alpha}_{\text{true } \alpha} + \underbrace{\beta^{unobs} E(F_t^{unobs} | t \in \text{estimation window})}_{\text{Misspecifaction}}, \tag{77}$$

where the model misspecification captures the return premium over this period that is not included in our model. In this setting, we view true alpha as zero, especially 280 days prior to the inclusion event. As a result, the positive alpha likely suggests model misspecification. The implications of this misspecification depend on the stability of this misspecification term. In Panel B of fig. 6, we see that the inclusion of alpha ensures that the various linear factor models do as well the synthetic control method in removing almost all pre-inclusion drift. This suggests that the trend beforehand is not due to front-running, but instead differential return profiles for included stocks. However, failing to include alpha for the CAPM, FF3F and gsynth fail to remove the pre-inclusion drift.

5.3 Empirical Example 3: Mergers and Acquisitions

5.3.1 Empirical Setting

We examine acquirer returns around merger announcements using deal data from SDC Platinum. Following Malmendier (2018) and Savor and Lu (2009), we implement several sample restrictions to ensure clean identification. We require targets to be classified as "Public," "Private," or "Subsidiary" and restrict to completed deals with all-cash or all-stock payment structures, as mixed consideration complicates the interpretation of market-timing effects. To ensure economic materiality, we require the target's pre-announcement market value to exceed 5% of the acquirer's market capitalization. We exclude repurchases, self-tenders, and minority stake purchases by requiring deal types to be "Disclosed Dollar Value" or "Undisclosed Dollar Value," and mandating that acquirers hold less than 50% of the target six months before announcement.

We match acquirers to CRSP using six-digit CUSIPs, restricting to U.S. common shares (share codes 10 or 11) traded on NYSE, NASDAQ, or AMEX. Our event window spans days −280 to +250 relative to announcement. For the 20% of deals announced on non-trading days, we define $t = 0$ as the next trading day. Control firms comprise all CRSP-listed firms without contemporaneous merger announcements that have complete returns data over the event window. Our final sample contains 14,847 merger events across 6,625 unique dates, providing substantial variation in event timing for identification.

5.3.2 Short-Term Announcement Returns

To assess whether event timing can be treated as random, we examine the distribution of factor returns on announcement days. Appendix Figure D.1 shows that the distribution of daily market returns on S&P 500 index inclusion announcement days is virtually indistinguishable from the distribution on non-announcement days. The small-minus-big (SMB) factor exhibits similar distributional stability (unreported) These results support treating announcement timing as conditionally random with respect to factor realizations, satisfying a key identification assumption for our short-horizon analysis.

Table E.1 reports CAPM and Fama-French three-factor betas for firms with merger announcements, estimated using daily returns from days −250 to −100 relative to announcement.[14] We also examine how the betas change following the announcement as well, and

[14]We exclude the immediate pre-announcement period to avoid contamination from potential information leakage.

show that there are statistically significant changes after announcement, but they are small economically.

We first examine three-day announcement returns $[-1, +1]$ to test whether short-horizon estimates are robust to model specification, as predicted by Theorem 2.1. We compare two approaches: market-adjusted returns using the CRSP value-weighted index (including distributions) following Malmendier (2018), and gsynth estimates using the generalized synthetic control method of Xu (2017).

For the synthetic control approach, we estimate a separate model for each of the 6,625 event dates, treating all firms announcing mergers on that date (typically one or two firms) as the treatment group. We construct factor loadings using returns from days -280 to -31, excluding the immediate pre-announcement period to avoid contamination. Control firms consist of all CRSP securities without merger announcements that satisfy our data requirements.

Table 6 reports cumulative abnormal returns by target type and payment method. Consistent with our theoretical predictions, the difference between market-adjusted and synthetic control estimates is economically negligible—less than 10 basis points in most specifications. This robustness to model choice confirms that short-horizon merger announcement effects are identified regardless of the specific factor adjustment employed.

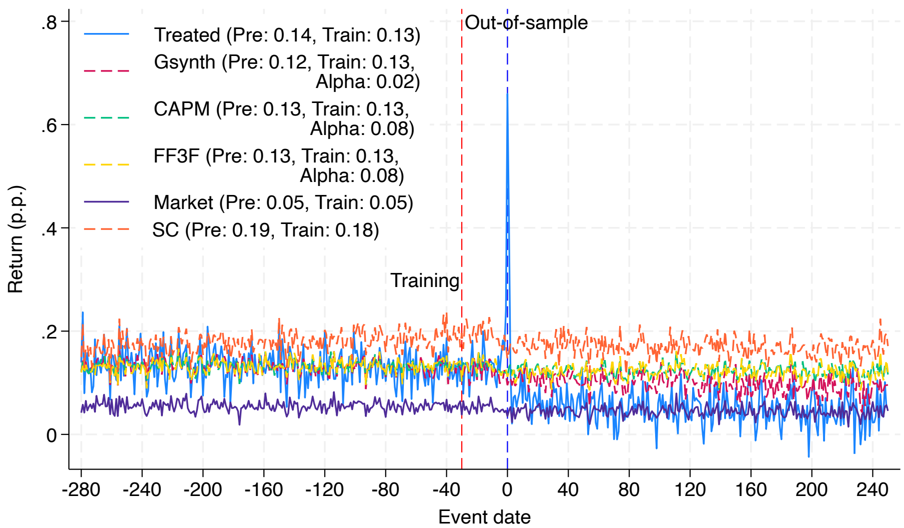
Table 6: Average three day cumulative p.p.

	Full sample	Public targets	Private targets	Other targets	Cash merger	Stock merger
Market mean	0.8	-1.2	1.3	1.6	1.1	0.4
Gsynth mean	0.7	-1.3	1.0	1.5	1.0	0.2
Count	14,847	3,297	7,030	4,520	9,261	5,592

In Figure 7, we plot the daily returns for the treated firms and the various control returns. Similar to the preannoucnment drift for the S&P index inclusion, the treated firm has significant daily returns prior in the year prior to the announcement, with 0.14 p.p. daily average returns (contrast with the market having 0.05 p.p. daily average returns). Again, the CAPM and FF3F models have significant alpha (0.08 p.p.), while gsynth has remarkably small alpha (0.02 p.p.). However, both gsynth and synthetic control do a poorer job matching the overall pre-period return, with 0.12 p.p. and 0.19p.p. returns respectively, relative to 0.14 for the treated group.

The path of the daily return line for the treated group spikes significantly on the event date, and then declines precipitously to a new steady state. It is worth remarking that within 30 days the event announcement, the treated firms' returns appear to line up almost exactly with the market returns. This is suggestive that there is a structural shift in the underlying

Figure 7: Per-Period Treated and Counterfactual Returns by event date. Event date= $[-120, 250]$



return performance to these acquiring firms, perhaps due to change in true alpha, or perhaps due to factor loadings.

In Figure 8, we see the long-run implications of these alphas in in that the difference counterfactual predictions have wildly different long-run cumulative ATT a year after the event. In the literature, the presence of a negative post-acquisition event is often pointed to as evidence in favor of Shleifer and Vishny (2003), but the modeling assumptions to make these types of assessments seem quite strong. This type of analysis also has implications for papers studying over- and under-reaction in the stock market.

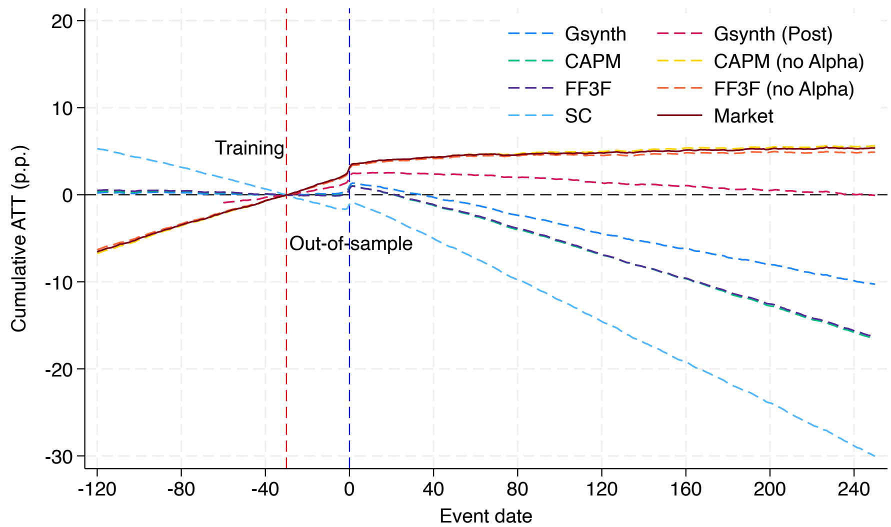
What can we do to deal with this specification error? A crucial alternative is to examine the effect in a setting where treatment is as-if randomly assigned.

5.4 Empirical Example 4: Close Merger Contests as a Quasi-Experimental Benchmark

Our previous empirical examples rely on model-based identification strategies that require correct specification of the factor structure. To validate these approaches, we now examine a setting with quasi-experimental variation: close merger contests where assignment to treatment (winning the contest) is plausibly random conditional on observables. Malmendier et al. (2018) show that in protracted bidding contests, winners and losers are ex ante similar firms competing for the same target, making losers natural counterfactuals for winners.

This setting provides a unique opportunity to assess the performance of different estima-

Figure 8: Cumulative ATT by event date. Event date= $[-120, 250]$



tors. Since losing bidders offer a design-based counterfactual, we can compare our model-based estimates (market adjustment, factor models, synthetic controls) against this quasi-experimental benchmark. Agreement between model-based and design-based estimates would validate our econometric approaches; divergence would suggest specification problems in the model-based methods.

5.4.1 Empirical Setting

Following Malmendier et al. (2018), we analyze close merger contests defined as those with above-median duration.[15] Protracted contests involving multiple rounds of bids and counterbids suggest that participants had similar ex ante winning probabilities, supporting the identifying assumption that contest outcomes are quasi-random.

We construct event-time at the monthly frequency, with $t = 0$ marking the month-end before the initial bid announcement. The pre-contest period spans months $t = -35$ to $t = 0$. The contest period ($t = 1$) encompasses all months from initial bid through completion, averaging 361 days in our sample. The post-merger period runs from $t = 2$ to $t = 36$. This structure accommodates contests of varying duration while maintaining a consistent event-time framework.

We match contest participants to CRSP monthly returns, filling missing observations with market returns following Malmendier et al. (2018). For synthetic control estimation, we

[15]We thank the authors for providing data on winning and losing bidders, announcement and completion dates, and contest duration.

augment the sample with all CRSP common shares (share codes 10 or 11) traded on NYSE, NASDAQ, or AMEX that have complete returns over the event window. This expanded control group allows the synthetic control algorithm to construct appropriate counterfactuals even when losing bidders may themselves be poor matches due to contest-specific shocks affecting all participants.

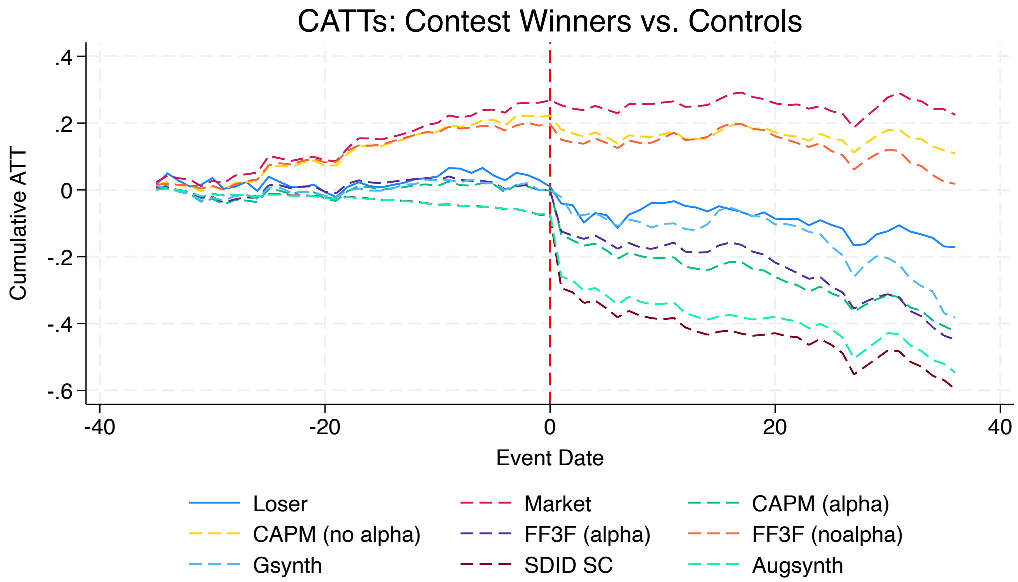
To check how well the as-if random counterfactual losing bidders match to winners, In Table 7, we compare the risk exposures on different factors of winners and losers. We see that these two groups are relatively similar, although the losers have slightly higher HML beta than winners.

Table 7: **Beta Distributions of Winners and Losers in Close Merger Contests** This table presents the average CAPM and Fama-French three-factor betas for winner and losers in close contest mergers. We estimate firm-level betas using daily stock returns using event month -35 to -1 before the start of the contest. We provide the mean and median of CAPM market beta and betas in Fama-French three-factor model. We test if the betas of treated and control firms are statistically different using a two-sided t-test with Welch (1947) approximation.

	Winner		Loser		Mean t-test
	Mean	Median	Mean	Median	Loser - Winner
CAPM Beta	1.148	0.922	1.114	0.968	-0.034
FF3F Mkt Beta	1.157	0.945	1.172	1.010	0.015
FF3F SMB Beta	0.369	0.291	0.232	0.168	-0.138
FF3F HML Beta	-0.055	-0.171	0.313	0.393	0.368*

In Figure 9, we plot the cumulative ATTs for the winners relative to our various controls. Our benchmark is the "Loser" control, in solid blue. We see that in the pre-period, there is reasonable balance between the two groups, and then a small but significant decline following the announcement, suggesting a negative effect. Notably, this counterfactual has the smallest and least trending of the different counterfactuals. The only alternative model-based portfolio that is meaningfully close is the Gsynth control.

Figure 9: Cumulative ATT by event date. Event month= $[-35, 36]$



It is worth noting that the synthetic control methods have much larger declines during the merger battle, and afterwards as well. The abnormal return models fit well in the pre-period, but then predict significant and continuing declines in cumulative returns. Overall, this evidence suggests that most of the models (especially abnormal return) do poorly in the longer run, although gsynth is the exception.

6 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 if the factor distribution is non-stationary. The key insight is that misspecification bias compounds over longer horizons if the average risk premium is different during the post-event window than in the pre-event window.

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 test the distribution of the events across time and 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.

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., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, *93*(1), 113–132.

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.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, *24*(2), 3–30.

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.

Barber, B. M., & Lyon, J. D. (1997a). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of financial economics*, *43*(3), 341–372.

Barber, B. M., & Lyon, J. D. (1997b). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, *43*(3), 341–372. https://doi.org/10.1016/S0304-405X(96)00890-2

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of financial economics*, *49*(3), 307–343.

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. (1987). Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research*, *25*(1), 1–48.

Boehmer, E., Musumeci, J., & Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, *30*(2), 253–272. https://doi.org/10.1016/0304-405X(91)90032-F

Brown, S. J., & Warner, J. B. (1985a). 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

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

Bryzgalova, S., Pelger, M., & Zhu, J. (2025). Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance*, *80*(5), 2447–2506.

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).

Cohn, J. B., Johnson, T. L., Liu, Z., & Wardlaw, M. (2025). Past is prologue: Inference from the cross section of returns around an event. *Available at SSRN 4296657*.

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

Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics*, *23*(2), 385–395. https://doi.org/10.1016/0304-405X(89)90064-0

Corrado, C. J., & Zivney, T. L. (1992). The specification and power of the sign test in event study hypothesis tests using daily stock returns. *Journal of Financial and Quantitative Analysis*, *27*(3), 465–478. https://doi.org/10.2307/2331331

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

Daniel, K., Mota, L., Rottke, S., & Santos, T. (2020). The cross-section of risk and returns. *The Review of Financial Studies*, *33*(5), 1927–1979.

Edmans, A. (2012). The link between job satisfaction and firm value, with implications for corporate social responsibility. *Academy of Management Perspectives*, *26*(4), 1–19.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306. https://doi.org/10.1016/S0304-405X(98)00026-9

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

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.

Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, *129*(7), 1947–1990.

Giglio, S., Xiu, D., & Zhang, D. (2025). Test assets and weak factors. *The Journal of Finance*, *80*(1), 259–319.

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

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, *54*(6), 2143–2184.

Huberman, G., Kandel, S., & Stambaugh, R. F. (1987). Mimicking portfolios and exact arbitrage pricing. *The Journal of Finance*, *42*(1), 1–9.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

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

Kolari, J. W., & Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *The Review of Financial Studies*, *23*(11), 3996–4025. https://doi.org/10.1093/rfs/hhq072

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

Kramer, L. A. (2001). Alternative methods for robust analysis in event study applications. *Advances in Investment Analysis and Portfolio Management*, *8*(1), 109–132.

Kwon, S. Y., & Tang, J. (2025). Extreme categories and overreaction to news. *Review of Economic Studies*, rdaf037.

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

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, 604–620.

Loughran, T., & Vijh, A. M. (1997). Do long-term shareholders benefit from corporate acquisitions? *The Journal of finance*, *52*(5), 1765–1790.

Lyu, T., & Xu, Z. W. (2024). Taking the road less traveled? market misreaction and firm innovation directions. *Market Misreaction and Firm Innovation Directions (August 30, 2024)*.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of economic literature*, *35*(1), 13–39.

Malmendier, U. (2018). Behavioral corporate finance. In *Handbook of behavioral economics: Applications and foundations 1* (pp. 277–379). Elsevier.

Malmendier, U., Moretti, E., & Peters, F. S. (2018). Winning by losing: Evidence on the long-run effects of mergers. *The Review of Financial Studies*, *31*(8), 3212–3264.

Mitchell, M. L., & Stafford, E. (2000). Managerial decisions and long-term stock price performance. *The Journal of Business*, *73*(3), 287–329.

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

Rau, P. R., & Vermaelen, T. (1998). Glamour, value and the post-acquisition performance of acquiring firms. *Journal of financial economics*, *49*(2), 223–253.

Ross, S. A. (2013). The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 11–30). World Scientific.

Roth, J., & Sant'Anna, P. H. (2023). When is parallel trends sensitive to functional form? *Econometrica*, *91*(2), 737–747.

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. (1986). Do demand curves for stocks slope down? *The Journal of Finance*, *41*(3), 579–590.

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

Welch, B. L. (1947). The generalization of 'student's'problem when several different population varlances are involved. *Biometrika*, *34*(1-2), 28–35.

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 Proofs

Proof of Proposition 1. Let $n_s = \#\{i : T_i = s\}$ and $n_c = \#\mathcal{C}$ denote the number of treated and control securities, respectively. Throughout the proof we maintain Assumptions 1 and 2 and impose the following regularity conditions:

(i) For each i, t, the idiosyncratic component ε_{it} satisfies

$$\mathbb{E}(\varepsilon_{it} \mid \mathbf{F}_t, \mathbf{X}_i, T_i) = 0, \quad \mathbb{E}(\varepsilon_{it}^2) < \infty.$$

(ii) Within each cohort s and the control group $\mathcal{C}$, $\{\varepsilon_{it}\}$ are independent across i (for fixed t) with uniformly bounded second moments, so that the law of large numbers applies to cross-sectional averages.

(iii) The usual OLS regularity conditions hold for the pre-treatment regressions used to construct the abnormal-returns estimator (e.g. fixed K, non-singular regressor covariance matrix, etc.).

Define the cohort- and control-group average idiosyncratic shocks

$$\varepsilon_{st} \equiv \frac{1}{n_s} \sum_{i:T_i=s} \varepsilon_{it}, \qquad \varepsilon_{\infty t} \equiv \frac{1}{n_c} \sum_{i \in \mathcal{C}} \varepsilon_{it}.$$

By (i)–(ii), for each fixed t,

$$\varepsilon_{st} \xrightarrow{p} 0 \quad \text{and} \quad \varepsilon_{\infty t} \xrightarrow{p} 0 \quad \text{as } n_s, n_c \to \infty.$$

Recall the algebraic decompositions in equations (29)–(31):

$$\begin{aligned}
\tau^{AR}(s,t) - \tau^{ATT}(s,t) &= (\alpha_s - \hat{\alpha}_s) + (\beta_s \mathbf{F}_t - \hat{\beta}_s \mathbf{F}_t^o) + \varepsilon_{st}, \\
\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) &= (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\mathbf{F}_t + (\varepsilon_{st} - \varepsilon_{\infty,t}), \\
\hat{\tau}^{alt}(s,t) - \tau^{ATT}(s,t) &= (\alpha_s - \hat{\alpha}_s^{alt}) + (\beta_s - \hat{\beta}_s^{alt})\mathbf{F}_t + \varepsilon_{st},
\end{aligned}$$

where $\hat{\alpha}_s, \hat{\beta}_s$ are the cohort averages of the OLS estimates from the abnormal-returns model and $(\hat{\alpha}_s^{alt}, \hat{\beta}_s^{alt})$ denote the implied factor loadings from an alternative estimator (synthetic control or GSC) at the cohort level.

(1) Probability limits of the three estimators. *Abnormal returns estimator.* By definition, $\tilde{\alpha}_s$ and $\tilde{\beta}_s$ are the probability limits of the cohort-average OLS coefficients:

$$\hat{\alpha}_s \xrightarrow{p} \tilde{\alpha}_s, \qquad \hat{\beta}_s \xrightarrow{p} \tilde{\beta}_s \quad \text{as } T_{pre} \to \infty,$$

where the limit is the linear projection of $R_{it}(\infty)$ onto $\mathbf{F}_t^o$ in the pre-treatment window $\{t < s - \delta\}$.[16] Combining this with the fact that $\varepsilon_{st} \xrightarrow{p} 0$ as $n_s \to \infty$ yields, from (29),

$$\tau^{AR}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} (\alpha_s - \tilde{\alpha}_s) + (\beta_s \mathbf{F}_t - \tilde{\beta}_s \mathbf{F}_t^o),$$

which is equation (32).

Difference-in-means estimator. The difference-in-means estimator does not involve any pre-event estimation, so T_{pre} is irrelevant here. Using (30) and the fact that $\varepsilon_{st} - \varepsilon_{\infty,t} \xrightarrow{p} 0$ as $(n_s, n_c) \to \infty$, we obtain

$$\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\mathbf{F}_t,$$

which is equation (33).

Synthetic control estimator. For the synthetic control estimator, we specialize $\hat{\tau}^{alt}(s,t)$ in (31) to $\hat{\tau}^{synth}(s,t)$ and denote the implied loadings by $(\hat{\alpha}_s^{synth}, \hat{\beta}_s^{synth})$.

Under Assumption 1 together with a standard interactive fixed-effects structure for the untreated potential outcomes,

$$R_{it}(\infty) = \alpha_i + \boldsymbol{\beta}_i' \mathbf{F}_t + \varepsilon_{it},$$

the average untreated return for cohort s can be written as

$$\mathbb{E}(R_{it}(\infty) \mid T_i = s) = \alpha_s + \boldsymbol{\beta}_s' \mathbf{F}_t,$$

with an analogous representation for each control unit $j \in \mathcal{C}$.

Ferman (Ferman, 2021) shows that, under such a factor structure, with sufficiently many pre-treatment periods and control units, the synthetic control weights constructed by minimizing the pre-treatment mean squared error recover the factor loadings of the treated unit (here, the treated cohort) in probability. Formally, applying their Theorem 1 to the cohort-level treated unit $R_{s,t}$, we obtain

$$\hat{\alpha}_s^{synth} \xrightarrow{p} \alpha_s, \qquad \hat{\beta}_s^{synth} \xrightarrow{p} \beta_s \quad \text{as } n_c, T_{pre} \to \infty.$$

[16] Limited anticipation (Assumption 2) guarantees that $R_{it} = R_{it}(\infty)$ for $t < T_i - \delta$, so pre-event returns identify the no-event process.

Combining this with $\varepsilon_{st} \xrightarrow{p} 0$ and (31) yields

$$\hat{\tau}^{synth}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0,$$

which is equation (34).

Remark (Gsynth). The same logic applies to the generalized synthetic control estimator in Definition 6. Under the interactive fixed-effects model

$$R_{it}(\infty) = \alpha_i + \boldsymbol{\lambda}_i'\mathbf{F}_t + \varepsilon_{it}$$

and the regularity conditions in Xu (2017), Xu shows that the counterfactual returns $\hat{R}_{it}^{GS}(\infty)$ are consistent for $R_{it}(\infty)$, uniformly over post-treatment periods. Aggregating over i within cohort s then implies that the cohort-period ATT estimated by Gsynth is also consistent:

$$\hat{\tau}^{GS}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0.$$

(2) Consistency of the difference-in-means estimator under random assignment. Assume now the random assignment condition in Assumption 3:

$$p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{F}),$$

so that T_i is independent of $\mathbf{X}_i = (\alpha_i, \boldsymbol{\beta}_i)$. Then the distribution of $\mathbf{X}_i$ is the same in every treatment cohort s and in the never-treated group $\mathcal{C}$. In particular,

$$\alpha_s = \mathbb{E}(\alpha_i \mid T_i = s) = \mathbb{E}(\alpha_i \mid i \in \mathcal{C}) = \alpha_\infty,$$

and similarly $\beta_s = \beta_\infty$.

Substituting these equalities into the probability limit in (33) gives

$$\hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0.$$

Note that this result only relies on large n_s, n_c; it does not require $T_{pre} \to \infty$ because the difference-in-means estimator does not use pre-event estimation.

(3) Consistency of the abnormal returns estimator under correct specification. Finally, suppose that the factor model is correctly specified in the abnormal-returns regression, i.e. $\mathbf{F}_t^o = \mathbf{F}_t$ for all t. For each security i,

$$R_{it}(\infty) = \alpha_i + \boldsymbol{\beta}_i'\mathbf{F}_t + \varepsilon_{it}, \quad t < T_i - \delta.$$

By Assumption 2, these pre-event observations coincide with the no-event potential outcome, and standard OLS consistency arguments imply that, as $T_{pre} \to \infty$,

$$\hat{\alpha}_i \xrightarrow{p} \alpha_i, \qquad \hat{\boldsymbol{\beta}}_i \xrightarrow{p} \boldsymbol{\beta}_i.$$

Averaging within cohort s and applying the law of large numbers as $n_s \to \infty$ gives

$$\hat{\alpha}_s \equiv \frac{1}{n_s} \sum_{i:T_i=s} \hat{\alpha}_i \xrightarrow{p} \frac{1}{n_s} \sum_{i:T_i=s} \alpha_i \xrightarrow{p} \alpha_s,$$

and analogously $\hat{\beta}_s \xrightarrow{p} \beta_s$. Hence, under correct specification,

$$\tilde{\alpha}_s = \alpha_s, \qquad \tilde{\beta}_s = \beta_s.$$

Substituting these equalities into (32) yields

$$\tau^{AR}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0,$$

so the abnormal returns estimator is consistent for $\tau^{ATT}(s,t)$ when the factor structure is correctly specified.

This completes the proof. $\square$

Proof of Theorem 2.1. Throughout, maintain Assumptions 1 and 2 and the auxiliary regularity conditions used in the proof of Proposition 1 (mean-zero idiosyncratic shocks with a law of large numbers across i, and standard OLS regularity for the abnormal-returns regressions, plus the spanning/interactive fixed effects conditions for synthetic control and gsynth).

Recall that for any estimator $\star \in \{AR, cont, synth, GS\}$ and horizon $\kappa \geq 0$,

$$\theta_\kappa^{ATT} = \sum_{s\in\mathcal{S}} w_s \tau^{ATT}(s, s+\kappa), \qquad \hat{\theta}_\kappa^\star = \sum_{s\in\mathcal{S}} w_s \hat{\tau}^\star(s, s+\kappa),$$

with weights $w_s = N_s / \sum_{s'\in\mathcal{S}} N_{s'}$. Hence

$$\hat{\theta}_\kappa^\star - \theta_\kappa^{ATT} = \sum_{s\in\mathcal{S}} w_s \big(\hat{\tau}^\star(s, s+\kappa) - \tau^{ATT}(s, s+\kappa)\big). \tag{78}$$

(1) Unbiasedness of synthetic control and gsynth. From Proposition 1, for each fixed cohort s and period t,

$$\hat{\tau}^{alt}(s,t) - \tau^{ATT}(s,t) \xrightarrow{p} 0 \quad \text{for } alt \in \{synth, GS\}$$

as $n_s, n_c, T_{pre} \to \infty$ under the conditions of Ferman (2021) (for synthetic control) and Xu (2017) (for gsynth). Setting $t = s + \kappa$ and substituting into (78) yields

$$\hat{\theta}_\kappa^{alt} - \theta_\kappa^{ATT} = \sum_{s \in \mathcal{S}} w_s \big(\hat{\tau}^{alt}(s, s+\kappa) - \tau^{ATT}(s, s+\kappa) \big).$$

Since $\mathcal{S} \subseteq \{1, \ldots, T\}$ is finite and the weights satisfy $0 \leq w_s \leq 1$ and $\sum_s w_s = 1$, a finite linear combination of terms that converge in probability to zero also converges to zero. Thus,

$$\hat{\theta}_\kappa^{alt} - \theta_\kappa^{ATT} \xrightarrow{p} 0,$$

which proves part (1).

(2) Bias of abnormal-returns and difference-in-means estimators. Assume $|\mathcal{S}| > 0$ and $1 > p_t(\mathbf{X}_i, \mathbf{F}) > \epsilon > 0$. The lower bound ϵ guarantees that each event time in $\mathcal{S}$ occurs with positive probability in the population, so N_s and $\sum_{s'} N_{s'}$ both diverge with N and the cohort weights converge:

$$w_s = \frac{N_s}{\sum_{s' \in \mathcal{S}} N_{s'}} \xrightarrow{p} \pi_s \equiv \Pr(T_i = s \mid T_i \in \mathcal{S}).$$

From Proposition 1, for each fixed s and t,

$$\begin{aligned} \tau^{AR}(s,t) - \tau^{ATT}(s,t) &\xrightarrow{p} (\alpha_s - \tilde{\alpha}_s) + (\beta_s \mathbf{F}_t - \tilde{\beta}_s \mathbf{F}_t^o), \\ \hat{\tau}^{cont}(s,t) - \tau^{ATT}(s,t) &\xrightarrow{p} (\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\mathbf{F}_t. \end{aligned}$$

Evaluating at $t = s + \kappa$ and plugging into (78), we obtain

$$\begin{aligned} \hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} &\xrightarrow{p} \sum_{s \in \mathcal{S}} \pi_s \Big[(\alpha_s - \tilde{\alpha}_s) + \big(\beta_s \mathbf{F}_{s+\kappa} - \tilde{\beta}_s \mathbf{F}_{s+\kappa}^o \big) \Big], \\ \hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT} &\xrightarrow{p} \sum_{s \in \mathcal{S}} \pi_s \Big[(\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\mathbf{F}_{s+\kappa} \Big]. \end{aligned}$$

Define a random event time S with $\Pr(S = s \mid T_i \in \mathcal{S}) = \pi_s$. Then the limits above can be written compactly as conditional expectations over treated cohorts:

$$\begin{aligned} \hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} &\xrightarrow{p} \mathbb{E}\left[(\alpha_S - \tilde{\alpha}_S) + \big(\beta_S \mathbf{F}_{S+\kappa} - \tilde{\beta}_S \mathbf{F}_{S+\kappa}^o \big) \,\middle|\, T_i \in \mathcal{S} \right], \\ \hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT} &\xrightarrow{p} \mathbb{E}\left[(\alpha_S - \alpha_\infty) + (\beta_S - \beta_\infty)\mathbf{F}_{S+\kappa} \,\middle|\, T_i \in \mathcal{S} \right]. \end{aligned}$$

Relabeling S as s inside the expectation gives the expressions stated in part (2).

(3) Random assignment across firms. Under random assignment across firms,

$$p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{F}),$$

so event assignment is independent of $\mathbf{X}_i = (\alpha_i, \beta_i)$. As in the proof of Proposition 1, this implies

$$\alpha_s = \alpha_\infty, \qquad \beta_s = \beta_\infty \quad \text{for all } s \in \mathcal{S}.$$

Substituting these equalities into the limit for $\hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT}$ in part (2) gives

$$\hat{\theta}_\kappa^{cont} - \theta_\kappa^{ATT} \xrightarrow{p} 0$$

as $n_s, n_c \to \infty$, even for fixed T_{pre}. This proves part (3).

(4) Random timing. Now assume random timing,

$$p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{X}_i),$$

so that event timing is independent of the factor path $\mathbf{F}$, and adopt the standard assumption that firm characteristics $\mathbf{X}_i$ (hence α_i, β_i) are independent of $\mathbf{F}$.

Start from the general bias expression for the abnormal-returns estimator in part (2):

$$\hat{\theta}_\kappa^{ar} - \theta_\kappa^{ATT} \xrightarrow{p} \mathbb{E}\left[(\alpha_s - \tilde{\alpha}_s) + \left(\beta_s \mathbf{F}_{s+\kappa} - \tilde{\beta}_s \mathbf{F}^o_{s+\kappa}\right) \,\middle|\, T_i \in \mathcal{S}\right].$$

Under random timing and independence between $(\alpha_s, \beta_s, \tilde{\alpha}_s, \tilde{\beta}_s, s)$ and the factor process $\mathbf{F}$, we can factor the cross terms:

$$\mathbb{E}\big(\beta_s \mathbf{F}_{s+\kappa} \mid T_i \in \mathcal{S}\big) = \mathbb{E}\big(\beta_i \mid T_i \in \mathcal{S}\big)\, \mathbb{E}(\mathbf{F}_{s+\kappa} \mid \text{post}),$$
$$\mathbb{E}\big(\tilde{\beta}_s \mathbf{F}^o_{s+\kappa} \mid T_i \in \mathcal{S}\big) = \mathbb{E}\big(\tilde{\beta}_i \mid T_i \in \mathcal{S}\big)\, \mathbb{E}(\mathbf{F}^o_{s+\kappa} \mid \text{post}),$$

where $\mathbb{E}(\cdot \mid \text{post})$ denotes the expectation over post-event periods.

Now we use the OLS population intercept identity. The estimated intercept $\tilde{\alpha}_i$ from regressing R_{it} on $\mathbf{F}^o_t$ in the pre-event window satisfies

$$\tilde{\alpha}_i = \mathbb{E}(R_{it} \mid \text{pre}) - \tilde{\beta}_i \mathbb{E}(\mathbf{F}^o_t \mid \text{pre}).$$

Since $\mathbb{E}(R_{it} \mid \text{pre}) = \alpha_i + \beta_i \mathbb{E}(\mathbf{F}_t \mid \text{pre})$ under Assumption 1, we obtain

$$\alpha_i - \tilde{\alpha}_i = -\beta_i \mathbb{E}(\mathbf{F}_t \mid \text{pre}) + \tilde{\beta}_i \mathbb{E}(\mathbf{F}^o_t \mid \text{pre}). \tag{79}$$

Substituting (79) and the factored expectations into the bias expression:

$$\begin{aligned}\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} \mathbb{E}\big[-\beta_i \mathbb{E}(\mathbf{F}_t \mid \text{pre}) + \tilde{\beta}_i \mathbb{E}(\mathbf{F}^o_t \mid \text{pre}) \mid T_i \in \mathcal{S}\big] \\ + \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\, \mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\tilde{\beta}_i \mid T_i \in \mathcal{S})\, \mathbb{E}(\mathbf{F}^o_t \mid \text{post}) \\ = \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\mathbf{F}_t \mid \text{pre})\big] \\ - \mathbb{E}(\tilde{\beta}_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}^o_t \mid \text{post}) - \mathbb{E}(\mathbf{F}^o_t \mid \text{pre})\big].\end{aligned}$$

Case (a): Stationary factors. If factors have constant means over time, i.e., $\mathbb{E}(\mathbf{F}_t \mid \text{post}) = \mathbb{E}(\mathbf{F}_t \mid \text{pre})$ and $\mathbb{E}(\mathbf{F}^o_t \mid \text{post}) = \mathbb{E}(\mathbf{F}^o_t \mid \text{pre})$, then both bracketed terms vanish and

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} 0.$$

Case (b): Non-stationary factors. If factor means drift between the pre-event and post-event windows, the asymptotic bias is

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\mathbf{F}_t \mid \text{pre})\big] - \mathbb{E}(\tilde{\beta}_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}^o_t \mid \text{post}) - \mathbb{E}(\mathbf{F}^o_t \mid \text{pre})\big]. \tag{80}$$

When the observed factors $\mathbf{F}^o_t$ are stationary but the true factors $\mathbf{F}_t$ include omitted components with drifting means, only the first term contributes to the bias.

For the difference-in-means estimator, starting from part (2):

$$\hat{\theta}^{cont}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} \mathbb{E}\left[(\alpha_s - \alpha_\infty) + (\beta_s - \beta_\infty)\mathbf{F}_{s+\kappa} \,\middle|\, T_i \in \mathcal{S}\right].$$

Under random timing, we can factor the second term:

$$\mathbb{E}\big((\beta_s - \beta_\infty)\mathbf{F}_{s+\kappa} \mid T_i \in \mathcal{S}\big) = \mathbb{E}(\beta_s - \beta_\infty \mid T_i \in \mathcal{S})\, \mathbb{E}(\mathbf{F}_t),$$

so

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

Note that unlike the abnormal-returns estimator, the difference-in-means estimator does not benefit from the intercept identity because α_∞ is a population parameter of the control group, not an OLS-estimated quantity. The bias vanishes only if $\alpha_s = \alpha_\infty$ and $\beta_s = \beta_\infty$, which requires random assignment (part 3), not merely random timing.

This establishes part (4).

(5) Random timing with non-stationary factors. Maintain the random timing assumption from part (4), so that $p_t(\mathbf{X}_i, \mathbf{F}) = p_t(\mathbf{X}_i)$ and firm characteristics are independent of the factor path. Now suppose that factor means may differ between the pre-event estimation window and the post-event period.

From the derivation in part (4), the asymptotic bias of the abnormal-returns estimator is

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\mathbf{F}_t \mid \text{pre})\big]$$
$$- \mathbb{E}(\tilde{\beta}_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}^o_t \mid \text{post}) - \mathbb{E}(\mathbf{F}^o_t \mid \text{pre})\big].$$

When the observed factors $\mathbf{F}^o_t$ are stationary but the true factor structure includes omitted factors with drifting means, the second term vanishes and the bias simplifies to

$$\hat{\theta}^{ar}_{\kappa} - \theta^{ATT}_{\kappa} \xrightarrow{p} \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\mathbf{F}_t \mid \text{pre})\big],$$

where $\mathbf{F}_t$ includes both observed and omitted factors and β_i denotes the corresponding full vector of loadings.

For cumulative effects over horizon H, the bias accumulates linearly:

$$\hat{\theta}^{CATT}_{H} - \theta^{CATT}_{H} \xrightarrow{p} H \cdot \mathbb{E}(\beta_i \mid T_i \in \mathcal{S})\big[\mathbb{E}(\mathbf{F}_t \mid \text{post}) - \mathbb{E}(\mathbf{F}_t \mid \text{pre})\big],$$

since the drift term applies to each post-event period.

This establishes part (5). $\square$

Proof of Lemma 1. Throughout, we work with the second–order Taylor expansion of $\log(1+x)$ around $x = 0$,

$$\log(1 + x) = x - \tfrac{1}{2}x^2 + r(x), \quad \text{with} \quad r(x) = O(x^3),$$

and omit the remainder term $r(x)$ for notational simplicity. All equalities below should be read as holding up to these higher–order terms in returns.

Step 1: Period–by–period geometric ATT. Fix an event cohort s and calendar period t. By definition,

$$\tau^{geo,ATT}(s,t) = E\big(\log(1 + R_{it}(s)) - \log(1 + R_{it}(\infty)) \mid T_i = s\big).$$

Using the second–order expansion,

$$\log(1 + R_{it}(s)) \approx R_{it}(s) - \tfrac{1}{2}R_{it}(s)^2,$$
$$\log(1 + R_{it}(\infty)) \approx R_{it}(\infty) - \tfrac{1}{2}R_{it}(\infty)^2,$$

so that

$$\tau^{geo,ATT}(s,t) \approx E\left[\big(R_{it}(s) - R_{it}(\infty)\big) - \tfrac{1}{2}\big(R_{it}(s)^2 - R_{it}(\infty)^2\big) \,\Big|\, T_i = s\right].$$

Let the individual treatment effect be

$$\tau_i(s,t) = R_{it}(s) - R_{it}(\infty),$$

so that $R_{it}(s) = R_{it}(\infty) + \tau_i(s,t)$. Then

$$R_{it}(s)^2 - R_{it}(\infty)^2 = \big(R_{it}(\infty) + \tau_i(s,t)\big)^2 - R_{it}(\infty)^2 = 2R_{it}(\infty)\tau_i(s,t) + \tau_i(s,t)^2.$$

Substituting this into the expression above,

$$\begin{aligned}
\tau^{geo,ATT}(s,t) &\approx E\left[\tau_i(s,t) - \tfrac{1}{2}\big(2R_{it}(\infty)\tau_i(s,t) + \tau_i(s,t)^2\big) \,\Big|\, T_i = s\right] \\
&= E\big(\tau_i(s,t) \mid T_i = s\big) - E\left(R_{it}(\infty)\tau_i(s,t) + \tfrac{1}{2}\tau_i(s,t)^2 \,\Big|\, T_i = s\right).
\end{aligned}$$

By definition of the arithmetic cohort–period ATT,

$$\tau^{ATT}(s,t) = E\big(\tau_i(s,t) \mid T_i = s\big),$$

so we have the key period–by–period relationship

$$\tau^{geo,ATT}(s,t) \approx \tau^{ATT}(s,t) - E\left(R_{it}(\infty)\tau_i(s,t) + \tfrac{1}{2}\tau_i(s,t)^2 \,\Big|\, T_i = s\right). \tag{81}$$

Step 2: From period ATT to horizon H. For cohort s, the geometric ATT over horizon H is

$$\tau^{geo,ATT}(s,H) = \sum_{\kappa=0}^{H} \tau^{geo,ATT}(s, s+\kappa),$$

and the corresponding arithmetic CATT is

$$\tau^{CATT}(s,H) = \sum_{\kappa=0}^{H} \tau^{ATT}(s, s+\kappa).$$

Summing (81) over $\kappa = 0, \ldots, H$ gives

$$\tau^{geo,ATT}(s,H) \approx \sum_{\kappa=0}^{H} \tau^{ATT}(s, s+\kappa) - \sum_{\kappa=0}^{H} E\left(R_{i,s+\kappa}(\infty)\tau_i(s, s+\kappa) + \tfrac{1}{2}\tau_i(s, s+\kappa)^2 \,\Big|\, T_i = s\right).$$

Now average across event cohorts with weights w_s:

$$\theta_H^{geo,ATT} = \sum_s w_s \tau^{geo,ATT}(s,H), \qquad \theta_H^{ATT} = \sum_{\kappa=0}^{H} \theta_\kappa^{ATT} = \sum_s w_s \sum_{\kappa=0}^{H} \tau^{ATT}(s,s+\kappa).$$

Thus,

$$\begin{aligned}
\theta_H^{geo,ATT} &\approx \sum_s w_s \sum_{\kappa=0}^{H} \tau^{ATT}(s,s+\kappa) \\
&\quad - \sum_s w_s \sum_{\kappa=0}^{H} E\left(R_{i,s+\kappa}(\infty)\tau_i(s,s+\kappa) + \tfrac{1}{2}\tau_i(s,s+\kappa)^2 \,\middle|\, T_i = s \right) \\
&= \theta_H^{ATT} - \sum_s w_s \sum_{\kappa=0}^{H} E\left(R_{i,s+\kappa}(\infty)\tau_i(s,s+\kappa) + \tfrac{1}{2}\tau_i(s,s+\kappa)^2 \,\middle|\, T_i = s \right),
\end{aligned}$$

which is the first expression in Lemma 1.

Step 3: Independence and simplification. Now impose the additional assumption stated in the lemma: for all s and κ,

- $R_{i,s+\kappa}(\infty)$ and $\tau_i(s,s+\kappa)$ are independent conditional on $T_i = s$; and
- the conditional mean of the no–event return is constant,

  $$\mu = E\big(R_{i,s+\kappa}(\infty) \mid T_i = s\big)$$

  does not depend on s or κ.

Then

$$E\big(R_{i,s+\kappa}(\infty)\tau_i(s,s+\kappa) \mid T_i = s\big) = \mu\, E\big(\tau_i(s,s+\kappa) \mid T_i = s\big) = \mu\, \tau^{ATT}(s,s+\kappa),$$

and the expression from Step 2 becomes

$$\begin{aligned}
\theta_H^{geo,ATT} &\approx \theta_H^{ATT} - \sum_s w_s \sum_{\kappa=0}^{H} \Big[\mu\, \tau^{ATT}(s,s+\kappa) + \tfrac{1}{2} E\big(\tau_i(s,s+\kappa)^2 \mid T_i = s\big) \Big] \\
&= \theta_H^{ATT} - \mu \sum_{\kappa=0}^{H} \sum_s w_s \tau^{ATT}(s,s+\kappa) - \tfrac{1}{2} \sum_{\kappa=0}^{H} \sum_s w_s E\big(\tau_i(s,s+\kappa)^2 \mid T_i = s\big).
\end{aligned}$$

Using $\sum_s w_s \tau^{ATT}(s, s+\kappa) = \theta_\kappa^{ATT}$ and $\sum_{\kappa=0}^H \theta_\kappa^{ATT} = \theta_H^{ATT}$, we get

$$\theta_H^{geo,ATT} \approx (1-\mu)\theta_H^{ATT} - \tfrac{1}{2}\sum_{\kappa=0}^{H}\sum_s w_s E\big(\tau_i(s, s+\kappa)^2 \mid T_i = s\big).$$

To rewrite the last term in terms of variances, consider a randomly drawn treated security i and define the individual treatment effect at event time κ as

$$\Delta_{i,\kappa} \equiv \tau_i(s, s+\kappa) \quad \text{for the (random) cohort } s = T_i.$$

Under the cohort weights w_s, the distribution of s among treated units satisfies

$$\Pr(s = r \mid T_i \in \mathcal{S}) = w_r,$$

so

$$\begin{aligned} E(\Delta_{i,\kappa} \mid T_i \in \mathcal{S}) &= \sum_s w_s E\big(\tau_i(s, s+\kappa) \mid T_i = s\big) = \sum_s w_s \tau^{ATT}(s, s+\kappa) = \theta_\kappa^{ATT}, \\ E(\Delta_{i,\kappa}^2 \mid T_i \in \mathcal{S}) &= \sum_s w_s E\big(\tau_i(s, s+\kappa)^2 \mid T_i = s\big). \end{aligned}$$

Denote the cross–sectional variance of individual treatment effects at event time κ by

$$\mathrm{var}(\theta_\kappa^{ATT}) \equiv \mathrm{var}\left(\Delta_{i,\kappa} \mid T_i \in \mathcal{S}\right).$$

Then

$$E(\Delta_{i,\kappa}^2 \mid T_i \in \mathcal{S}) = \mathrm{var}(\theta_\kappa^{ATT}) + \big(\theta_\kappa^{ATT}\big)^2,$$

so that

$$\sum_s w_s E\big(\tau_i(s, s+\kappa)^2 \mid T_i = s\big) = \mathrm{var}(\theta_\kappa^{ATT}) + \big(\theta_\kappa^{ATT}\big)^2.$$

Substituting into the expression for $\theta_H^{geo,ATT}$ yields

$$\theta_H^{geo,ATT} \approx (1-\mu)\theta_H^{ATT} - \tfrac{1}{2}\sum_{\kappa=0}^{H}\Big[\mathrm{var}(\theta_\kappa^{ATT}) + \big(\theta_\kappa^{ATT}\big)^2\Big],$$

which is the second expression in Lemma 1.

This completes the proof. $\square$

B 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 B.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 4.1. 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.

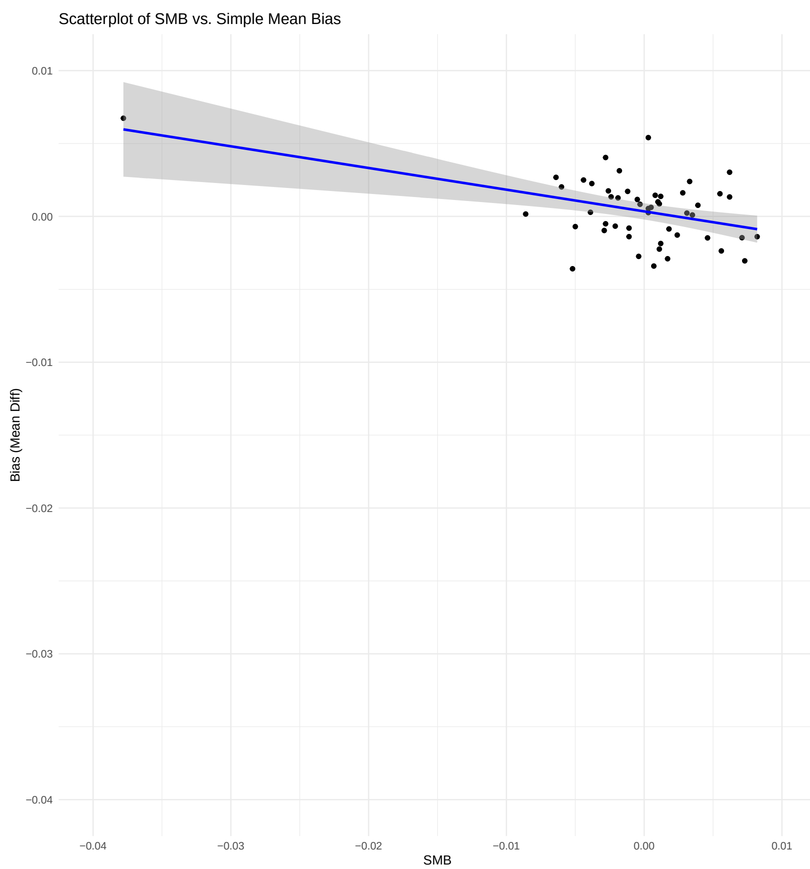
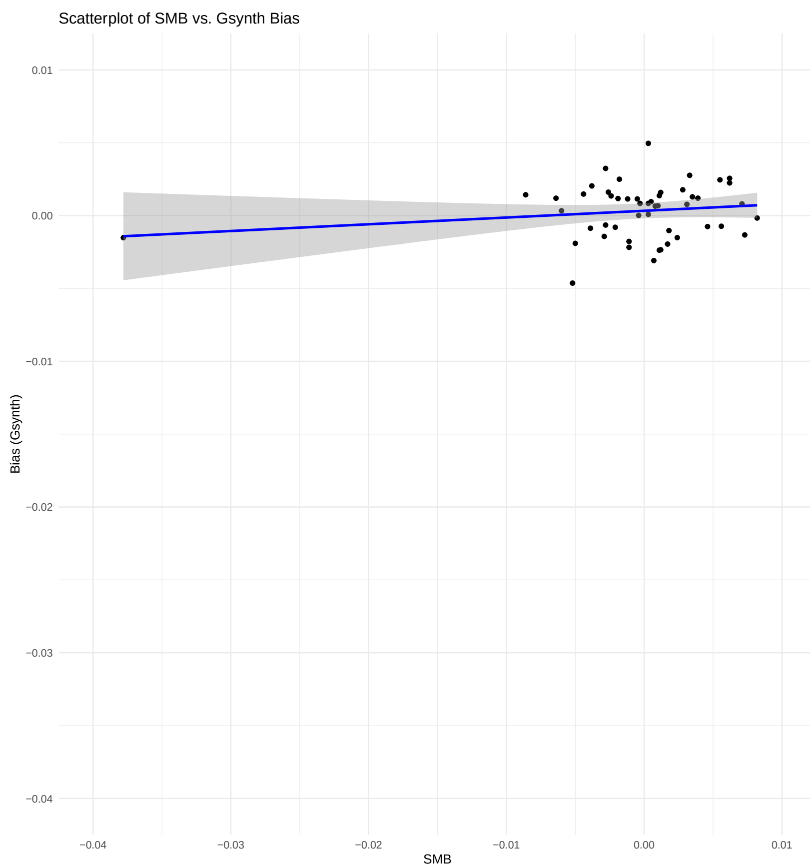


Figure B.2: Bias from Gsynth Model on SMB Returns with Assignment Selection

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 4.1. 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.



C Additional Results for Geithner

This section presents addition results for the Acemoglu et al. (2016) empirical example.

C.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 C.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 C.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 C.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

C.2 Placebo Period ATT

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

	(1) Average	(2) DID	(3) SC	(4) 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 C.6: Placebo Period ATT to Geithner Announcement (New York connections)

	(1) Average	(2) DID	(3) SC	(4) 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

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

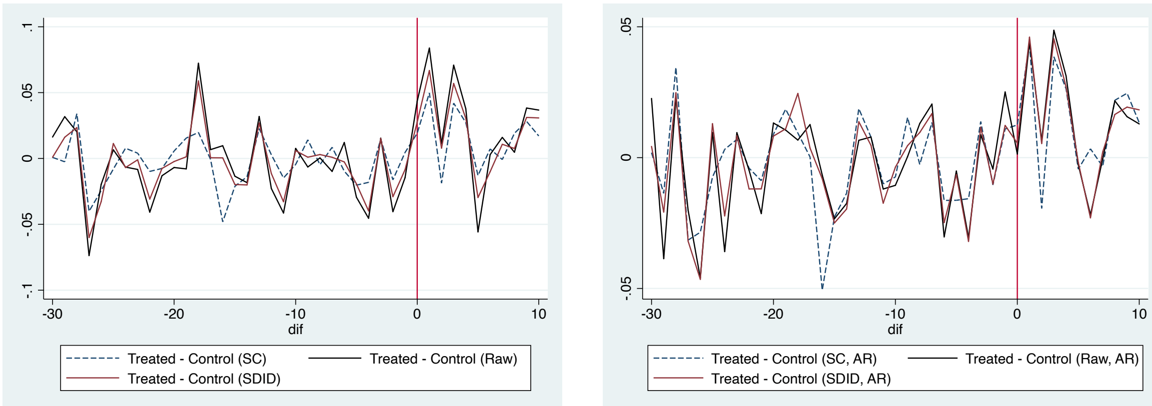
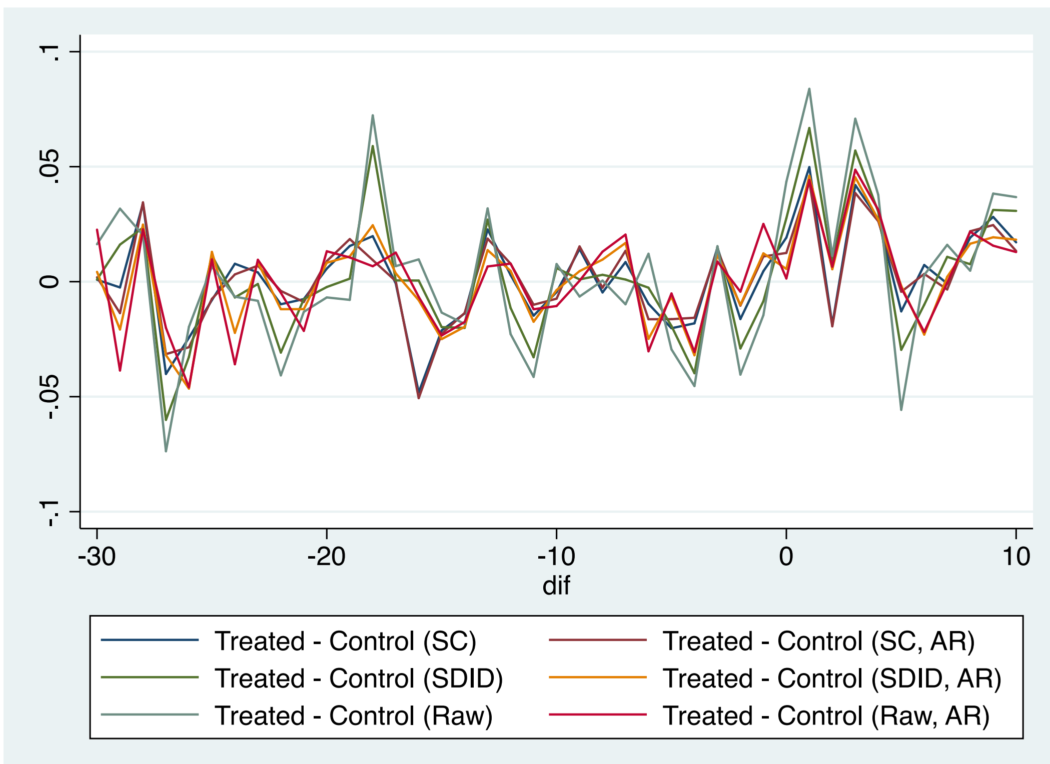


Figure C.2: Period-by-Period ATT in Placebo and Post Period (Schedule connections)



C.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 C.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.

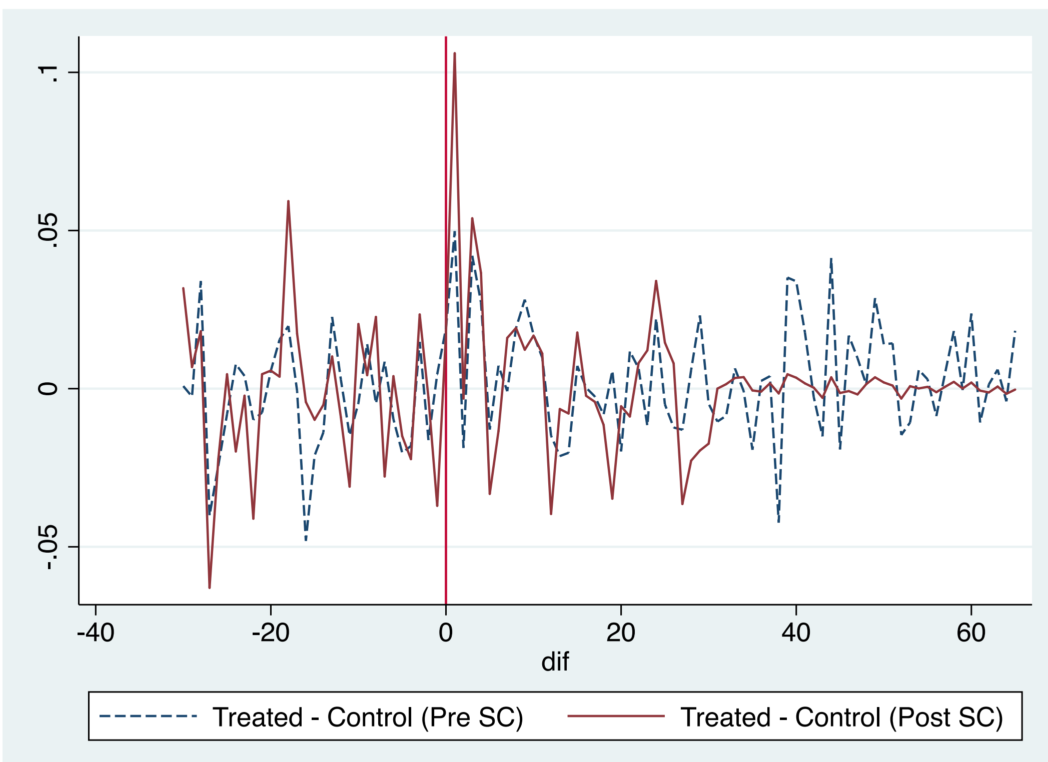
Table C.7: Pre-/Post-Event Market Beta from CAPM

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



C.5 Beta: All Public Firms as Control

Table C.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 C.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

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

Table C.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 C.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 C.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

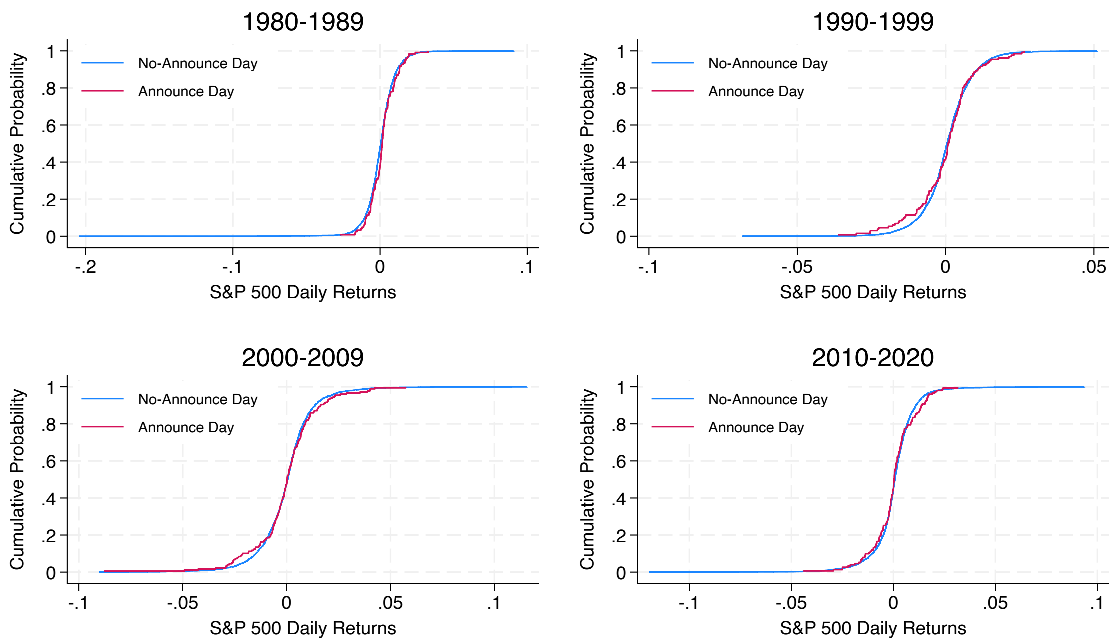
* p<0.10, ** p<0.05, *** p<0.01

D Additional Results for Index Inclusion

This section presents addition results for the S&P index inclusion empirical example.

Figure D.1: **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

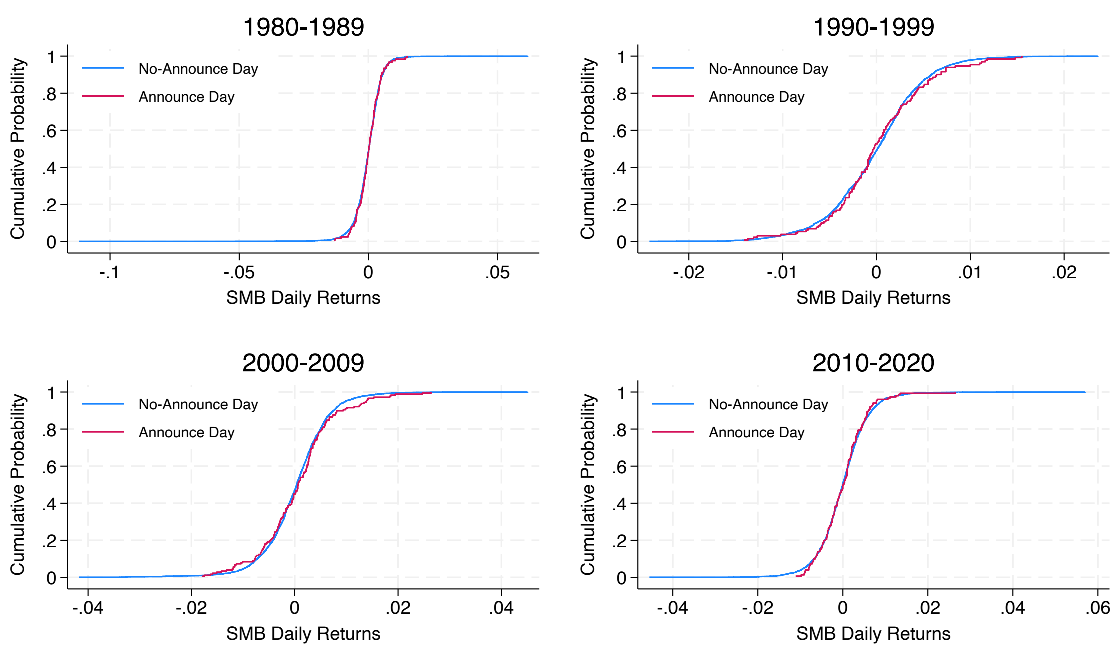
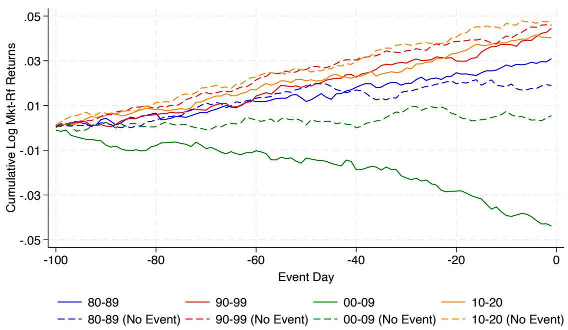


Figure D.2: **Pre-addition Cumulative Market Factor Returns (Inclusion vs. Randomized No-Inclusion Days)** This figure plots the average cumulative returns on the market and the SMB factor following index inclusion announcements in event time, averaged across inclusions for each decade. We also plot 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

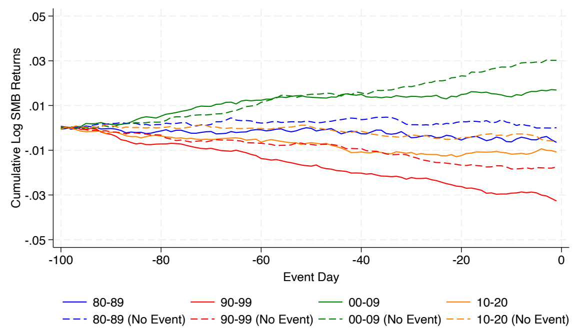


Figure D.3: Cumulative pre-addition market-adjusted returns (Treated vs. propensity score matched)

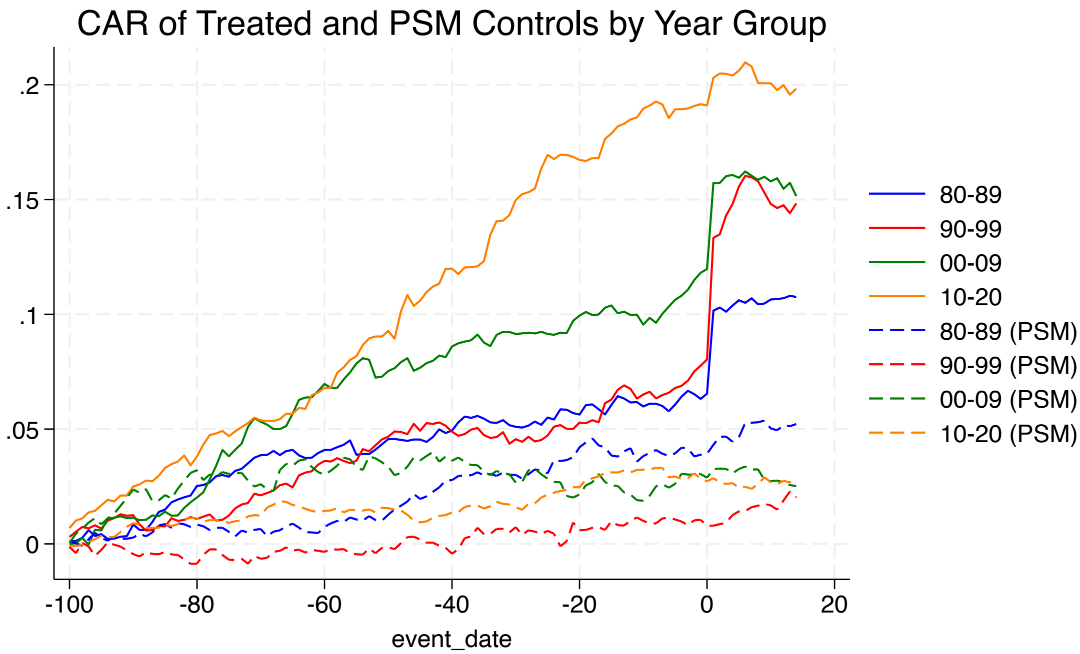
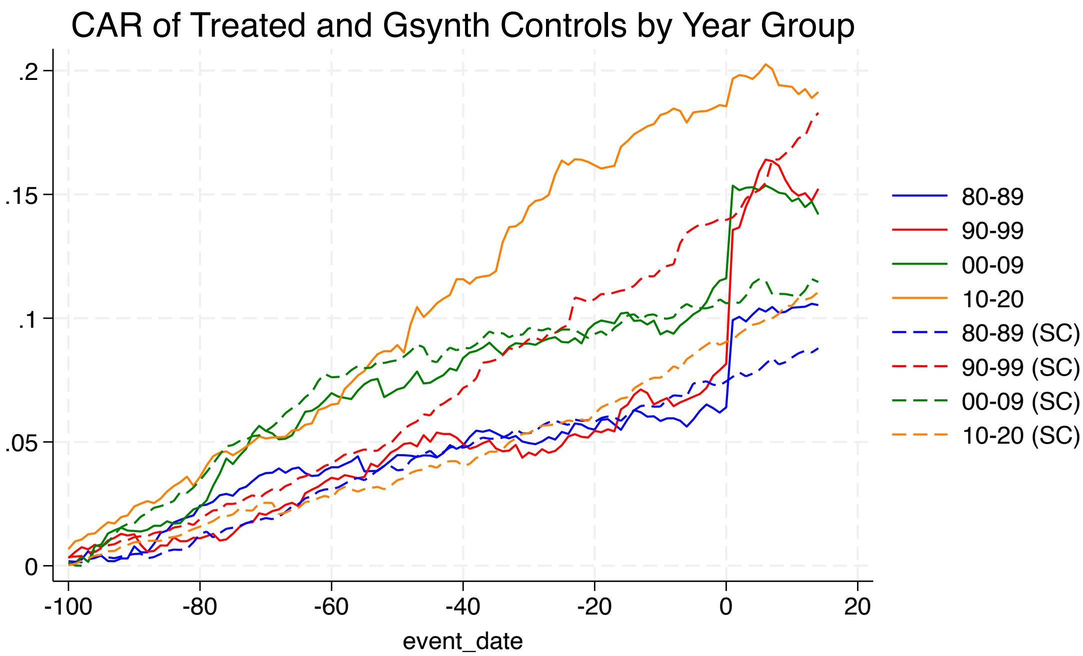


Figure D.4: Cumulative pre-addition market-adjusted returns (Treated vs. synthetic method)



E Additional results on empirical example 3, merger announcements

This section presents additional exhibits and results for our first mergers announcement example.

Figure E.1: **Cumulative Distributions of Factor Returns by Announcement Status** This figure plots the daily returns of the CRSP value-weighted index on the dates when there are merger 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 a merger annnouncment event.

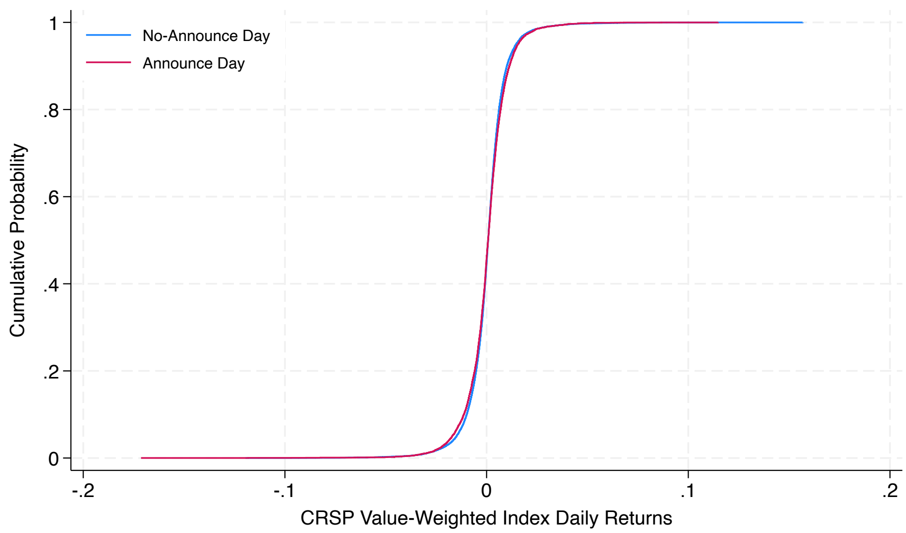


Table E.1: **Beta Distributions of Acquirers in Merger** This table presents the average CAPM and Fama-French three-factor betas for acquirers in merger transactions. We estimate firm-level betas using daily stock returns before and after the announcement. For pre betas, we use event date -280 to -30, while for post beta, we use event date 30-280 to estimate betas. We provide the mean and median of CAPM market beta and betas in Fama-French three-factor model. We test if the pre and post betas are statistically different using a two-sided t-test.

	Pre		Post		Mean t-test
	Mean	Median	Mean	Median	Pre - Post
CAPM Beta	0.952	0.909	0.968	0.926	-0.016***
FF3F Mkt Beta	1.020	0.993	1.022	1.001	-0.002
FF3F SMB Beta	0.690	0.620	0.678	0.613	0.012*
FF3F HML Beta	0.101	0.141	0.132	0.179	-0.031***

F Additional results on empirical example 4, close mergers with winners and losers

This section presents additional exhibits and results for our second mergers announcement example.

F.0.1 Estimation Window and Model Stability

We estimate different counterfactual models with different estimation window lengths. We vary the estimation window from 35 months as default to 12 month. We again leave the period $t = 0$ as the placebo period. We plot the average treatment effects in the estimation window, placebo period, treatment period, and post treatment windows, respectively.

Figure F.1: **Average Treatment Effects by Estimation Window Length** This figure plots the average treatment effects of winners in merger contests in the estimation, placebo, treatment, and post-treatment windows, by the length of the estimation window. We estimate betas and train synthetic control and gsynth models using pre-announcement periods from 12 months to 35 months. We then compute the counterfactual returns from different models with the design-based loser portfolio.