The long-run stock market performance of mergers and acquisitions

Sudheer Chava[†], Ágoston Reguly[†*]

[†] Georgia Institute of Technology

[*] Corvinus University of Budapest

Click for Latest Version.

February 15, 2024

Abstract

We investigate the long-run stock market performance of acquirer firms using a modified staggered synthetic control approach. Our methodology matches on multiple acquirer characteristics before mergers and acquisitions (M&As), with weights optimized to balance differences during pre-M&A time periods. Based on post-merger excess returns over three years, we find that, on average, M&As are neither value adding nor value destroying. Our heterogeneity analysis based on merger characteristics finds that larger acquirers tend to have higher three-year returns. Our methodology has applications for other long-run corporate event studies.

JEL: C31, C33, G14, G34

Keywords: M&A, long-run, synthetic control, matching, stock market performance

[*]Corresponding author

Email addresses: sudheer.chava@scheller.gatech.edu (S. Chava), agoston.reguly@uni-corvinus.hu (Á. Reguly).

1 Introduction

Do acquiring firms create or destroy stockholder value? In the last forty years, there were more than 250,000 Merger and acquisition (M&A) events in the US alone, accounting for almost 30 trillion dollars in transaction value[1]. M&A are some of the most important events in the life of a company and can have a significant impact on the firm's operations and activities. These transactions are fundamental to the interests of the leadership, board of directors, employees, investment banks, and regulators.

Finding an appropriate counterfactual for an acquirer firm to compute the long-run abnormal stock returns of the acquirer firms, a measure of the M&A value to stockholders, is challenging. Long-run abnormal returns may be biased due to unobserved differences between the firms that merge and the firms that do not (see, e.g., Loughran and Vijh 1997, Lyon et al. 1999, Bessembinder and Zhang 2013 or Malmendier et al. 2018). For instance, a decline in the acquirer's market valuation after a merger might not be caused by the merger but because highly valued firms choose to acquire less highly valued targets. If so, the following decline would have occurred even without acquisition.[2] Furthermore, this selection problem of acquirer firms at a selected point may accumulate over the long-run horizon, leading to large differences in the empirical findings.[3]

This paper proposes a novel method to create a control firm to understand acquiring firms' long-run stock market performance. As the firm characteristics of each M&A firm differ in multiple dimensions, we use a set of candidate firms for each event firm and weigh them so they match the characteristics of the event firm before the event occurs. We tailor the synthetic control methodology to create these synthetic-event firms that are the basis of our comparisons.[4] We also weigh the pre-event periods to get a better quality match closer

[1] In 2022, there were more than 8.3 thousand M&As with more than 1.25 trillion dollars in transaction value ($\sim 5\%$ of US GDP)

[2] This argument appears in multiple papers, such as Shleifer and Vishny (2003), Rhodes-Kropf and Viswanathan (2004), Dong et al. (2006), Savor and Lu (2009) or Rhodes-Kropf et al. (2005).

[3] The long-run event study literature faces the "bad-model" or "bad-comparison" problem, which produces conflicting results for the same type of event based on different samples and different approaches to modeling long-run stock returns. See e.g. Kothari and Warner (2007), Brav et al. (2000), Eckbo et al. (2000), Eckbo et al. (2007), Loughran and Ritter (2000), Liu et al. (2023).

[4] For instance, ABX Air Inc. bought Cargo Holdings International, a provider of outsourced air cargo services, for $350 million in 2007. In this deal, ABX Air Inc. said they wanted to expand their presence in the Americas, Asia, and Europe; hence, they bought Cargo Holdings International, whose two largest customers are BAX/Schenker, a logistics company owned by German rail operator Deutsche Bahn, and the U.S. government. To create a synthetic, ABX Air Inc., we used 29 publicly listed firms from the US and weighted them so that they match ABX Air Inc.'s size, book-to-market ratio, momentum, return-on-assets (ROA), and asset growth up to one year before the event. The firms used to create the synthetic firm are such as Bluelinx Holdings Inc (distributes specialty and commodity building products – with weight of 4.4%) or Northwestern Corp. (utility company providing electricity and natural gas – weight is 3,7%) and 27 more.

to the event. We show that each created "synthetic" firm has statistically indistinguishable firm characteristics from its event firm. Based on these weights, we calculate the excess returns on the three-year horizon and find that *on average* there is no positive or negative market return in the last 40 years. Although the overall effect is zero, we find considerable amounts of heterogeneity across the M&A events that allow us to analyze the time evolution of excess returns and test some of the popular hypotheses, such as the change in the market conditions, prior over-or undervaluation of the acquirer, whether the target firm is private or public, competition for the target firm, form of payments, effect of leverage or unanticipated integration costs.

Our paper contributes to the rich collection of academic studies investigating firm performance around and after mergers. The evaluation of the market performance of the acquirer firm requires that researchers define normal or benchmark returns. Following Renneboog and Vansteenkiste (2019), we may classify papers into two broad categories how they define abnormal returns. The first type directly searches for a firm that is the basis for stock market performance comparison, whereas the second category uses some pricing model to calculate the "normal" returns for each event firm.

The first type is based on the similarity of firm characteristics, such as size and market-to-book ratio; or acquisition characteristics, such as cancellation or bidding contest. Comparing acquirer firms' stock returns to a control group based on financial variables has been a popular method in the past few decades (see, e.g., Loughran and Ritter 1995, or Bessembinder and Zhang 2013, Kolari et al. 2021). There is, more or less, an agreement on which financial variables to use as primarily: market capitalization (as firm size) and market-to-book ratio.[5] The advantage of this approach is that it does not restrict its sample to specific types of M&As, thus results have higher external validity to M&A events in general. The disadvantage is the well-known bad comparison problem of using bad controls (see e.g., Kothari and Warner 2007, Bessembinder and Zhang 2013[6]).

Specific circumstances related to M&As have also been employed to select controls. For instance, Davidson et al. (1989) uses cancellation of mergers, Malmendier et al. (2018) compares post-merger returns for the winner of a bidding contest to a looser firm, while Boyson et al. (2017) focuses on failed bids of hedge fund activism. The main strength of these papers is the higher internal validity of the comparison group, whereas the investigated set of mergers is somewhat limited. In our paper, we do not restrict our attention to special

[5]Table A2 presents different matching methods used for corporate events.

[6]Bessembinder and Zhang (2013) shows poor matching quality, which highlights mismatch for market beta, firm size, book-to-market ratio, momentum, idiosyncratic risk, illiquidity, and investment variables. Bessembinder and Zhang (2013) and Kolari et al. (2021) argue how to control for such differences in a regression framework.

M&A events to create controls but articulate which assumptions are required to create valid controls to avoid bad comparisons and inducing bias in our estimate.

The second category consists of articles that obtain alpha coefficients from regressing event firm returns[7] on market-wide factor models such as the capital asset pricing model (Sharpe 1964, Lintner 1965), Fama-French three- or five-factor models (Fama 1998, Fama and French 2015), the four-factor model proposed by Hou et al. (2015) or other portfolio constructions such as Daniel et al. (1997), Eckbo et al. (2000), or Bessembinder et al. (2018) In this paper, we discuss some obstacles to utilize such pricing models to create benchmark returns and compare them with the actual returns of acquirer firms in the *long term* while interpreting it as a causal effect. Briefly, our main concern is the bad-control bias (Cinelli et al., 2022): these factors include information on variables that have been affected by the acquisition after some lagged periods.[8]

The paper also contributes to the methodologies of event study designs in finance. We propose a framework that bridges commonly used matching techniques in corporate finance. We cover "classical" matching method from M&A literature, which uses one period before the event and matches on certain variables; the time-series cross-sectional matching, proposed by Imai et al. (2019); and we propose a stacked synthetic control method. We show the differences in the identification assumptions and the possible bridges between the methods. We briefly discuss our method's connection to the fast-growing difference-in-differences literature.

Our proposed stacked synthetic control method contributes to the synthetic-control methodology (see, e.g., Abadie et al., 2010, Abadie et al., 2015, Abadie, 2021, Arkhangelsky et al., 2021, Ben-Michael et al., 2022, Porreca, 2022, Abadie and L'hour, 2021, Cattaneo et al., 2021, Cattaneo et al., 2023). We modify the synthetic difference-in-difference method Arkhangelsky et al. (2021) to fit the context of M&As: match on *multiple pre-event variables* and weight candidate firms and pre-event periods. Similarly to Ben-Michael et al. (2022), Porreca (2022),

[7]Not only M&As, but used to evaluate other corporate events as well.

[8]For example, Bessembinder et al. (2018) proposes a characteristic-based benchmark return approach, where the authors specifically incorporate firm characteristics to control for pre-event characteristics and to analyze the post-event performance of the (event) firms. The paper introduces two sets of firm characteristics: 'C5' contains the firm size, book-to-market ratio, momentum, ROA, and asset growth. The second, broader set of characteristics, 'C14', includes C5 variables and market beta, accrual, dividend, cumulative 2-year returns, idiosyncratic risk, illiquidity, turnover, leverage, and sales over price. Based on these variables, Bessembinder et al. (2018) construct benchmark returns (which they call 'CBBR5' and 'CBBR14') by subtracting the benchmark return from the realized return. They employ a Fama-MacBeth regression (Fama and MacBeth, 1973) along with pooled OLS to regress on event dummies, and the results are interpreted as abnormal stock returns after specific events. If one uses a monthly panel on firms, even if the benchmark returns are lagged (typically, a one-month lag is used), e.g., the book-to-market ratio will change due to price changes that have been affected by the merger. If the long-term performance is measured in the three-year horizon, the lagged factor values will contain these effects, resulting in bad-control bias.

and Cattaneo et al. (2023), we extend the synthetic control method to a staggered setup. Ben-Michael et al. (2022) allows for matching on multiple pre-event characteristics but does not allow for varying time-related weights. We find it important in our case to allow for pre-event period weighting due to two reasons: i) the acquirer firms are present in the sample with different time horizons, and some of them are not in the sample for one year before the event happens; ii) the matching quality is poorer if we do not allow pre-event period weights. Incorporating these facts, one can see our contribution as we allow for event-specific time-related weights while creating the firm-specific weights based on multiple covariates.[9] Cattaneo et al. (2023) propose principled prediction intervals to quantify the uncertainty of a large class of synthetic control predictions. Our method relates to their general framework as an applied procedure allowing event-specific time-related weights. Porreca (2022) and Clarke et al. (2023) generalize Arkhangelsky et al. (2021) to a staggered setup with time and individual weights. Our paper extends their approach by creating weights using multiple variables.

Our main result shows that, on average, M&As are neither value-added nor value-destroying. The proposed stacked synthetic control method with different specifications shows zero average excess returns on the 36-month horizon. If we use other methods to evaluate the long-run market performance, we can find negative effects showing under-performance. With the classical matching method, we find a negative 11 basis points significant at 5%, while time-series cross-sectional matching gives different values based on the specification, between -51 and -33 basis points. In a simulation study, we verify that our stacked synthetic control method has superior properties and finds the null or positive effects better than its competitors. We also compare our method to the factor or benchmark-based return methods (see, e.g., Bessembinder et al. 2018) and combine it with the difference-in-differences estimator proposed by Callaway and Sant'Anna (2021). Under the assumptions of these models, we find negative average excess returns between 2.1 and 47 basis points. However, these values are only significant for the Fama-MacBeth estimation.[10]

As a last exercise, we use cumulative abnormal returns (CARs) from our stacked synthetic control method to analyze heterogeneity and test some popular hypotheses. First, we investigate if there are any periods where negative CARs are realized, but we find no

[9]From another angle, our staggered setup is a case of Ben-Michael et al. (2022) procedure as we do the synthetic procedure on each event individually. In their terminology, our method is a special case of 'separate SCM'.

[10]To be more specific, we find significant values in the case of raw returns and Fama-French 5 factor model-based benchmark returns with pooled OLS. None of the parameters are significant with the Callaway and Sant'Anna (2021) estimator; however, the negative effects are stable across different specifications around the -10 and -13 basis points.

significantly different periods from zero. However, when we condition on investor sentiment, using the measure proposed by Baker and Wurgler (2006), we find suggestive evidence that during cold market conditions (investor sentiment is below median), there is a cumulative ~ 1 basis point negative effect. We also test some of the most frequent hypotheses, following Malmendier et al. (2018). First, we find no evidence that acquirers' overvaluation via Tobin-q would significantly lower market performance. Second, we test if the target firm is private or public, which would lead to significant differences in CARs. Despite the common argument of private information used for private targets or liquidity discounts, we find no support for these theories in our case. Also, we investigate bidding competition among potential acquirers (number of bidders) and find no significant differences in the long-run performance, similarly to Eckbo et al. (2018).

Finally, we have revisited some hypotheses regarding the possible mechanisms governing CARs. Cash holdings offer more flexibility than any other less liquid asset and may provide some advantages, whereas loss of flexibility plays a negative role. We use the type of payments (all cash, all stocks, or mixed) along with a leverage measure of the acquirer firm to investigate these questions. We find no evidence that any of these variables affect CARs, even though the sign of the coefficients is aligned with the theory. The last hypothesis refers to integration costs, often cited as a key reason for poor post-merger performance. We considered relatedness (in terms of industry) and acquirer absolute and relative size to target as different factors. We find no evidence of differences in CARs conditional on whether the acquirer and target are from the same industry or not but show suggestive evidence on the acquirer's absolute and relative size: the larger the acquirer firm, the CARs tend to be higher over the three-year horizon.

The structure of the paper is as follows. Section 2 describes the data. Section 3 outlines the framework and the three methods that we use to analyze the long-run performance of the acquirer firms. At the end of this section, we also show the connections between these methods and their relation to other methods used in the potential outcome literature. Section 4 shows the empirical strategies, the matching quality, the main results, and the robustness checks. Section 5 contains the heterogeneity analysis, and section 6 concludes.

2 Data

We focus on acquirer firms located in the US and events that happened between 1980 and 2019. In order to evaluate the long-run performance of these firms, we used returns and other market-based variables from the Center for Research in Security Prices (CRSP), as well as quarterly accounting data for firms from Compustat fundamentals. We extend the time window by 12 months backward and 36 months forward, thus firms have records between 1979 January and 2022 December. This extension is due to two reasons: i) extending 12 extensions backward allows our matching methods to find good quality matches up to 12 months before the M&A announced; ii) extending the time period forward helps to study post-event return horizons in 36 months that broadly match the existing empirical practice to consider long-run as three years (see, e.g., Bessembinder and Zhang, 2013, Malmendier et al., 2018 or Bessembinder et al., 2018).

2.1 Mergers and Acquisitions

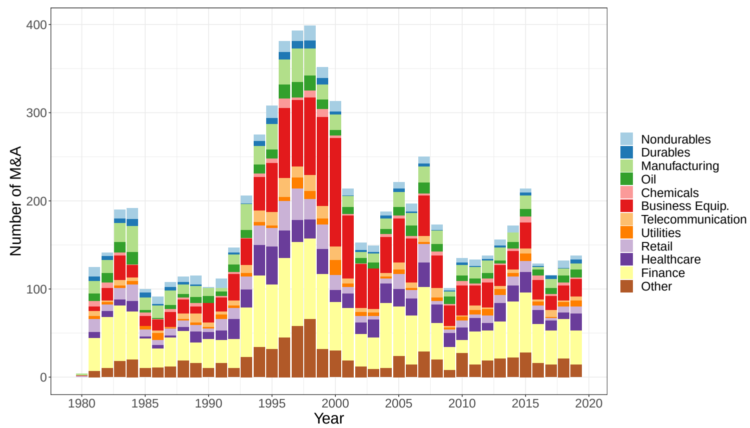
We consider mergers and acquisitions from the SDC Platinum database. We follow Netter et al. (2011), Phillips and Zhdanov (2013), Ewens et al. (2018) and Bessembinder et al. (2018) to select the pool of mergers and acquisitions. We filter for disclosed and undisclosed deals where the deal is completed, and the percentage of shares acquired in the transaction grants majority in the firm (the percentage of shares acquired in the transaction is more than 50% and the percentage of shares held by the acquirer six months prior to the announcement is less than 49%). The number of unique events that meet these criteria is 223,780. Following Bessembinder et al. (2018), we select deals with types of merger (SDC form "M"), acquisition of majority interest ("AM"), acquisition of remaining interest ("AR") or acquisition of partial interest ("AP").

Furthermore, we exclude small transactions without material impact on the acquirer: the transaction value must be greater than $5 million or more than 5% of the acquirer's market capitalization before the deal announcement. This restricted sample contains 11,033 mergers and acquisitions. We require that the event firm has not done merger or acquisition activities in the past 36 months to avoid overlapping event bias. Finally, we require that acquirer firms have records in CRSP and Compustat. After merging the events with these databases[11], we have 7,312 events. Figure 1. shows the number of M&As for each year, with the Fama-French 12 industry categorization.

To highlight the heterogeneity in mergers and acquisitions, Table 2. shows descriptive statistics on the characteristics of the M&As. One can see i) the imposed restrictions (e.g., shares owned before are capped at 49% and shares owned after having a minimum of 50%. Moreover, the minimum transaction value is 5.05 million dollars); ii) there is a significant amount of heterogeneity in the events captured by the standard deviation. It might be

[11] For merging SDC with Compustat and CRSP, we have used (historical) CUSIP identifier and company names. We have also used Ewens et al. (2018) for merging M&A events from SDC to Compustat. The used resource is available from Michael Ewens' github page. Merging CRSP and Compustat, we used the CCM link file.

Figure 1: Number of Mergers and Acquisition with Fama-French 12 industry categorization for the acquirer firm



interesting to note that 42% of the target firms were private firms, in 27% of the cases, the method of payment was all-cash, and in 75% of the cases, the acquirer and target firms were in the same Fama-French 12 industry. The last two rows of Table 2 show that the median acquirer exists before and after the event. The median firm has at least 12 months before the event (note that it is censored at 12 months), and the average survival rate in our sample is more than 33 months after the event.

2.2 Firms characteristics and other variables

Bessembinder et al. (2018) proposes a characteristic-based benchmark return approach, where the authors specifically incorporate firm characteristics to control for pre-event characteristics and to analyze the post-event market performance of the event firms. Following their work, we define two sets of firm characteristics: 'C5' contains the firm size, book-to-market ratio, momentum, ROA, and asset growth, whereas the second, broader set of characteristics, 'C14', includes C5 variables and market beta, accrual, dividend, cumulative 2-year returns, idiosyncratic risk, illiquidity, turnover, leverage, and sales over price. Table A1 shows the definitions of these variables. We use these variables to create the synthetic firms that is similar in these characteristics to the acquirer firm. The summary statistics for the variables are given in Table 2.

Table 1: Descriptive statistics for Mergers and Acquisitions related variables

	N	Mean	Median	SD	Min	P25	P75	Max
Shares owned before (%)	7312	0.77	0.00	4.80	0.00	0.00	0.00	49.00
Shares owned after (%)	7312	98.52	100.00	7.23	50.00	100.00	100.00	100.00
Shares acquired during transaction (%)	7312	97.75	100.00	8.62	50.00	100.00	100.00	100.00
Transaction Value (Billion $)	7312	0.72	0.07	3.31	0.01	0.02	0.30	89.56
Equity Value (Billion $)	7312	0.77	0.07	7.74	0.00	0.02	0.28	602.63
Target is private firm	6650	0.43	0.00	0.50	0.00	0.00	1.00	1.00
Target is public firm	6650	0.42	0.00	0.49	0.00	0.00	1.00	1.00
Method of payment: All-Stock	6650	0.30	0.00	0.46	0.00	0.00	1.00	1.00
Method of payment: All-Cash	6650	0.27	0.00	0.44	0.00	0.00	1.00	1.00
Method of payment: Mixed	6650	0.26	0.00	0.44	0.00	0.00	1.00	1.00
Number of bid	6650	1.03	1.00	0.20	1.00	1.00	1.00	5.00
Same industry (FF12)	7312	75.15	100.00	43.22	0.00	100.00	100.00	100.00
Same industry (FF48)	7312	63.80	100.00	48.06	0.00	0.00	100.00	100.00
Number of pre-event periods (months)*	7312	7.77	12.00	5.19	0.00	1.00	12.00	12.00
Number of post-event periods (months)	7312	33.51	36.00	8.99	0.00	36.00	36.00	471.00

Data is from the SDC Platinum database. Payment method is calculated as in Eckbo et al. (2018), all-stock if "consideration structure = shares" in SDC and all-cash if it equals "casho". FF12 and FF48 stand for Fama-French 12 and 48 industry categorizations. M&As are between 1980-2019 that are merged with CRSP and Compustat.

** The number of pre-events is censored at 12 months; there can be more pre-events that are missing from these statistics. We detect only 4 cases where the number of pre-events is 0s.*

Table 2: Summary statistics of firm characteristics

	N	Mean	Median	SD	Min	P25	P75	Max
Return (%)	3427698	0.0106	0.0000	0.1737	-1.0000	-0.0621	0.0667	3.0000
C5 characteristics								
Log Size	3396880	5.0716	4.9406	2.1617	0.8771	3.4718	6.5301	10.5793
Log BM	2770054	0.5169	0.3533	0.5913	0.0025	0.1418	0.6673	3.5802
Momentum	3427698	-0.0109	0.0219	0.4952	-1.8197	-0.1966	0.2501	1.2408
ROA	2722705	-0.0097	0.0071	0.0911	-0.5676	-0.0057	0.0213	0.1700
Asset growth	2524757	0.1370	0.0747	0.3918	-0.9045	-0.0292	0.2252	1.8930
Additional 9 characteristics for the C14 model								
Beta	3270014	0.9840	0.9300	0.7390	-0.7900	0.4750	1.4070	3.3316
Accrual	2884914	-0.0004	0.0000	0.0861	-0.3807	-0.0189	0.0157	0.3146
Dividend	3427698	0.0014	0.0000	0.0031	0.0000	0.0000	0.0012	0.0196
Log LR Return	3427698	0.0236	0.0263	0.6697	-2.4681	-0.1823	0.3847	1.6188
Idio. risk	3209670	0.0254	0.0223	0.0141	0.0037	0.0147	0.0335	0.0685
Illiquidity	2566083	0.5839	0.0327	1.4093	0.0000	0.0035	0.3431	8.3090
Trunover	3005903	0.1174	0.0649	0.1528	0.0028	0.0293	0.1423	0.9630
Leverage	3396880	0.8991	0.0731	3.2286	0.0000	0.0000	0.4672	26.3978
Sales/Price	2852766	3.2988	0.6598	11.2353	0.0000	0.1955	1.9329	94.1383

We use quarterly Compustat data and monthly CRSP variables. We have winsorized returns at -100% and 300%, replacing 0.0001% of the observations. We have winsorized the rest of the variables at 1% and 99%.

We use Fama-French 5 factors (Fama and French, 2015) from Kenneth's R. French's website[12], and the characteristic-based benchmark returns: CBBR5 and CBBR14 proposed by Bessembinder et al. (2018) from the contributed data at WRDS to compare our results to factor based models.

3 Long-run counterfactuals for event firms

To find or create valid control firms for the acquirer, we use the causal inference literature to impute the outcome of the event firm for the case where the event would not have occurred. In the following, we propose a unified approach that tackles the imputation task from an event design perspective and nests the commonly used matching in finance, and economics.

We use the event-study design approach and notation instead of the classical panel ("it") approach for multiple reasons. First, with firms over a long period of time (between 1980-2022), we have highly unbalanced panel data: firms appear and disappear in the dataset. Second, M&A events can be substantially different from each other; therefore, the treatment effect is heterogeneous across firms, and the event-study design approach allows us to mimic this difference better. Although we consider each event case by case, we will embed our method in the classical it notation.

Let $i = 1, \ldots, N$ denote the firms (units), while $t = 1, \ldots, T$ the time periods. In corporate event studies, in most cases, the event (treatment) is not an absorbing state; thus, after a fixed period of time, the event firms' treatment status (can) change. Let us denote t_i^* the time period when an event happens for firm i and the number of periods after event status changes back by $E_{reverse}$[13]. Let us define D_{it} as a vector taking the value of 1 between the time periods t_i^* and $t_i^* + E_{reverse}$ if an event occurs for the firm i.[14] Finally, let us define the event time: $e = -E_{pre}, \ldots, -1, 0, 1, \ldots, E_{post}$, where $E_{pre} > 0$, $E_{post} \geq 0$ and both are integers, defining the lower and upper bounds of the event window.[15] Now, the set of event firms contains all firms that have been treated $\mathcal{T} = \{i \,|\, \exists D_{it} = 1\}$ with $N^{tr} = |\mathcal{T}|$ elements. One of the main tasks is to select the set of control units for the event firm $i \in \mathcal{T}$. We define

[12] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[13] For simplicity, we use the same $E_{reverse}$ for all firms. Note that Callaway and Sant'Anna (2021) notes t_i^* by g.

[14] For simplicity, we restrict our attention to the case when each firm are treated only once. However, our method extends to cases where multiple events occur for the same firm.

[15] Note that if $E_{pre} = E_{post} = \infty$, we will get back the classical panel "it" setup. Furthermore, in many cases E_{pre} and E_{post} can vary between event firms since the event firm can appear or disappear after or before the date values given by E_{pre}, E_{post}.

the most generic set of control firms for the event firm i as

$$
\begin{aligned}
\mathcal{C}_{i|i\in\mathcal{T}}(e_{pre}, e_{post}) &= \mathcal{C}_i(e_{pre}, e_{post}) \\
&= \left\{ i' : i' \neq i,\ D_{i't} = 0\,, \quad \forall t \in [t_i^* - e_{pre}, \ldots, t_i^* + e_{post}] \right\},
\end{aligned} \tag{1}
$$

where we require that all potential control firms are not treated between the event time minus a pre-set period of time ($t_i^* - e_{pre}$) and after the event time up to e_{post} depending on the method we use. Note that e_{pre} and e_{post} can be the same as E_{pre} and E_{post}, or different. In the following, we show different methods that further restrict the set of candidate firms to improve the quality of controls.

Example: Bessembinder and Zhang (2013) uses firms as controls that have not gone through M&As at least 3 years before the event occurs, setting e_{pre} as 3 years. Allowing control firms to exit implies $e_{post} = 0$, so some controls may not have values for D_{i',t_i^*}, but requires at least one firm to be in the sample up to $t_i^* + E_{post}$.

In many cases (see, e.g., Loughran and Vijh, 1997, Eckbo et al., 2007, Bessembinder and Zhang, 2013, Bessembinder et al., 2018), the set of control firms is further filtered by particular criterion such as the market capitalization is between 70%-130% before the event occurs. One can express these types of conditions as

$$
\mathcal{C}_i^F(e_{pre}, e_{post}) = C_i(e_{pre}, e_{post}) \cap \left(i' : b_L \leq \Lambda^F(X_{i,t_i^*-e}, X_{i',t_i^*-e}) \leq b_H\right) \tag{2}
$$

where $\mathcal{C}_i^F(e_{pre}, e_{post})$ represents a filtered set of controls for event firm i, X_{it} are covariates (with $NT \times K$ dimension) that we want to use for filtering. We use event periods such as X_{i,t_i^*-e} stands for periods between $t_i^* - e_{pre}$ and $t_i^* + e_{post}$, thus the dimension of X_{i,t_i^*-e} is $(N(e_{post}+e_{pre}+1)\times K)$. b_L and b_H are the lower and upper bounds with the same dimensions as X_{i,t_i^*-e}. Λ^F is a function defined for the filtering.

Example: Imposing the market capitalization to be between 70% and 130% of the event firm, we use X_{it} as the market capitalization, set $e = 1$, $b_L = -0.3, b_H = 0.3$ and use $\Lambda^F(x_{i'}, x_i) = \frac{x_{i'} - x_i}{x_i}$.

3.1 Average long-run effect

This paper aims to give a credible estimate of the average long-run effect of the performance of the acquirer firms. It is a natural choice to focus on the average treatment effects of treated units on a specific horizon E_{post},

$$
\tau(E_{post}) = \mathbb{E}\left[Y_{ie}(1) - Y_{ie}(0) | D_{ie} = 1\right]\,, \qquad e \in [0, E_{post}]\,. \tag{3}
$$

To identify $\tau(E_{post})$, different methods use different identification assumptions, which we will cover along with the methodologies. If an identification can be established for $\tau(E_{post})$, we will express the sample counterparts in the following generic formula,

$$\begin{aligned} \hat{\tau}(E_{post}) &= \frac{1}{N^{tr} E_{post}} \sum_{i \in \mathcal{T}} \sum_{e=0}^{E_{post}} (Y_{ie} - \tilde{Y}_{ie}) \\ \tilde{Y}_{ie} &= \sum_{i' \in \mathcal{C}_i} \hat{\omega}_{i'} Y_{i'e} - \hat{\delta}_{i,pre} \end{aligned} \tag{4}$$

where $\tilde{Y}_{ie}$ is the imputed value for $Y_{ie}(0)\,,\ i \in \mathcal{T},\ e \in [0, E_{post}]$. We will consider methods that express $\tilde{Y}_{ie}$ as the weighted sum of control firms' value, where $\hat{\omega}_i$ are the method-dependent weights. Furthermore, we consider difference-in-differences type of estimators; thus $\hat{\delta}_{i,pre}$ stands for the pre-event differences in the event firm and pool of control firms.

In the following, we overview three methods, one commonly used in the corporate event literature and two methods from the causal inference literature, that allow our highly imbalanced panel to estimate heterogeneous treatment effects.

3.2 Classical Matching

The most commonly used matching method for corporate events is to find a firm that has similar market capitalization and the closest but larger book-to-market ratio one period before the event (see, e.g., Loughran and Ritter, 1995, Eckbo et al., 2007, Bessembinder and Zhang, 2013, Bessembinder et al., 2018, or Liu et al., 2023.[16]) We call this method "classical matching" as it is used as a benchmark in evaluating the long-run effects of corporate events. This method chooses only one firm as a control.[17] To formalize "classical matching" method, let us first review the identification assumptions needed for $\tau(E_{post})$.

Assumption 1 (Sequential ignorability, Robins et al. 2000)

$$\begin{aligned} &\{Y_{i,t_i^*+e}\left(D_{i,t_i^*}=1, D_{i,t_i^*-1}=0\right), \\ &\ Y_{i,t_i^*+e}\left(D_{i,t_i^*}=0, D_{i,t_i^*-1}=0\right)\} \perp\!\!\!\perp D_{i,t_i^*} | D_{i,t_i^*-1}, X_{i,t_i^*-1}\,, \qquad \forall e \in [0, E_{post}]\,. \end{aligned}$$

Assumption 1 means that there are no unobserved confounders that impact the (long-run) outcome. In fact, Bessembinder and Zhang (2013) shows that this assumption is violated

[16]Note, that the implemented method in the cited paper differs to some degree. For example, Loughran and Ritter (1995) does not match on book-to-market and matches on the following December 31. Eckbo et al. (2007), Bessembinder and Zhang (2013) matches on December 31 values with both variables.

[17]The classical method furthermore allows the chosen firm to be delisted. If this is the case, the method imputes after delisting the second-best candidate and iterates further if needed. To simplify formalism, we stick to one control firm, but one can generalize to this case.

when investigating multiple (different) corporate events. Another implication of this assumption is that the treatment, outcome, and covariate histories before one period of the event confound the causal relationship between the event and the outcomes realized after the events. This assumption may be too restrictive in our case, as it is reasonable to think that multiple periods before the actual M&A announcement, the management has already started to form its strategy.
The second assumption is the lack of spillover effects,

Assumption 2 (No spillover effects)

$$Y_{i,t_i^*+e}(D_{i,t_i^*+e}) \perp\!\!\!\perp D_{i',t}, D_{i,t_i^*-1}\,, \forall t, i \neq i', 0 \leq e \leq E_{post}\ ,$$

thus, the potential outcome of unit i after the event up to E_{post} is independent of other units' treatment status and history, up to one period before the event happens. Under Assumptions 1. and 2. $\tau(E_{post})$ given by Equation (3) can be identified by the "classical matching" method.[18]

As a second step, let us define the set of control firms as

$$\mathcal{C}_i^{CM}(-1,0) = \underset{i' \in \mathcal{C}_i^F(-1,0)}{\arg\min}\ |X_{i',t_i^*-1} - X_{i,t_i^*-1}| \quad s.t. \quad X_{i',t_i^*-1} \geq X_{i,t_i^*-1}\,,$$

where $\mathcal{C}_i^F(-1,0)$ refers to filtered units having market capitalization between 70% and 130% of the event firm. Here X_{i',t_i^*-1} stands for the book-to-market ratio one period before the event occurs only. The weight vector, in this case, simplifies to

$$\omega_{i'}^{CM} = \mathbf{1}(i' \in \mathcal{C}_i^{CM}(-1,0)) = \begin{cases} 1, & \text{if } i' \in \mathcal{C}_i^{CM}(-1,0) \\ 0, & \text{otherwise}\,. \end{cases}$$

This method relies on sequential ignorability and does not use pre-event differences; thus, $\hat{\delta}_{i,pre} = 0$. Equation (4) boils down to

$$\hat{\tau}^{CM}(E_{post}) = \frac{1}{N^{tr} E_{post}} \sum_{i \in \mathcal{T}} \sum_{e=0}^{E_{post}} (Y_{ie} - Y_{\mathcal{C}_i^{CM},e})\,. \tag{5}$$

[18] The proof of identification is similar to the proof of identification under unconfoundedness assumption see, e.g., Imbens and Wooldridge (2009).

3.3 TSCS Matching

Our second method is used more frequently in the economic literature. We follow Imai et al. (2019), who proposes matching methods for time-series cross-sectional (TSCS) data. Compared to the "classical matching" method, the main advantage is that it allows multiple candidate firms to be used and weighted instead of picking one firm. The use of multiple firms allows the control firm's trajectories to be averaged out and, thus, can be more robust to the choice of an individual control firm. It also allows for matching on multiple time periods prior to the event, not only one period, while accounting for level differences in the outcome during the pre-event periods. After defining the number of pre-periods before the event, one can match multiple firm characteristics and weigh those firms that are the closest to the event firm defined by these characteristics. Furthermore, one can relax the sequential ignorability assumption and use the parallel trend assumption instead to identify the parameter of interest that is more reasonable in our case.

To identify $\tau(E_{post})$ Imai et al. (2019) uses a form of parallel trend assumption,

Assumption 3 (Parallel trend assumption - TSCS)

$$\begin{aligned}
&\mathbb{E}\Big[Y_{i,t_i^*+E_{post}}\left(D_{i,t_i^*}=0, D_{i,t_i^*-1}=0, \left\{D_{i,t_i^*-e}\right\}_{e=2}^{E_{pre}}\right) \\
&\quad - Y_{i,t_i^*-1}|D_{i,t_i^*}=1, D_{i,t_i^*-1}=0, \left\{Y_{t_i^*-e}, D_{t_i^*-e}\right\}_{e=2}^{E_{pre}}, \left\{X_{i,t_i^*-e}\right\}_{e=1}^{E_{pre}}\Big] \\
=&\mathbb{E}\Big[Y_{i,t_i^*+E_{post}}\left(D_{i,t_i^*}=0, D_{i,t_i^*-1}=0, \left\{D_{i,t_i^*-e}\right\}_{e=2}^{E_{pre}}\right) \\
&\quad - Y_{i,t_i^*-1}|D_{i,t_i^*}=0, D_{i,t_i^*-1}=0, \left\{Y_{t_i^*-e}, D_{t_i^*-e}\right\}_{e=2}^{E_{pre}}, \left\{X_{i,t_i^*-e}\right\}_{e=1}^{E_{pre}}\Big]
\end{aligned}$$

along with Assumption 2.

We can apply the methodology of Imai et al. (2019) by selecting control units up to a pre-defined number J:

$$\mathcal{C}_i^{TSCS}(E_{pre},0) = \left\{i' : i' \in \mathcal{C}_i(E_{pre},0),\ S_i(i') \leq S_i^{(J)}\right\}. \tag{6}$$

where $S_i(i')$ is a selected distance measure between i and i' and $S_i^{(J)}$ is the J-th-order statistics of $S_i(i')$ among the control units in the original control set $\mathcal{C}_i(E_{pre},0)$. Imai et al. (2019) proposes the Mahalanobis distance measures (or other distance measures as in Rubin, 2006 or Stuart, 2010) or to use the estimated propensity score instead of X_{it} in $S_i(i')$. To show how matching variables $\{X_{i,t_i^*-e}\}_{e=1}^{E_{pre}}$ play a role, let us use the Mahalanobis distance measure,

$$S_i(i') = \frac{1}{E_{pre}} \sum_{e=1}^{E_{pre}} \sqrt{(X_{i,t_i^*-e} - X_{i',t_i^*-e})' \Sigma^{-1}_{i',t_i^*-e} (X_{i,t_i^*-e} - X_{i',t_i^*-e})}\,,$$

where $\Sigma_{i't}$ is the sample covariance matrix of $X_{i't}$.

One can have multiple options to define the weights for each control unit based on TSCS matching. The simplest is to provide equal weights,[19]

$$\hat{\omega}_{i'}^{TSCS} = \begin{cases} \frac{1}{|\mathcal{C}_i^{TSCS}|}\,, & \text{if}\, i' \in \mathcal{C}_i^{TSCS} \\ 0 & \text{otherwise}\,. \end{cases} \tag{7}$$

Alternatively, one can use the propensity scores to create weights based on units in $\mathcal{C}_i^{TSCS}$ or other calibration weights based on $S_i(i')$, see more in details Imai et al. (2019). As a last step, the difference-in-differences estimator of Imai et al. (2019) can be written as

$$\begin{aligned} \hat{\tau}^{TSCS}(E_{post}) &= \frac{1}{N^{tr} E_{post}} \sum_{i \in \mathcal{T}} \sum_{e=0}^{E_{post}} (Y_{ie} - \tilde{Y}_{ie}^{TSCS}) \\ \tilde{Y}_{ie}^{TSCS} &= \sum_{i' \in \mathcal{C}_i^{TSCS}} \hat{\omega}_{i'}^{TSCS} Y_{i'e} - \hat{\delta}_{i,pre}^{TSCS} \\ \hat{\delta}_{i,pre}^{TSCS} &= \frac{1}{E_{pre}} \sum_{e=1}^{E_{pre}} \left(Y_{i,t_i^*-e} - \sum_{i' \in \mathcal{C}_i^{TSCS}} \hat{\omega}_{i'}^{TSCS} Y_{i'e} \right) \end{aligned} \tag{8}$$

Although this method is more robust than the classical matching method, there are two potential threats in our case. First, it assumes that the weights on the pre-event periods are the same.[20] In our case, it is problematic, as we would like to weigh firm characteristics closer to the event and put less weight on periods further away. Put differently; equal weighting can result in fair matching on the *average* past characteristics but poor matching on more recent characteristics. Second, this method matches the levels and thus favors candidate firms that have a more volatile variable but around the event firm in contrast to a firm with the same pre-trend but with a slight level shift.

[19] We denote $\mathcal{C}_i^{TSCS} = \mathcal{C}_i^{TSCS}(E_{pre}, 0)$

[20] Imai et al. (2019) notes that the more pre-periods are taken into account, the more credible the analysis in terms of similarity; however, the fewer candidate matching firms and the quality of the match tend to decrease.

3.4 Stacked Synthetic Control

We introduce our version of the staggered synthetic difference-in-differences method that better describes corporate event requirements. We call it "stacked synthetic control" (SSC) to emphasize that we use the synthetic control method to create weights for each event firm, repeat this process for each M&A case, and stack the results together. Our basis for finding weights is based on the synthetic difference-in-differences method proposed by Arkhangelsky et al. (2021) and adopted for staggered treatment timing by Porreca (2022) and Clarke et al. (2023).[21] We extend their methodology by using multiple covariates to find firm and pre-event-specific weights. The main advantage of this method compared to the previous ones is that i) it allows multiple firm characteristics to match; ii) does not require specifying the exact number of pre-event periods but estimates the pre-event period weights data-driven; iii) it (sparsely) weights candidate firms such that it matches the variation in the event firm allowing for individual and time fixed effects differences.

To get the stacked synthetic control estimator for $\tau(E_{post})$, we follow Arkhangelsky et al. (2021) and Ben-Michael et al. (2022) defining the assumptions.

Assumption 4 (No anticipation) $Y_{ie}(1) = Y_{ie}(0),\ \forall i \in \mathcal{T}, e < 0$

Assumption 5 (Latent factor model) *There are* $\mathbf{L}$ *latent time-varying factors, where* $\mathbf{L} << NT$. *L is bounded. Each unit has a vector of time-invariant factor loadings* $\phi_i \in \mathbb{R}^{\mathbf{L}}$ *and time-varying component* $f_t \in \mathbb{R}^{\mathbf{L}}$ *that is bounded.*[22] *The outcomes can be written by the latent factor model as* $Y_{i,t_i^*-e} = \phi_i f_t + D_{i,t_i^*-e}\tau + \epsilon_{it}$, *where* ϵ_{it} *are mean zero, independent across units and time, and* $\epsilon_{it} \perp\!\!\!\perp t_i^*,\ \forall i, t$.

Assumption 6 (sub-Gaussian noise) ϵ_{it} *are sub-Gaussian random variables with scale parameter* σ *bounded away from zero.*

Assumption 7 (Properties of weights and L) *The control unit specific (oracle) weights* $\hat{\omega}_{i'}^{SSC}$, *the pre-event period specific (oracle) weights, and the latent time-varying factors satisfy the requirements listed in Arkhangelsky et al. (2021) Assumption 4. for each event.*

Under these assumptions and some asymptotic requirements on the sample size (see Arkhangelsky et al., 2021, Assumption 2), the sample analog of SSC will converge to $\tau(E_{post})$. Note that these assumptions are weaker than TSCS-matching or comparable two-way fixed-effects methods that require some form of parallel trend assumption.

[21] Note, they are using one outcome variable and not estimating the weights on multiple characteristics. In contrast, Abadie and L'hour (2021), Ben-Michael et al. (2022), and Cattaneo et al. (2023) using covariates but not weights for pre-event periods.

[22] Arkhangelsky et al. (2021) in Assumption 3. works out the exact bounds

As a second step, we construct the estimator for SSC. Candidate control units for SSC are the same as the generic control group (or the filtered) that requires treatment history between E_{pre} and E_{post} $\mathcal{C}_i^{SSC}(E_{pre}, E_{post}) = \mathcal{C}_i(E_{pre}, E_{post})$. For simplicity let us use $\mathcal{C}_i^{SSC} = \mathcal{C}_i^{SSC}(E_{pre}, E_{post})$. An important part of SSC is that it estimates the unit-specific weights for each control group unit and the pre-event period weights. In contrast to the synthetic difference-in-differences approach, we use a set of covariates to get the weights (similar to other synthetic control methods such as Abadie et al., 2010, Abadie et al., 2015, Doudchenko and Imbens, 2016 or Abadie and L'hour, 2021),

$$\begin{aligned}\hat{\omega}_{i'}^{SSC} = \left(\hat{\omega}_0^K, \hat{\omega}_{i'}\right) = \underset{\omega_0 \in \mathbb{R}, \omega \in \Omega}{\arg\min} \sum_{e=1}^{E_{pre}} \Bigg[& \Big(\omega_0 + \sum_{i' \in \mathcal{C}_i^{SSC}} \omega_{i'} X_{i', t_i^* - e} - X_{i, t_i^* - e}\Big)' \hat{\Sigma}_{t_i^* - e}^{-1} \\ & \Big(\omega_0 + \sum_{i' \in \mathcal{C}_i^{SSC}} \omega_{i'} X_{i', t_i^* - e} - X_{i, t_i^* - e}\Big) \Bigg] + \zeta^2 E_{pre} \sum_{i' \in \mathcal{C}_i^{SSC}} \omega_{i'}^2 \\ \Omega = \Big\{ \omega \in \mathbb{R}_+^N : \sum_{i' \in \mathcal{C}_i^{SSC}} \omega_i = 1 \Big\}. & \end{aligned} \tag{9}$$

We want to find firm-specific weights that minimize this distance measure across units before the event happens. There are some important remarks; similarly to the synthetic control literature, we require the weights to sum up to one and to be zeros or positive. Furthermore, there is a discussion in the literature on how to scale the covariates (see, e.g., Abadie, 2021). With our staggered setup, case-by-case cross-validation methods are not particularly appealing[23]; we restrict our attention to scaling the variables by their variance-covariance matrix ($\hat{\Sigma}_{t_i^* - e}^{-1}$) calculated across units for each $t_i^* - e$ period to simplify the estimation. Second, similarly to Doudchenko and Imbens (2016) and Arkhangelsky et al. (2021), we use the regularization parameter ζ, which helps to increase the dispersion and ensure the uniqueness of the weights. In the spirit of Arkhangelsky et al. (2021), we use the scaled trace of the variance-covariance matrix of the differentiated covariates across time and control units.[24] Finally, let us highlight that with the use of ω_0, we allow for individual fixed effects in the covariates X_{it}, which helps to match the time-varying trend rather than on the levels.

As a second step, we weigh pre-event periods. We emphasize that firms may have different characteristics before the event but account for the possibility of becoming more similar closer

[23]Although with staggered design, one may use cross-validation method to select variables across different events.

[24]Let $\zeta = E_{post}^{1/4} K^{-1} tr(\tilde{\sigma})$, with $\tilde{\sigma}^2 = \frac{1}{|\mathcal{C}_i^{SSC}| E_{pre} - 1} \sum_{i' \in \mathcal{C}_i^{SSC}} \sum_{e=1}^{E_{pre}-1} \left(\Delta X_{i', t_i^* - e} - \overline{\Delta X}_{i', t_i^* - e}\right)$, where K is the number of covariates used for finding the weights, $tr()$ stands for the trace of the matrix, $\Delta X_{i', t_i^* - e} = X_{i', t_i^* - e + 1} - X_{i', t_i^* - e}$ and is the averaged value across controls and pre-event periods. One may consider using other measures, such as Fröbenius or nuclear norm, instead of trace.

to the event and weighing them accordingly. After we have created the synthetic event firm, we search for optimal pre-event period weights,

$$
\begin{aligned}
\left(\hat{\lambda}_0, \hat{\lambda}\right) = \underset{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda}{\arg\min} \sum_{i' \in \mathcal{C}_i^{SSC}} \Bigg[& \left(\lambda_0 + \sum_{e=1}^{E_{pre}} \lambda_{t_i^*-e} X_{i', t_i^*-e} - \frac{1}{E_{post}} \sum_{e=0}^{E_{post}} X_{i', t_i^*-e}\right)' \tilde{\Sigma}_{i'}^{-1} \\
& \left(\lambda_0 + \sum_{e=1}^{E_{pre}} \lambda_{t_i^*-e} X_{i', t_i^*-e} - \frac{1}{E_{post}} \sum_{e=0}^{E_{post}} X_{i', t_i^*-e}\right) \Bigg] \\
\Lambda = \left\{ \lambda \in \mathbb{R}_+^T : \sum_{e=1}^{E_{pre}} \lambda_t = 1 \right\}. &
\end{aligned} \tag{10}
$$

The pre-event weights are similar as in Arkhangelsky et al. (2021); they optimize over the differences between pre-and post-event periods. The main difference is the scaling parameter $\tilde{\Sigma}_{i'}^{-1}$, that is, the variance-covariance matrix over time periods for each control unit.

The resulting stacked synthetic control estimator can be written as

$$
\begin{aligned}
\hat{\tau}^{SSC}(E_{post}) &= \frac{1}{N^{tr} E_{post}} \sum_{i \in \mathcal{T}} \sum_{e=0}^{E_{post}} (Y_{ie} - \tilde{Y}_{ie}^{SSC}) \\
\tilde{Y}_{ie}^{SSC} &= \sum_{i' \in \mathcal{C}_i^{SSC}} \hat{\omega}_{i'}^{SSC} Y_{i'e} - \hat{\delta}_{i,pre}^{SSC} \\
\hat{\delta}_{i,pre}^{SSC} &= \sum_{e=1}^{E_{pre}} \left(\hat{\lambda}_{t_i^*-e} Y_{i, t_i^*-e} - \sum_{i' \in \mathcal{C}_i^{SSC}} \hat{\omega}_{i'}^{SSC} \hat{\lambda}_{t_i^*-e} Y_{i'e} \right) - \hat{\lambda}_{0,i}\,.
\end{aligned} \tag{11}
$$

Remarks: First, the stacked synthetic control estimator is a weighted difference-in-differences estimator. It takes the differences in pre-event periods and after the event. Furthermore, it weights both units by $\hat{\omega}_{i'}^{SSC}$ and the pre-event periods by $\hat{\lambda}_{t_i^*-e}$. In the case $\hat{\omega}_{i'}^{SSC} = \frac{1}{|\mathcal{C}_i^{SSC}|}$ and $\hat{\lambda}_{t_i^*-e} = 1/E_{pre}$, we get a difference-in-differences estimator that is the same as the event study parameter proposed by Callaway and Sant'Anna (2021).[25] The second remark is that the stacked synthetic control can be viewed as an extension of the TSCS matching by Imai et al. (2019) in multiple dimensions. i) it uses pre-event period weights, ii) it solves for the weights for each unit $\hat{\omega}_{i'}$, instead of finding a matched set and setting the weights. iii) It can match not only on the levels and variation but can use fixed effects during the matching that favor time variation over level matching. As a next remark, the stacked synthetic control method augments the classical matching by using weights in the last pre-event period and

[25]For this type of difference-in-differences estimator to be identified, one needs the parallel trend assumption for post-treatment only, that was introduced by Callaway and Sant'Anna (2021).

using only one control firm. Finally, let us note that Ben-Michael et al. (2022) proposes a pooled and separate synthetic control method in the staggered treatment setup. Our case is the pure separate synthetic method that mimics our problem better as the heterogeneous treatment effect and changing the pool of control group makes it questionable to weigh the average of the acquirer firms.

3.5 Further remarks

One may consider other imputation estimators from the field of difference-in-differences (see e.g., Wooldridge, 2021, Borusyak et al., 2023 or Liu et al., 2022 or for on overview Roth et al., 2023, De Chaisemartin and d'Haultfoeuille, 2023 or Baker et al., 2022). For these types of estimators to work, one needs to use the assumption of no anticipation effects (Assumption 4) and be willing to assume that parallel trends hold for all event firms through all time periods, which is not too appealing in our case.

So far, we have not discussed standard errors of the $\hat{\tau}(E_{post})$ estimators. Following the literature, it is reasonable to consider the conditional and unconditional variances, where we condition the weights. For unconditional variance, Imai et al. (2019) proposes a bootstrap method that we employ with TSCS matching. We will follow Arkhangelsky et al. (2021) and Clarke et al. (2023) with the stacked synthetic control method and employ placebo variance estimation instead of the alternatives.[26]. Lastly, with classical matching, we can also use placebo variance estimation using non-event firms from the control group.

As a last theoretical point, let us consider the factor or benchmark-based approaches for evaluating the long-run performance of the firms. We follow Bessembinder et al. (2018), where they use the following specification,

$$Y_{it} - \mathbb{E}\left[Y_{it} | f(X_{it-1})\right] = \alpha + \tau^{FM} D_{it} + \epsilon_{it} \,, \tag{12}$$

where $\mathbb{E}\left[Y_{it} | f(X_{it-1})\right]$ stands for benchmark returns or factor-based returns, using covariates one period before, X_{it-1} and τ^{FM} stands for the long-run performance measure using factor models. To identify τ^{FM} as $\tau(E_{post})$, we can take two approaches. The common assumptions for both approaches are the sequential ignorability assumption (Assumption 1) with or without conditioning covariates and the assumption of no spillover effects (Assumption 2).

[26] Arkhangelsky et al. (2021) note that placebo is the proper method with one treated unit, although they offer a bootstrap and a jackknife procedure. However, placebo variance estimation requires homogeneity assumption across controls. Also, note that Cattaneo et al. (2023) proposes principled prediction intervals that offer precise non-asymptotic coverage probability guarantees.

The first approach considers $Y_{it} - \mathbb{E}\left[Y_{it}|f(X_{it-1})\right]$ as the outcome variable and refers to τ^{FM} as the (modified) parameter of interest that tells the long-term performance of the excess return compared to the factor/benchmark returns. Using different outcome variables implies that the long-run effect measure will refer to different quantities depending on the factor/benchmark model used; thus, comparing these results is not particularly helpful.

The second approach uses X_{it-1} as a conditioning variable. In this case, the outcome is the same everywhere; however, one must be careful with the "bad control problem". To avoid "bad controls", X_{it-1} must not be affected by treatment at any time (see: Zeldow and Hatfield, 2021, Caetano et al., 2022 or Cinelli et al., 2022). However, the commonly used variables as factors or characteristics of the firm are affected by the treatment. A simple example is the book-to-market ratio. At the time of the event for unit i, the merger affects the firm's market prices. Through price, it affects both: the returns r_{i,t_i^*} and the book-to-market ratio BM_{i,t_i^*}. In the next time period, we are interested in the changes of r_{i,t_i^*+1}; however, it is affected through BM_{i,t_i^*} as well, not only through D_{i,t_i^*}. Figure 2 shows this case.

$$
\begin{tikzcd}
\mathrm{BM}_{i,t_i^*-1} \arrow[r] & D_{i,t_i^*} \arrow[r] \arrow[d] \arrow[dr] & r_{i,t_i^*} \arrow[dl, dashed, red] \\
 & \color{red}{\mathrm{BM}_{i,t_i^*}} \arrow[r, red] & r_{i,t_i^*+1}
\end{tikzcd}
$$

Figure 2: Book-to-Market ratio (BM) is a bad control, while investigating the effect of M&As in the market returns (r_{it})

4 Results

4.1 Empirical strategy

Whenever possible, we compare three sets of covariates to calculate the weights. We use the following abbreviations for each set of covariates[27]:

1. C2: only log of size and log of book-to-market ratio.

2. C5: based on C5 variables of Bessembinder et al. (2018): log of size, log of book-to-market ratio, momentum, ROA and asset growth (non-lagged values)[28]

[27]Definition of covariates is defined in Table A1.

[28]Bessembinder et al. (2018) proposes to lag C5 or C14 variables, when creating CBBR5 or CBBR14.

3. C14: C14 variables of Bessembinder et al. (2018). In addition to C5 variables, market beta, accrual, dividend, log of long-run return, idiosyncratic risk, illiquidity, turnover, leverage, and sales over price (non-lagged values).

Classical matching

For the classical matching method, we follow Bessembinder et al. (2018) and require that the matching firm has a log of market capitalization between 70% and 130% of the acquirer firm and its log of book-to-market ratio is larger than the closest to the event firm. Thus, we will use only C2 variables, not C5 or C14, with classical matching. To be aligned with the literature, we allow for delisting; thus, if the control firm gets delisted, we will use the second, third firm, and so on after the date of delisting.

TSCS-matching

For TSCS matching, we use 12 months of data before the event and match all three sets of firm characteristics: i) C2, ii) C5, and iii) C14 variables. Before matching, we use a filter similar to the "classical method": restrict our attention to firms that have log market capitalization and log of book-to-market ratio between 50% and 150% of the event firm before one month of the announcement and C14 variables observed before the event for 12 months. We allow for a broader range as we do not want to restrict our matching sample too much in the case of C5 and C14 variables. We also require that each candidate firm has at least the same data record as the acquirer firm, so we do not allow delisting. We use the Mahalanobis distance measure to evaluate the distance from the event firm and choose a maximum of 50 firms closest to the event firm.[29] We used equal weights for the selected pool of control firms. We will call the results of this method "TSCS-" and add the variable set abbreviations accordingly.

Stacked Synthetic Control

Our empirical strategy for the stacked synthetic control method is similar to TSCS: first, we do the same filtering that allows a maximum of 100 candidate matches[30]. In the second step, we use our SSC method for each event, one by one, to calculate the weights. This

Lagging the variables is essential to create a benchmark return based on market information. However, we would like to create a good quality match before the event that allows us to use contemporaneous variables before the event.

[29]The number of matched firms can be smaller if the number of firms with the same treatment pattern before the event differs.

[30]If there are more than 100 candidates, then we prefer those that are in the same FF48 industry categorization and have a smaller euclidean distance in terms of log book-to-market ratio.

procedure generates a synthetic firm for each event firm. We require a record of variables at least two months before the event but allow up to 12 months. During solving for Equations (9) and (10), we standardize the covariates and set $\hat{\Sigma}^{-1}_{t_i^*-e}$ and $\tilde{\Sigma}^{-1}_{i'}$ to a diagonal matrix with $1/K$ that further simplifies our calculations. To consider a pool of control firms valid, we require that there are at least four candidate firms ($|\mathcal{C}_i^{SSC}| \geq 4$) and after we estimated the weights, we consider a synthetic firm valid if the weights are less than 0.5, which mimics the requirements of Assumption 7.

4.2 Weights of stacked synthetic control

To evaluate the stacked synthetic control method, we analyze the estimated weights, both for pre-event periods and for the control firms. These are important as weight irregularities may indicate a violation of the assumptions behind stacked synthetic control.

First, we analyze the control firm-related weights $\hat{\omega}_{i'}^{SSC}$. Table 3. reports the summary statistics for $\hat{\omega}_{i,e}$ across events. The first row, "Number of candidate firms", shows the distribution for the number of candidate firms in $\mathcal{C}_i^{SSC}$. On average, each event firm has 82 candidate firms. We report the number of non-null weighted firms after estimating the weights. The distribution is quite stable across the different matching variable sets, with 35-46 weighted firms on average. Not surprisingly, as the number of matched variables increases, the number of non-null weighted firms also increases to capture the increased complexity in the data. Note that we have truncated the maximum number of candidate firms in the control group to 100. The weight distribution has a mean of 2-3%, meaning most of the control firms have a relatively low weight. Furthermore, the distribution is skewed to the right and truncated at 100.[31]

[31] Figure A1 shows the distribution of non-null weighted firms.

Table 3: Descriptive statistics for the number of firms used during stacked synthetic control method

		N	Mean	Median	SD	Min	P25	P75	Max
No candidate firms		3634	82.15	100.00	30.77	4.00	73.00	100.00	100.00
All cases									
No. non-null weight firms	SM-C2	3634	35.80	32.00	23.17	1.00	17.00	51.75	100.00
	SM-C5	2968	43.09	26.00	37.39	1.00	12.00	99.25	100.00
	SM-C14	2966	46.09	26.00	38.43	1.00	14.00	100.00	100.00
Firm weights (non-null)	SM-C2	130103	0.03	0.02	0.03	0.00	0.01	0.03	1.00
	SM-C5	127884	0.02	0.01	0.03	0.00	0.01	0.02	1.00
	SM-C14	136711	0.02	0.01	0.03	0.00	0.01	0.02	1.00
Events with $\hat{\omega}_{i,e} > 0.5$ removed									
No. non-null weight firms	SM-C2	3610	36.02	32.00	23.09	4.00	17.00	52.00	100.00
	SM-C5	2924	43.70	26.00	37.33	4.00	12.00	100.00	100.00
	SM-C14	2947	46.37	26.00	38.40	4.00	14.00	100.00	100.00
Firm weights (non-null)	SM-SBM	130077	0.03	0.02	0.03	0.00	0.01	0.03	0.49
	SM-C5	127849	0.02	0.01	0.03	0.00	0.01	0.02	0.49
	SM-C14	136688	0.02	0.01	0.03	0.00	0.01	0.02	0.49

Descriptive statistics for the distribution of firm-specific weights for the SSC method ($\hat{\omega}_{i'}^{SSC}$). Each event firm can have a maximum of 100 candidate firms, but the number of actual candidate firms varies, as we require a log of size and book-to-market ratio within the 50% and 150% range. The number of candidate firms is the same regardless of the variables used. The number of non-null weight firms shows the distribution of the estimated non-zero weights over each event firm. The firms' weights are conditional on having non-null weights, thus we remove firms that are not used to create the synthetic firm for the event firm.

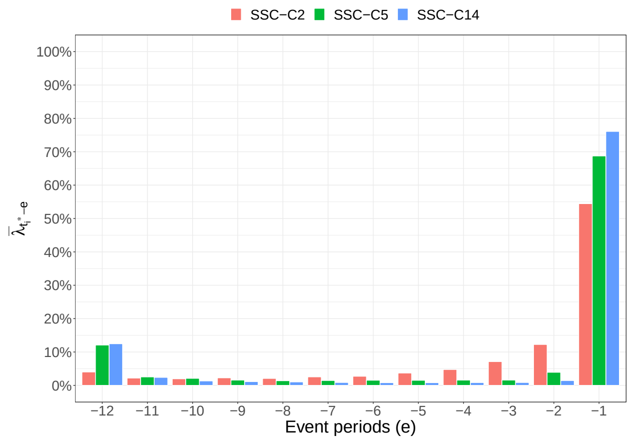
As the synthetic control approach requires that the event firm's characteristics be in the candidate firms' convex hull, we check the distribution of weights. Events are removed if any firm has a larger weight than 50%. Note that it is not a sufficient condition, but increases the quality of the overall match. Overall, 24-49-19 events dropped for C2-C5-C14 cases; the descriptive statistics regarding weights have stayed the same.

In Figure 3, we report the average of pre-event period weights. These are calculated by $\bar{\hat{\lambda}}_{t_i^*-e} = \frac{1}{N^{tr}} \sum_{i \in \mathcal{T}} \hat{\lambda}_{t_i^*-e}$, where $e = 0, \ldots, 12$. We can see that the trajectory of time weights is similar for all variable sets: they put around 60-80% of the weights on one period before the announcement, and then the weights exponentially decay.

4.3 Matching quality on firm characteristics variables

We provide visual and quantitative evidence to understand the quality of matching of the different methods. We report here the matching results for the Classical Matching, TSCS-

Figure 3: Average of time weights



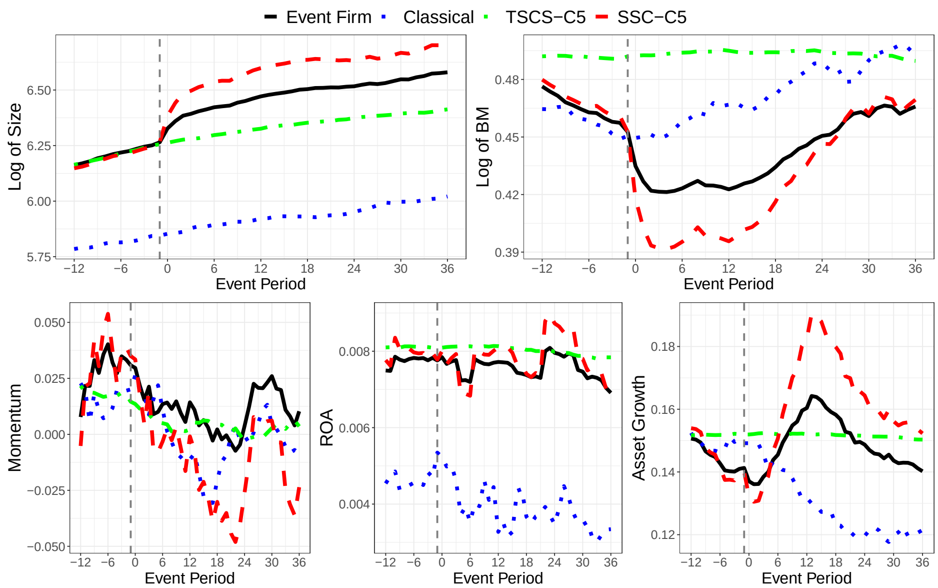
Each bar represents the average $\hat{\lambda}_t$ across events for each time period up to -12 (one year).

Matching with C5 variables, and the SSC method, also with C5 variables. In the appendix, we provide further figures showing the different variable sets for SSC; however, the results are similar to those using C5 variables. Figure 4. shows how different methods capture the pre-event variation in the averaged control compared to the averaged treated firms for C5 variables. For the TSCS and SSC methods, we adjusted for the differences before the event with $\hat{\delta}_{pre}$.

Figure 5. shows the remaining nine variables of C14. None of the methods have used these variables to match; however, the SSC method could capture the pre-event trends considerably well. The classical matching method performs relatively poorly, except on the log of book-to-market ratio, which is similar to the findings of Bessembinder and Zhang (2013).

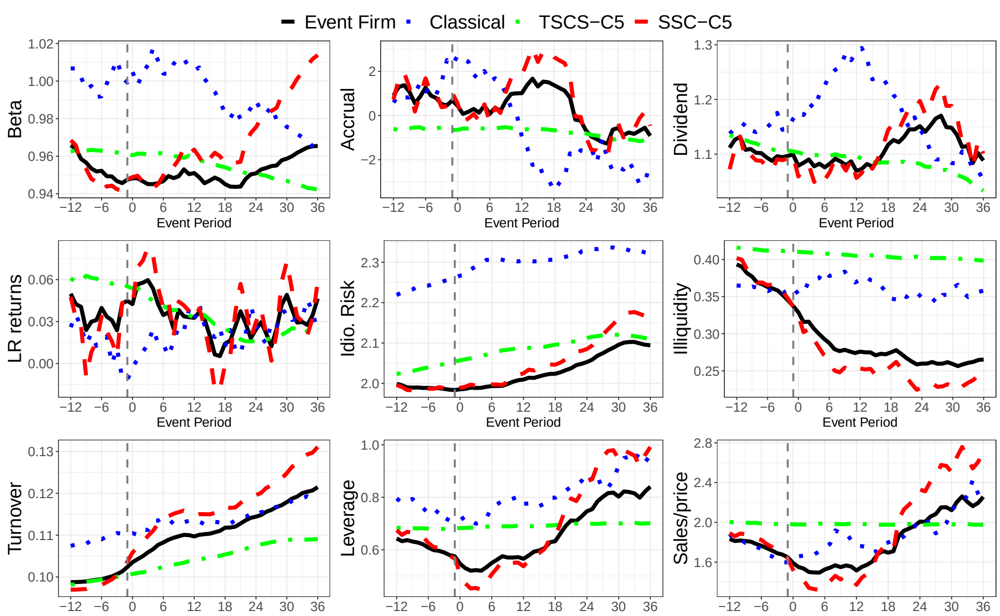
To quantify the match quality, we check the *differences* between the event firms and created synthetic firms. We carried out two tests reported in Table 4. The first test compares the weighted differences in the variables and is shown by columns 'Weighted'. This test shows that if the assumed number of pre-event periods is correct by assumption, what the average differences are between the control and treated firms. For "Classical", we take only one period before the event. For TSCS-C5, we take 12 months before the event and weigh equally. We use the time weights $\hat{\lambda}_{t_i^*-e}$ for stacked synthetic control. Results show that while Classical and TSCS-C5 methods reject the H_0, which means they are the same for almost all cases, with stacked synthetic methods, we can not reject the null in any case. The second block of columns shows the unweighted differences for 12 months before the event. The general pattern is similar: for Classical and TSCS-C5, we reject the H_0 at a lower significance level.

Figure 4: C5 variables for treated and created control firms by different methods methods



Averaged firm characteristics before 12 months and after the event for 36 months. Classical stands for "classical matching", and TSCS–C5 uses "TSCS-matching" with Mahalanobis distance measure and C5 variables. SSC-C5 stands for the "stacked synthetic control" method using C5 variables. TSCS and SSC methods are accordingly adjusted with the pre-event differences ($\hat{\delta}_{pre}$). The dashed vertical line is placed one period ahead of the event.

Figure 5: Remaining C14 variables for treated and different control firms by methods



Averaged firm characteristics before 12 months and after the event for 36 months. Classical stands for "classical matching," TSCS–C5 uses "TSCS-matching" with Mahalanobis distance measured with the C5 variables. SSC-C5 stands for the "stacked synthetic control" method using C5 variables. TSCS and SSC methods are accordingly adjusted with the pre-event differences ($\hat{\delta}_{pre}$). The dashed vertical line is placed one period ahead of the event.

Table 4: Differences in the means for pre-event periods

	Weighted					12 months				
Variable	Classical	TSCS-C5	SSC-C2	SSC-C5	SSC-C14	Classical	TSCS-C5	SSC-C2	SSC-C5	SSC-C14
C5 variables										
Log of BM	0.0000	0.103***	0.0000	0.0000	0.0000	0.004	0.103***	0.0006	0.0024	0.0017
Log of size	0.4649***	0.4164***	0.0000	0.0000	0.0000	0.4587***	0.4164***	-0.0019	-0.0095	-0.0106
Momentum	0.0269	-0.0175	0.0000	0.0000	0.0000	0.0265	-0.0175	-0.003	0.0055	0.0016
ROA	0.0031	-0.0019	0.0000	0.0000	0.0000	0.0037	-0.0019	0.0000	0.0002	0.0001
Asset growth	-0.0104	0.0526***	0.0000	0.0000	0.0000	-0.006	0.0526***	0.0012	0.0009	0.00002
Additional 9 variables for C14										
Beta	-0.057	0.2457***	0.0000	0.0000	0.0000	-0.0592	0.2457***	0.0011	-0.0002	-0.0004
Accrual	-0.0017	-0.0004	0.0000	0.0000	0.0000	-0.0005	-0.0004	0.0001	0.0000	0.0002
Dividend	0.0000	0.0003***	0.0000	0.0000	0.0000	0.0000	−0.0003***	0.0000	0.0000	0.0000
Log of long run return	0.0722	-0.0373	0.0000	0.0000	0.0000	0.0351	-0.0373	-0.0028	-0.0015	-0.0028
Idiosyncratic risk	−0.0032***	0.0048***	0.0000	0.0000	0.0000	−0.0029***	0.0048***	0.0001	0.0000	-0.0001
Illiquidity	−0.0911*	0.045	0.0000	0.0000	0.0000	-0.0667	0.045	-0.0051	-0.0037	-0.002
Turnover	-0.0091	0.0276***	0.0000	0.0000	0.0000	-0.0095	0.0276***	-0.0002	-0.0015	-0.0021
Leverage	-0.1509	0.1066	0.0000	0.0000	0.0000	-0.1493	0.1066	0.0071	0.0131	0.0097
Sales / price	0.0155	0.5565***	0.0000	0.0000	0.0000	0.014	0.5565***	0.0191	0.0294	0.0275
Returns	0.0044	0.0029*	0.0000	0.0000	0.0000	0.0013	0.0029*	0.0000	0.0000	0.0000

*Differences in the variables for M&A firms and (created) control firms pre-event periods. Different matching methods are reported along different time horizons. Weighted stands for the assumed/pre-set pre-event periods. For classical matching, it is one period before the event. TSCS-C5 is used 12 months prior to the event with equal weights. SSC methods use the estimated weights $\hat{\lambda}_{t_i^*-e}$. Stars are corresponding Bonferroni-adjusted p-values for pooling the same means over all events. *** stands for 0.1%, ** stands for 1% and * for 5% significance levels. Standard errors are clustered at firm and calendar time levels.*

Although the stacked synthetic control method does not rely on the parallel trend assumption, we have run the test proposed by Roth (2022) to check if parallel trend assumptions would hold with the (created) control group. The results show that the stacked synthetic control method outperformed both alternatives. See the results in the Appendix, Table A3.

4.4 Average performance of M&A events

To evaluate the long-run performance of the average M&A firm, we calculate the excess returns for each event firm using the matched / synthetic firms. We use raw monthly excess returns $\hat{r}_{i,e}^{excess} = r_{i,e} - \tilde{r}_{i',e}$, where $\tilde{r}_{i',e}$ is given by the method.[32]. Table 5. shows the main results on the long-run excess returns for acquirer firms up to 36 months. The classical matching method suggests an average decrease of 11 basis points per month, which is significant at the 5% level. Its associated cumulative abnormal returns on the three-year horizon is $\left(CAR(36) = \sum_{e=0}^{36} \hat{r}_{ie}^{excess}\right)$ is -379 basis points. Similar results are found by Bessembinder and Zhang (2013), Bessembinder et al. (2018) for classical matching. TSCS method with different sets of variables provides the largest average negative effects in the long run for both excess returns and CARs. However, this method had the worst pre-event matching quality, and thus, its results should be handled cautiously. Finally, the stacked synthetic control method provides the smallest excess returns along with the CARs, whereas the standard errors are the largest in this case. All this suggests that M&As are neither value adding nor value destroying *on average*. A final but important remark for Table 5 is that the methods use different event samples. The number of events is conditional on the characteristics of the available firm and the filtering process that we imposed in the empirical strategy. We report the different samples even if they refer to different cases.

[32] We use raw returns instead of log returns as Chen and Roth (2023) shows that $log(1 + y)$ type of transformation with zero-valued outcomes resulting ATE to be arbitrarily scale-dependent.

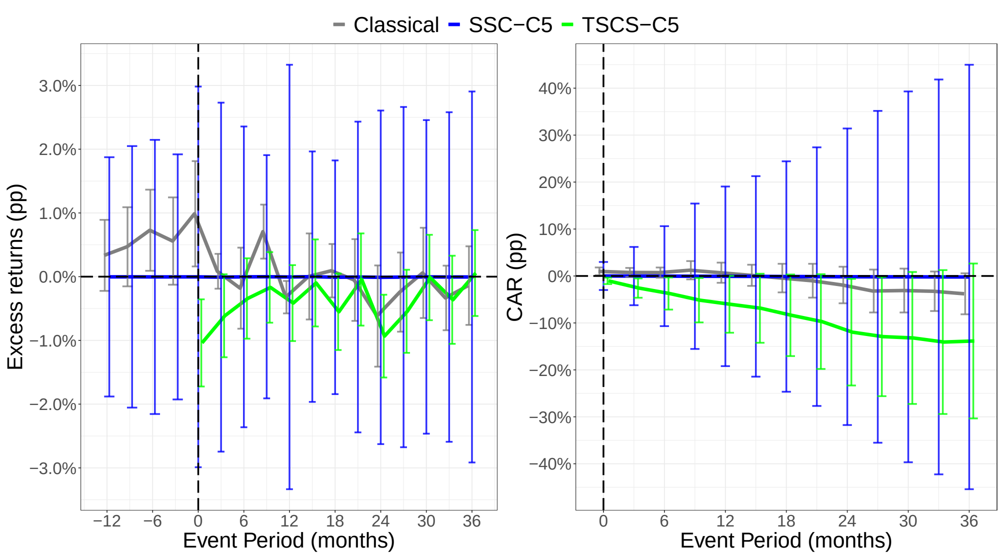
Table 5: Average excess returns (pp) after announcement up to 36 months

Method	$\hat{r}_{i,e}^{excess}$	$SE\left[\hat{r}_{i,e}^{excess}\right]$	$\widehat{CAR}(36)$	$SE\left[\widehat{CAR}(36)\right]$	Num. Events
Classical Matching	-0.1108^{*}	0.0487	-3.7940	2.2299	4,250
TSCS-C2	-0.3364	0.2176	-12.4452	8.0499	3,289
TSCS-C5	-0.3742	0.2275	-13.8443	8.4174	3,289
TSCS-C14	-0.5144^{*}	0.2164	-19.0337^{*}	8.0057	3,289
SSC-C2	-0.0023	1.3649	-0.0935	15.0766	2,445
SSC-C5	-0.0056	1.5474	-0.2151	23.0625	2,295
SSC-C14	-0.0047	1.5616	-0.1766	25.2008	2,267

**** stands for 0.1%, ** stands for 1% and * for 5% significance levels. Standard errors are clustered at the industry-year level for classical matching and are the conditional SEs on finding the proper counterfactual. TSCS and SSC methods report unconditional SEs.*

Figure 6. shows the same results with plotting the excess returns before and after the event. An interesting point from the left panel is that there is a systematic difference from zero in the excess returns for the classical matching method before the event, that is, the result of the poor matching pre-event periods.

Figure 6: Time-varying excess returns with different matching methods



Error bars are showing 95% CI based on unconditional standard errors.

Table 6: Simulation results

	Model	Mean	SD	Min	Median	Max
		No excess return				
$\hat{r}^{excess}_{i,e}$	Classical	-0.40	0.28	-0.64	-0.54	0.01
	TSCS-C5	0.54	1.53	-2.06	1.16	1.87
	SSC-C5	0.00	0.02	-0.03	0.01	0.02
$\widehat{CAR}(36)$	Classical	-14.79	8.92	-21.85	-17.62	0.64
	TSCS-C5	20.16	56.77	-76.21	42.82	69.19
	SSC-C5	0.04	0.70	-1.19	0.28	0.59
		With 25bp excess return per period – 9pp CAR(36)				
$\hat{r}^{excess}_{i,e}$	Classical	0.04	0.58	-1.68	0.05	1.84
	TSCS-C5	0.05	2.39	-9.41	0.01	7.50
	SSC-C5	0.23	0.60	-3.69	0.30	4.81
$\widehat{CAR}(36)$	Classical	1.83	21.56	-61.28	1.90	67.13
	TSCS-C5	1.91	88.31	-348.14	0.19	277.52
	SSC-C5	8.61	22.46	-136.16	11.13	177.40

Classical stands for classical matching method, TSCS-C5, and SSC-C5 are the time-series cross-sectional matching and stacked synthetic control method using C5 variables. Descriptive statistics refer to the distribution of point estimates. E.g., mean stands for the average value of the estimated $\hat{r}^{excess}_{i,e}$ *that is all excess returns between event periods 0-36 or* $\widehat{CAR}(36)$ *that refers to the average of CARs after 36 periods of the event.*

4.5 Simulation results

We run two sets of simulations to see how the different methods work in a controlled environment. In the first simulation, we randomly select 100 firms[33] and check if we find any effect. This simulation mitigates the scenario when there is no excess return (no M&A event happened). We only report Classical Matching and TSCS and SSC methods with C5 variables to simplify our analysis. We have used 500 repetitions, and Table 6 shows the summary statistics of the estimated effects for each method. One can infer from this exercise that all methods provide close averages to zero; thus there is no systematic bias in any of the used methods under the null of zero effect. All standard deviations are greater than the difference of the average from zero. However, the stacked synthetic control method outperforms both the classical and the TSCS methods regarding the magnitude of sample bias and standard deviation. This shows us that SSC gives more reliable results than the alternatives if there is no effect.

As a next exercise, we use the same setup as before, but now, at and after the event date, we add 25 basis points for each monthly return on top of the observed returns. This means nine percentage points CAR on the 36-month horizon, which mitigates our larger

[33]We require that the selected firms are present in the data for at least 48 months. This is required to ensure that we have at least 12 months before the event and 36 months after. If the selected firm is present for more than 48 months, we randomly select the event month between 12 months after and 36 months before the first and last date.

positive effects in the estimated sample with the SSC-C5 method. Table 6 shows that the SSC method finds 23 basis point excess returns and 8.61 pp CAR at the end of 36 months. On the contrary, classical and TSCS methods miss the positive effect and get 0.4-0.5 basis points per month and 1.8-1.9 pp as the CAR at the end of 36 months.

4.6 Other methods

We compare our matching results with the methods proposed in Bessembinder et al. (2018) and add a difference-in-differences estimator proposed by Callaway and Sant'Anna (2021). We follow Bessembinder et al. (2018) description, and apart from Bessembinder et al. (2018) estimator, we use Fama-MacBeth and pooled OLS to estimate the model given by Equation 12,

$$Y_{it} - \mathbb{E}\left[Y_{it}|f(X_{it-1})\right] = \alpha + \tau^{FM} D_{it} + \epsilon_{it} \, .$$

As we have mentioned before, to identify the treatment effect with Fama-MacBeth and pooled OLS, we need sequential ignorability assumption along with no spillover effects. Furthermore, let us emphasize that with different sets of benchmark returns or factor models ($\mathbb{E}\left[r_{it}|f(X_{it})\right]$), we use bad controls embedded in $f(X_{it})$.

We use four types of benchmarks: "None" stands for raw returns, where we do not adjust. CBBR-5 and CBBR-14 use the characteristic-based benchmark returns with five and fourteen variables proposed by Bessembinder et al. (2018). FF5 stands for the Fama-French 5-factor model's predicted returns. We use three sets of estimators; the first column for each benchmark estimate is the Fama and MacBeth (1973) type of estimation (denoted by "FMB") that weighs each time period equally. The second "Pool" uses a pooled OLS that weights each observation equally. "C-S'A" method uses the estimator proposed by Callaway and Sant'Anna (2021). Table 7 shows the results.

Table 7: Average excess returns (in pp) calculated with panel calendar-time methods

	None			**FF5**		
	FMB	Pooled	C-S'A	FMB	Pooled	C-S'A
$\hat{\alpha}$	1.510***	1.284***	1.164***	0.607***	0.402*	0.302
SE	(0.260)	(0.249)	(0.188)	(0.194)	(0.185)	(0.186)
$\hat{\tau}^{FM}$	−0.465***	−0.4049*	-0.132	−0.474***	−0.351**	-0.126
SE	(0.067)	(0.158)	(0.105)	(0.067)	(0.127)	(0.094)
	CBBR-5			**CBBR-14**		
	FMB	Pooled	C-S'A	FMB	Pooled	C-S'A
$\hat{\alpha}$	−0.244	−0.059	0.0875	−0.241	−0.061	0.062
SE	(0.294)	(0.2670)	(0.203)	(0.294)	(0.267)	(0.200))
$\hat{\tau}^{FM}$	0.309***	−0.038	-0.127	0.291***	−0.021	-0.102
SE	(0.066)	(0.165)	(0.101)	(0.067)	(0.166)	(0.109)
Observations: 2,196,831						

*Each column block refers to the benchmark return used. "None" uses raw returns, "CBBR-5" and "CBBR-14" use the characteristic-based benchmark returns with 5 and 14 variables proposed by Bessembinder et al. (2018), and FF5 uses the predicted returns of the Fama-French 5-factor model. FMB stands for the estimation performed by Fama and MacBeth (1973), pooled stands for pooled OLS estimation, and C-S'A for estimator proposed by Callaway and Sant'Anna (2021). Standard errors for pooled OLS and C -S 'A are clustered at the firm level and calendar time. Fama-MacBeth standard errors incorporate Newey-West correction with four lags. *** stands for 0.1%, ** stands for 0.5% and * stands for 1% significance levels.*

With FMB and pooled OLS estimations, the sign, and the magnitude change when we use different benchmark returns. We get similar results as Bessembinder et al. (2018), as the $\hat{\tau}^{FM}$ is insignificant in the case of the CBBR-5 and CBBR-14 estimations for the pooled OLS. Interestingly, when we use the "C-S'A", none of the estimated τ^{FM} are significant even at 5%, but the estimated coefficients are pretty stable across different benchmark returns and in its magnitude is close to the value of classical matching (−0.1108). We would interpret these results cautiously as we believe bad controls may bias the results.

5 Heterogeneity analysis

Our approach to estimating long-run abnormal returns to mergers allows us to address concerns about unobserved determinants that affect estimates. In this section, we first focus on questions commonly explored in existing merger studies: How do market conditions affect the market performance of the acquirer? (see e.g. Bessembinder et al. 2018 or Malmendier et al. 2018) Are the return estimates affected by the prior over-or undervaluation of the acquirer? (see e.g. Raghavendra Rau and Vermaelen 1998, Malmendier et al. 2018) Are the return estimates different for private versus public targets? (see e.g. Fuller et al. 2002, Betton et al. 2008, Malmendier et al. 2018). As a second step, we investigate some commonly concerned mechanisms. Which mergers are particularly likely to generate negative or positive cumulative abnormal returns?

In the following analysis, we use CAR values from the SSC-C5 model and take the values as given, thus neglecting modeling uncertainties of CARs. We model these CARs with baseline average and deterministic event-time trends. In all cases, we report these findings, and one shall keep in mind, that significant results only refer to conditional on CARs given, implying rather suggestive evidence not definite.[34]

5.1 Differences in market performances

Different time periods and market conditions

Let us consider when different circumstances were given during the M&A event. First, we revisit the results of Bessembinder et al. (2018), where they investigate if there are differences along different time periods. Instead of using pre-defined time periods we do year-by-year rolling regressions with α_t and event-time deterministic trend δ_t,

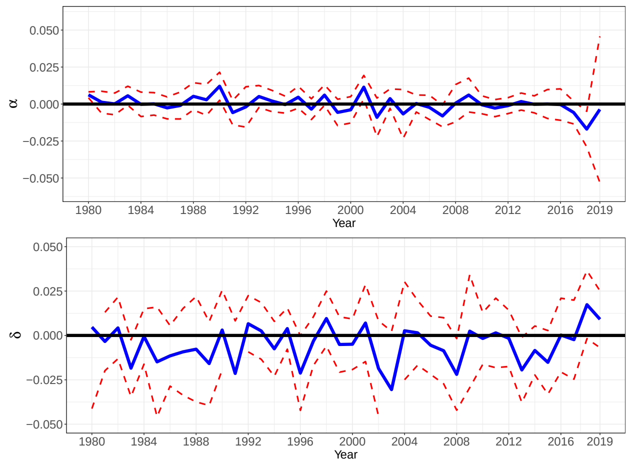
$$CAR_{iet} = \alpha_t + \delta_t e + u_{iet} \,,$$

where CAR_{iet} stands for the cumulative abnormal returns for M&A event i at event-time period e in year t. We use event-time deterministic trend e to capture systematic increases or decreases in CARs. Figure 7 shows the resulting estimates via OLS. We find no evidence of systematic negative α value or trends in CARs even if we neglect modeling uncertainties of the used CARs. When we pool over all time periods, we find a negative and significant overall event-time trend in CARs as Table 8, column 1 shows. Although it is significant, we do not consider it as evidence of negative time-event trends as this measure neglects the modeling uncertainties of CARs. However, we find it necessary to report as in the following results we find significant event-time trends for the benchmark cases, where the same issue holds.

Secondly, we revisit the exercise by Bessembinder et al. (2018), who investigates how "cold" vs. "hot" market conditions affect CAR outcomes using the investor sentiment intro-

[34]Note that these findings do not contradict previous results of no systemic differences in CARs as e.g. Table 5 includes the modeling uncertainties as well.

Figure 7: Time-varying parameters for CARs (pp)



Year-by-year rolling regressions are used to estimate the parameters: $CAR_{iet} = \alpha_t + \delta_t \times e + u_{iet}$*, where e stands for event-time deterministic trend. Estimation is via OLS, standard errors are clustered at the event level. Dashed red lines show 95% CI based on conditional standard errors, where we condition CAR values as given.*

duced by Baker and Wurgler (2006). Table 8, columns 2-3 show our results, where we find negative deterministic trends, for cold market conditions (column 2). This means that when investor sentiment by Baker and Wurgler (2006) is below the median, CARs with each month tend to be lower by one basis point. In the 36-month horizon, it means 33.7 basis points. Note, again we neglect the modeling uncertainties for CARs, thus this is rather suggestive evidence.

As a second exercise, we revisit Malmendier et al. (2018) to see if there is any heterogeneity in CARs among i) acquirers with different Tobin-q values, ii) if the target firm is public or private, and iii) number of bidders.

Acquirer's Tobin-q

First, we address concerns about the prior finding that highly valued acquirers tend to underperform in the long run. The argument is that the subsequent reversal in acquirers' market valuation might not have been caused by the merger but would have occurred even in the absence of the takeover. For example, temporarily overvalued firms might choose to acquire less highly valued targets to attenuate the reversal in their (over-) valuation. (see e.g. Raghavendra Rau and Vermaelen 1998, Shleifer and Vishny 2003 or Rhodes-Kropf and Viswanathan 2004). Following Malmendier et al. (2018), we define a dummy variable indicating acquirers with above-median market-to-book ratio. The estimation results are shown in column 1 of Table 9. Similarly to Malmendier et al. (2018), we find no difference in the merger effects estimated for highly and less highly valued acquirers. This result also suggests that early results on the acquirer's Tobin-q might be affected by the lack of a proper

Table 8: Market conditions and differences in CARs (pp)

Model:	Full sample	Market condition	
		Cold	Hot
	(1)	(2)	(3)
α	-0.0015	0.0321	-0.0347*
	(0.0073)	(0.0206)	(0.0178)
Event-time	-0.0060***	-0.0102***	-0.0017
	0.0016	(0.0023)	(0.0023)
Fit statistics			
Observations	78,662	38,115	40,547
R^2	0.00165	0.00443	0.00015

*The regression equation is $CAR_{ie} = \alpha + \delta \times e + u_{ie}$, where e stands for event time. A month is regarded as a hot market if the Baker and Wurgler (2006) investor sentiment is above the median over the period 1980–2021. Clustered standard errors are in parentheses clustered at the event level, with significance codes: ***: 0.01, **: 0.05, *: 0.1. Note that we neglect modeling uncertainty for getting CAR values.*

counterfactual and may have to be interpreted cautiously.

Public vs. private target

We distinguish between the return implications of public and private acquisitions, a standard robustness check, in prior studies. Many papers argues that announcement (or short-term) returns are significantly lower in acquisitions of public targets (see, e.g., Fuller et al. 2002, Betton et al. 2008). They attribute this finding to private information about private targets (Makadok and Barney 2001, or Capron and Shen (2007)) or liquidity discounts for private targets (Fuller et al., 2002).

In Table 9, column 2, we regress public target firms and their interaction with the deterministic time-event trend on CARs. We found no evidence that acquisitions of public firms would destroy value, nor that private firms would generate any.

Number of bidders

Moeller et al. (2004), Malmendier et al. (2018) or Eckbo et al. (2018) investigate if the number of bidders that indicates more contest for the target firm, and leads the winner to underperform in the long run. Similar to the literature, we found no evidence that bidding competition would decrease the acquirers' return. Column 3 of Table 9 shows insignificant, but positive cumulative average returns compared to bidding competitions where there is one more competitor on average.[35]

[35] We run several other specifications compared to using the number of bidders as a continuous variable:

Table 9: Acqurier's market conditions, M&A conditions and differences in CARs (pp)

Model:	Acq. q (1)	Publ./priv. (2)	No. bidders (3)
α	0.0040	-0.0143	0.0241
	(0.0103)	(0.0141)	(0.0232)
Event-time (e)	-0.0040*	-0.0057*	-0.0127
	(0.0024)	(0.0032)	(0.0101)
High MB	-0.0109		
	(0.0146)		
High MB$\times e$	-0.0039		
	(0.0033)		
Public		0.0156	
		(0.0172)	
Public$\times e$		0.0009	
		(0.0039)	
No. bidders			-0.0247
			(0.0207)
No. bidders$\times e$			0.0065
			(0.0098)
Fit statistics			
Observations	78,662	59,552	78,662
R^2	0.00248	0.00137	0.00197

*The regression equation is $CAR_{ie} = \alpha + \delta_0 \times e + \gamma x_{ie_0} + \delta_1(e \times x_{i,e_0}) + u_{ie}$, where e stands for event time, variables included are x_{ie_0} referring for fixed values with pre-announcement value or property of M&A. Clustered standard-errors are in parentheses clustered at the event level, with significance codes: ***: 0.01, **: 0.05, *: 0.1. Note that we neglect modeling uncertainty for getting CAR values.*

5.2 Possible Mechanisms

We now ask which mergers are particularly likely to generate negative abnormal returns. That is, can we reveal the channels or possible mechanisms that determine the returns to mergers? We discuss and test hypotheses that may explain the estimated return implications. We also discuss to what extent the potential explanations might imply positive abnormal returns.

Form of payment

First, we look at the form of payment: cash, stock, or other. Theory suggests that equity-financed deals should earn significantly lower returns relative to cash-financed deals, as the fact that management opts for equity financing hints to the market that the firm's stock is overvalued. Eckbo et al. (2018), considers the method of payment as a sign of trust. They find that the fraction of stock financing is higher when targets are better informed about the bidder, consistent with the idea that bidders offer stock when they are concerned about target adverse selection. In addition, they report that the composition of the payment method over time is strongly correlated with the presence of private bidders who exert pressure on public bidders to pay in cash.

Similarly to Eckbo et al. (2018), we separate two subsamples: firms with high market-to-book ratios (high MB) and low market-to-book ratios (low MB) firms. A high (above median) market-to-book indicates a high potential for overvaluation of bidder shares, and the test is whether all-stock financed deals underperform all-cash deals within the sample of high MB bidders. Table 10, columns 1-2 show the results for both subsamples, where the baseline scenario is paying with a mixed solution (both cash and stocks). Similarly to Malmendier et al. (2018) or Eckbo et al. (2018), we find no evidence for under-performance in none of the sub-groups.

Leverage

Another proxy for low strategic and financial flexibility is high leverage. Financial flexibility is a key driver of capital structure decisions in firms both empirically (Lang et al. 1996; Marchica and Mura 2010) and theoretically DeAngelo et al. (2011). Acquisitions, through an increase in acquiring firm's leverage, may reduce future growth (see, e.g., Penman et al. 2007). Method of payment may be connected to an increase in leverage, that is, a heavier burden in cash-financed deals that rely on increasing debt obligations, but may also be driven

using as a factor variable or modeling one/two/more than two bidders separately. The results are the same in all cases.

Table 10: Acqurier's market conditions, M&A conditions and differences in CARs (pp)

Model	Type of payment		Leverage	Same Ind.	Acq. size	Rel. size
	High MB	Low MB				
	(1)	(2)	(3)	(4)	(5)	(6)
α	-0.0085	0.0136	-0.0012	-0.0211	0.0061	-0.0106
	(0.0159)	(0.0146)	(0.0073)	(0.0154)	(0.0115)	(0.0099)
Only stocks	-0.0022	-0.0280				
	(0.0265)	(0.0271)				
Only cash	0.0069	-0.0128				
	(0.0231)	(0.0239)				
e	-0.0087***	-0.0026	-0.0060***	-0.0063*	-0.0088***	-0.0069***
	(0.0031)	(0.0033)	(0.0017)	(0.0033)	(0.0026)	(0.0021)
Only stocks $\times e$	-0.0006	-0.0044				
	(0.0055)	(0.0065)				
Only cash $\times e$	0.0028	-0.0018				
	(0.0053)	(0.0057)				
Leverage			-0.0003			
			(0.0004)			
Leverage $\times e$			-0.0000			
			(0.0003)			
Same Industry (FF12)				0.0003		
				(0.0002)		
Same Industry (FF12) $\times e$				0.0000		
				(0.0001)		
Large					-0.0145	
					(0.0145)	
Large $\times$ e					0.0054*	
					(0.0033)	
Rel. Large						0.0180
						(0.0146)
Rel. Large $\times$ e						0.0018
						(0.0033)
Fit statistics						
Observations	39,366	39,296	78,662	78,662	78,662	78,662
R^2	0.00365	0.00145	0.00166	0.00174	0.00266	0.00193

*The regression equation is $CAR_{ie} = \alpha + \delta_0 \times e + \gamma x_{ie_0} + \delta_1(e \times x_{i,e_0}) + u_{ie}$, where e stands for event time, variables included are x_{ie_0} referring for fixed values with one month/period pre-announcement value or property of M&A. High MB stands for acquirer firms that have higher market-to-book ratio than median value, low MB stands for firms that have lower values. The market-to-book ratio is the inverse of the book-to-market ratio defined in Table A1. Only stocks and only cash transactions stand for 'consideration structure = shares' and 'consideration structure = casho' from SDC, following Eckbo et al. (2018). Ind. stands for the same industry for acquirer and target firm according to Fama-French 12 industry categorization similarly to Malmendier et al. (2018). Acq. size investigates the possible effect of acquirer's size: a firm is considered 'Large' if its market capitalization is higher than the median value. Rel. size stands for relative size, defined as the transaction value ratio to the market size of the acquirer firm. Rel. Large takes the value of 1 if the relative size is larger than the median and takes 0 otherwise. Clustered standard errors are in parentheses clustered at the event level, with significance codes: ***: 0.01, **: 0.05, *: 0.1. Note that we neglect modeling uncertainty for getting CAR values.*

by the leverage of the target. For instance with cash deals, acquirers may use their cash holdings and take on additional debt in order to finance the deal.

We compute leverage as the ratio of debt (current liabilities plus long-term debt) to market value (total shares outstanding times the price of shares at the end of the last month). Similarly to Malmendier et al. (2018), we have found no regression-based evidence that higher leverage on average would imply lower CARs.

Integration costs

The cost of post-merger integration is often cited as a key reason for poor post-merger performance, and the underestimation of these costs is one of the top mistakes companies make in acquisitions (see, e.g., Bereskin et al. 2018 or Renneboog and Vansteenkiste 2019). The most commonly mentioned underlying factor is "cultural differences" and Weber and Camerer (2003) have illustrated experimentally how different organizational cultures introduce merger costs. If integration costs and their underestimation are important explanations for the merger effect, mergers, where post-merger integration issues are more likely to arise, should experience stronger underperformance.

We consider some possible factors contributing to the cost of integration: relatedness (in terms of industry) and size (in terms of acquirer absolute and relative size).

Relatedness

A large strand of management and corporate-finance literature has shown that mergers between related firms tend to generate higher value than diversifying mergers (Chatterjee 1986; Singh and Montgomery 1987; Morck et al. 1990; Cartwright and Cooper 1993). Their definitions of relatedness are mainly concerned with similarities in production technology, scientific research, products, and industries. However, more recent studies find different results (Schneider and Spalt 2016; Akbulut and Matsusaka 2010) that point to identification as an essential issue: firms that participate in diversifying mergers are different from firms that engage in concentrating acquisitions.

Our approach is similar to Malmendier et al. (2018); we use the Fama-French 12-industry classification to distinguish related versus diversifying mergers. An acquisition bid is related if the acquirer is in the same industry as the target, and it is diversifying otherwise. The total sample has 1,686 cases for related M&A events and 548 unrelated ones.

Table 10, column 4, shows our results. In contrast to Malmendier et al. (2018) – who find negative but non-significant effects in case of diversified events – we find no evidence of either value adding or destroying based on FF-12 classification.

Size

As a second correlate of the integration cost, we look at size. Integration costs tend to be more severe the larger the target is relative to the acquirer, and the more difficult it is to transform the target's corporate culture. At the end of the spectrum, a small firm acquiring a large target will incur significant costs training the target firm's employees to adhere to the acquirer's business practices. Indeed, sizeable relative target size has been associated with significantly lower returns to mergers, at least in the short run. Here, too, causal interpretation is complicated because acquirers of large target firms may differ from acquirers of small firms. For instance, mature firms with declining profits may acquire large firms, whereas young growth firms may tend to acquire small firms. (Malmendier et al., 2018)

We hypothesize that, if integration costs play a role in explaining our findings, acquirers underperform more when they are relatively small compared to the target. We start from the absolute size of the acquirer. In column 5 of Table 10, we estimate the merger separately for acquirers with above-and below-median market capitalization. We find that large firms tend to outperform smaller ones in absolute value by 0.5 basis points at each period after the event. This is aligned with Malmendier et al. (2018), who find smaller firms tend to perform worse. Note that this result is significant only at 10% level, neglecting uncertainty on constructing CAR values.

A more direct test is based not on absolute acquirer size, but on the relative sizes of target and acquirer. We calculate the relative target size as the ratio of the transaction value and the acquirers' market capitalization. Column 6 of Table 10 shows the results of the corresponding estimation, where we split the sample into above-median and below-median relative target sizes. Consistent with the integration-cost hypothesis, we find that deals involving relatively large targets induce larger adverse effects, though the differences are insignificant.

6 Conclusion

This paper proposes a new method – stacked synthetic control – to create better counterfactuals for acquirer firms. Stacked synthetic control provides flexible framework with unbalanced panel data, where it is important to match on multiple (firm) characteristics with a flexible weighting of pre-event periods. Currently, existing methods that use one-period ahead matching or other alternatives from the potential outcome framework, such as time-series cross-sectional matching or difference-in-differences estimators, do not allow for such flexibility. We show how our method connects to them in terms of identification and estimation. Finally, our method allows us to investigate heterogeneity in cumulative abnormal returns and test some popular hypotheses from the finance literature.

Our findings show that M&As are, on average, neither value-adding nor value-destroying, as the long-run market excess returns are on expectation zero. Other competing methods show zero or negative monthly excess returns between 10-50 basis points. We also find these results in the existing literature, where different papers claim no or slightly negative effects (Renneboog and Vansteenkiste, 2019).

At the end of the paper, we analyze the sources of heterogeneity and test some popular hypotheses in the literature (see, e.g., Malmendier et al. 2018). Although we do not find decisive evidence, our results suggest that when market conditions are cold (investor sentiment based on Baker and Wurgler (2006) below the median), CARs tend to be lower by one basis point. Furthermore, we find suggestive evidence that integration costs are relevant; when larger firms acquire relatively more minor firms, the expected CARs are slightly higher.

References

A. Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, 2021.

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

A. Abadie, A. Diamond, and J. Hainmueller. 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, 2010.

A. Abadie, A. Diamond, and J. Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015.

M. E. Akbulut and J. G. Matsusaka. 50+ years of diversification announcements. *Financial Review*, 45(2):231–262, 2010.

Y. Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Mcarkets*, 5(1):31–56, 2002.

A. Ang, R. J. Hodrick, Y. Xing, and X. Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.

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

A. C. Baker, D. F. Larcker, and C. C. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022.

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680, 2006.

B. M. Barber and J. D. Lyon. Detecting abnormal operating performance: The empirical power and specification of test statistics. *Journal of Financial Economics*, 41(3):359–399, 1996.

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

F. Bereskin, S. K. Byun, M. S. Officer, and J.-M. Oh. The effect of cultural similarity on mergers and acquisitions: Evidence from corporate social responsibility. *Journal of Financial and Quantitative Analysis*, 53(5):1995–2039, 2018.

H. Bessembinder and F. Zhang. Firm characteristics and long-run stock returns after corporate events. *Journal of Financial Economics*, 109(1):83–102, 2013.

H. Bessembinder, M. J. Cooper, and F. Zhang. Characteristic-Based Benchmark Returns and Corporate Events. *The Review of Financial Studies*, 32(1):75–125, 04 2018.

S. Betton, B. E. Eckbo, and K. S. Thorburn. Chapter 15 - Corporate Takeovers. In B. E. Eckbo, editor, *Handbook of Empirical Corporate Finance*, Handbooks in Finance, pages 291–429. Elsevier, San Diego, 2008.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2023.

N. M. Boyson, N. Gantchev, and A. Shivdasani. Activism mergers. *Journal of Financial Economics*, 126(1):54–73, 2017.

A. Brav, C. Geczy, and P. A. Gompers. Is the abnormal return following equity issuances anomalous? *Journal of Financial Economics*, 56(2):209–249, 2000.

C. Caetano, B. Callaway, S. Payne, and H. S. Rodrigues. Difference in differences with time-varying covariates. *arXiv preprint arXiv:2202.02903*, 2022.

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

L. Capron and J.-C. Shen. Acquisitions of private vs. public firms: Private information, target selection, and acquirer returns. *Strategic Management Journal*, 28(9):891–911, 2007.

S. Cartwright and C. L. Cooper. The role of culture compatibility in successful organizational marriage. *Academy of Management Perspectives*, 7(2):57–70, 1993.

M. D. Cattaneo, Y. Feng, and R. Titiunik. Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536):1865–1880, 2021.

M. D. Cattaneo, Y. Feng, F. Palomba, and R. Titiunik. Uncertainty quantification in synthetic controls with staggered treatment adoption. *arXiv preprint arXiv:2210.05026*, 2023.

S. Chatterjee. Types of synergy and economic value: The impact of acquisitions on merging and rival firms. *Strategic Management Journal*, 7(2):119–139, 1986.

J. Chen and J. Roth. Log-like? ATEs defined with zero outcomes are (arbitrarily) scale-dependent. *arXiv preprint arXiv:2212.06080*, 2023.

C. Cinelli, A. Forney, and J. Pearl. A crash course in good and bad controls. *Sociological Methods & Research*, 2022.

D. Clarke, D. Pailañir, S. Athey, and G. Imbens. Synthetic difference in differences estimation. *arXiv preprint arXiv:2301.11859*, 2023.

M. J. Cooper, H. Gulen, and M. J. Schill. Asset growth and the cross-section of stock returns. *The Journal of Finance*, 63(4):1609–1651, 2008.

K. Daniel, M. Grinblatt, S. Titman, and R. Wermers. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3):1035–1058, 1997.

W. N. Davidson, D. Dutia, and L. Cheng. A re-examination of the market reaction to failed mergers. *The Journal of Finance*, 44(4):1077–1083, 1989.

C. De Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26 (3):C1–C30, 2023.

H. DeAngelo, L. DeAngelo, and T. M. Whited. Capital structure dynamics and transitory debt. *Journal of Financial Economics*, 99(2):235–261, 2011.

M. Dong, D. Hirshleifer, S. Richardson, and S. H. Teoh. Does investor misvaluation drive the takeover market? *The Journal of Finance*, 61(2):725–762, 2006.

N. Doudchenko and G. W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.

B. E. Eckbo, R. W. Masulis, and Ø. Norli. Seasoned public offerings: Resolution of the 'new issues puzzle'. *Journal of Financial Economics*, 56(2):251–291, 2000.

B. E. Eckbo, R. W. Masulis, and Ø. Norli. Security offerings. *Handbook of empirical corporate finance*, pages 233–373, 2007.

B. E. Eckbo, T. Makaew, and K. S. Thorburn. Are stock-financed takeovers opportunistic? *Journal of Financial Economics*, 128(3):443–465, 2018.

M. Ewens, R. Peters, and S. Wang. Acquisition prices and the measurement of intangible capital. *Working Paper*, 2018.

E. F. Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306, 1998.

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

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

K. Fuller, J. Netter, and M. Stegemoller. What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions. *The Journal of Finance*, 57(4):1763–1793, 2002.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

K. Imai, I. S. Kim, and E. H. Wang. Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 2019.

G. W. Imbens and J. M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86, 2009.

J. W. Kolari, S. Pynnonen, and A. M. Tuncez. Further evidence on long-run abnormal returns after corporate events. *The Quarterly Review of Economics and Finance*, 81: 421–439, 2021.

S. P. Kothari and J. B. Warner. Econometrics of event studies. In *Handbook of empirical corporate finance*, pages 3–36. Elsevier, 2007.

L. Lang, E. Ofek, and R. Stulz. Leverage, investment, and firm growth. *Journal of Financial Economics*, 40(1):3–29, 1996.

J. Lewellen. The cross-section of expected stock returns. *Critical Finance Review*, 4(1):1–44, 2015.

X. Li and X. Zhao. Propensity score matching and abnormal performance after seasoned equity offerings. *Journal of Empirical Finance*, 13(3):351–370, 2006.

J. Lintner. Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4):587–615, 1965.

L. Liu, Y. Wang, and Y. Xu. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 2022.

Y. Liu, W. Wang, and F. Zhang. Replicating long-run event studies. Working paper, Krannert Graduate School of Management at Purdue University and Cox School of Business at Southern Methodist University, 2023. URL https://shimengwang.com/__static/320572ad16a04747f708c511d7d43a2b/wp_liuwangzhang(3).pdf?dl=1.

T. Loughran and J. R. Ritter. The new issues puzzle. *The Journal of Finance*, 50(1):23–51, 1995.

T. Loughran and J. R. Ritter. The operating performance of firms conducting seasoned equity offerings. *The Journal of Finance*, 52(5):1823–1850, 1997.

T. Loughran and J. R. Ritter. Uniformly least powerful tests of market efficiency. *Journal of Financial Economics*, 55(3):361–389, 2000.

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

E. Lyandres, L. Sun, and L. Zhang. The New Issues Puzzle: Testing the Investment-Based Explanation. *The Review of Financial Studies*, 21(6):2825–2855, 12 2007.

J. D. Lyon, B. M. Barber, and C.-L. Tsai. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance*, 54(1):165–201, 1999.

R. Makadok and J. B. Barney. Strategic factor market intelligence: An application of information economics to strategy formulation and competitor intelligence. *Management Science*, 47(12):1621–1638, 2001.

U. Malmendier, E. Moretti, and F. S. Peters. Winning by Losing: Evidence on the Long-run Effects of Mergers. *The Review of Financial Studies*, 31(8):3212–3264, 03 2018.

M.-T. Marchica and R. Mura. Financial flexibility, investment ability, and firm value: evidence from firms with spare debt capacity. *Financial Management*, 39(4):1339–1365, 2010.

S. B. Moeller, F. P. Schlingemann, and R. M. Stulz. Firm size and the gains from acquisitions. *Journal of Financial Economics*, 73(2):201–228, 2004.

R. Morck, A. Shleifer, and R. W. Vishny. Do managerial objectives drive bad acquisitions? *The Journal of Finance*, 45(1):31–48, 1990.

J. Netter, M. Stegemoller, and M. B. Wintoki. Implications of data screens on merger and acquisition analysis: A large sample study of mergers and acquisitions from 1992 to 2009. *The Review of Financial Studies*, 24(7):2316–2357, 2011.

S. H. Penman, S. A. Richardson, and I. Tuna. The book-to-price effect in stock returns: accounting for leverage. *Journal of Accounting Research*, 45(2):427–467, 2007.

G. M. Phillips and A. Zhdanov. R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies*, 26(1):34–78, 2013.

Z. Porreca. Synthetic difference-in-differences estimation with staggered treatment timing. *Economics Letters*, 220:110874, 2022.

A. K. Purnanandam and B. Swaminathan. Are IPOs really underpriced? *The Review of Financial Studies*, 17(3):811–848, 2004.

P. Raghavendra Rau and T. Vermaelen. Glamour, value and the post-acquisition performance of acquiring firms. *Journal of Financial Economics*, 49(2):223–253, 1998. ISSN 0304-405X.

L. Renneboog and C. Vansteenkiste. Failure and success in mergers and acquisitions. *Journal of Corporate Finance*, 58:650–699, 2019.

M. Rhodes-Kropf and S. Viswanathan. Market valuation and merger waves. *The Journal of Finance*, 59(6):2685–2718, 2004.

M. Rhodes-Kropf, D. T. Robinson, and S. Viswanathan. Valuation waves and merger activity: The empirical evidence. *Journal of Financial Economics*, 77(3):561–603, 2005.

J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, pages 550–560, 2000.

J. Roth. Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322, 2022.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.

D. B. Rubin. *Matched sampling for causal effects*. Cambridge University Press, 2006.

P. G. Savor and Q. Lu. Do stock mergers create value for acquirers? *The Journal of Finance*, 64(3):1061–1097, 2009.

C. Schneider and O. Spalt. Conglomerate investment, skewness, and the CEO long-shot bias. *The Journal of Finance*, 71(2):635–672, 2016.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

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

H. Singh and C. A. Montgomery. Corporate acquisition strategies and economic performance. *Strategic Management Journal*, 8(4):377–386, 1987.

R. G. Sloan. Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, pages 289–315, 1996.

E. A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1, 2010.

R. A. Weber and C. F. Camerer. Cultural conflict and merger failure: An experimental approach. *Management Science*, 49(4):400–415, 2003.

J. M. Wooldridge. Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*, 2021.

B. Zeldow and L. A. Hatfield. Confounding and regression adjustment in difference-in-differences studies. *Health Services Research*, 56(5):932–941, 2021.

A Appendix

Table A1: Definition of the C5 and C14 firm characteristics as in Bessembinder et al. (2018)

Characteristics in the C5 model	
Log size	Natural log of market capitalization, which is stock price (prc in CRSP monthly stock file) times number of shares outstanding (shrout), at the end of the prior month
Log book-to-market ratio	Natural log of the book-to-market ratio at the end of the prior month. Book value is the firm's common equity (Compustat item ceq) in the latest annual report. Market value is the firm's market capitalization (prc times shrout) at the end of the prior month reported in CRSP.
Momentum	Buy-and-hold stock returns over months (-12,-2) before the month of interest
ROA	Income before extraordinary items (ib) divided by average total assets (at) in the year
Asset growth	Natural log of the ratio of total assets (at) at the end of the year to total assets at the beginning of the year, following Cooper et al. (2008)
Additional nine characteristics in the C14 model	
Beta	Market beta is estimated using monthly excess stock returns and market risk premiums over the preceding 60 months. We require a minimum of six data points for the accuracy of the estimation
Accrual	Change in working capital from the last year minus depreciation and amortization (dp), divided by average total assets (at) in the year, following Sloan (1996). Working capital equals current assets (act) minus cash and short-term investment (che) minus current liabilities (lct) plus debt in current liabilities (dlc) plus income taxes payable (txp). Missing act, che, lct, dlc, txp, and dp are replaced with zero
Dividend	Dividends per share over the prior 12 months divided by the price at the end of the prior month
Log LR return	Natural log of buy-and-hold stock returns over months (-13,-36) before the month of interest
Idiosyncratic risk	In each month, we compute the standard deviation of the residual daily stock returns in the Fama and French (1993) three-factor regression, following Ang et al. (2006). Idiosyncratic risk is the average standard deviation over the prior 12 months
Illiquidity	The average daily ratio of absolute stock return to dollar trading volume during the prior 12 months, as defined by Amihud (2002)
Turnover	Average monthly turnover (shares traded divided by shares outstanding) during the prior 12 months
Leverage	Debt in current liabilities (dlc) plus long-term debt (dltt), divided by market capitalization (prc times shrout in CRSP) at the end of the last month. Missing dlc and dltt are replaced with zero
Sales/price	Sales (sale) divided by market capitalization (prc times shrout in CRSP) at the end of the last month.

We measure these characteristics following Lewellen (2015). All variables are created using data from the CRSP stock price files and the Compustat quarterly data.

Table A2: Different matching methods for event firms

Matching method	Corporate Event	Replace if delisted	Further constraint	Papers introducing or using the method
Closest market capitalization on December 31	M&A, SEO, dividends	Y	-	Loughran and Ritter (1995)
Closest market capitalization on December 31	IPO	Y	For IPO matching firm must have been traded publicly for 5 years	Loughran and Ritter (1995), Bessembinder and Zhang (2013), Kolari et al. (2021)
Closest market capitalization and filter to Book-to-Market Ratio (BMR) on proceeding December 31	M&A, SEO, dividends	Y	Market capitalization between 70%-130%.	Bessembinder and Zhang (2013), Kolari et al. (2021)[36]
Market Capitalization and Book-to-Market Ratio (BMR) on one period before event time	M&A, SEO, dividends	Y	Market capitalization between 70%-130%.	Eckbo et al. (2007), Bessembinder et al. (2018), Liu et al. (2023)
Closest Book-to-Market Ratio after the event time is first available, the filter for the market capitalization (MC)	IPO	Y	MC is larger than the event firm's MC and less than 20 times. Matching firm is publicly traded for more than 3 years	Lyandres et al. (2007)[37], Bessembinder et al. (2018), Liu et al. (2023)
Closest operating income before depreciation and amortization (OIBD) relative to assets, with same industry	SEO	N	Not issued equity during the five years prior to the offering date. Same industry (2-digit SIC codes), with asset size as of the end of year 0 between 25 percent and 200 percent of the issuer are ranked by their year 0 OIBD relative to assets. The firm with the closest OIBD/assets ratio among these non issuing firms is picked as the matching firm. If no matching firm, neglect industry restriction and asset size within 90 percent to 110 percent of the issuer are ranked by OIBD/assets, and the firm with the closest, but higher, the ratio is chosen as the matching firm	Barber and Lyon (1996), Loughran and Ritter (1997)
Market capitalization in June and book-to-market ratio in December (t-1), using multiple firms	SEO (and other stock-related events)	N	order multiple times to decide and/or quantiles. Event firms are compared to firms in the same bracket.	Lyon et al. (1999)
Market capitalization and book-to-market ratio via (yearly) regressions	M&A	N	Estimate regressions of market capitalization and book-to-market ratio on cumulative returns. Get the F-value and rank firms. Closest to the event firm will be the matching firm.	Loughran and Vijh (1997)
Propensity score matching with market capitalization, book-to-market ratio, and cumulative returns	SEO	N	Matching firm has not initiated SEO in the past 3 years	Li and Zhao (2006)
Industry, sales, and EBITDA grid	IPO	N	Matching firm is in the same industry (FF48) and a 3x3 grid by sales and EBITDA is created to categorize. The closest in sales is chosen.	Purnanandam and Swaminathan (2004)

Figure A1: Distribution for the number of candidate control firms used for stacked synthetic control methods.

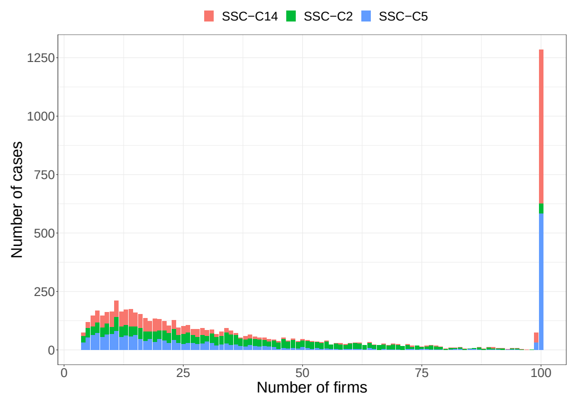
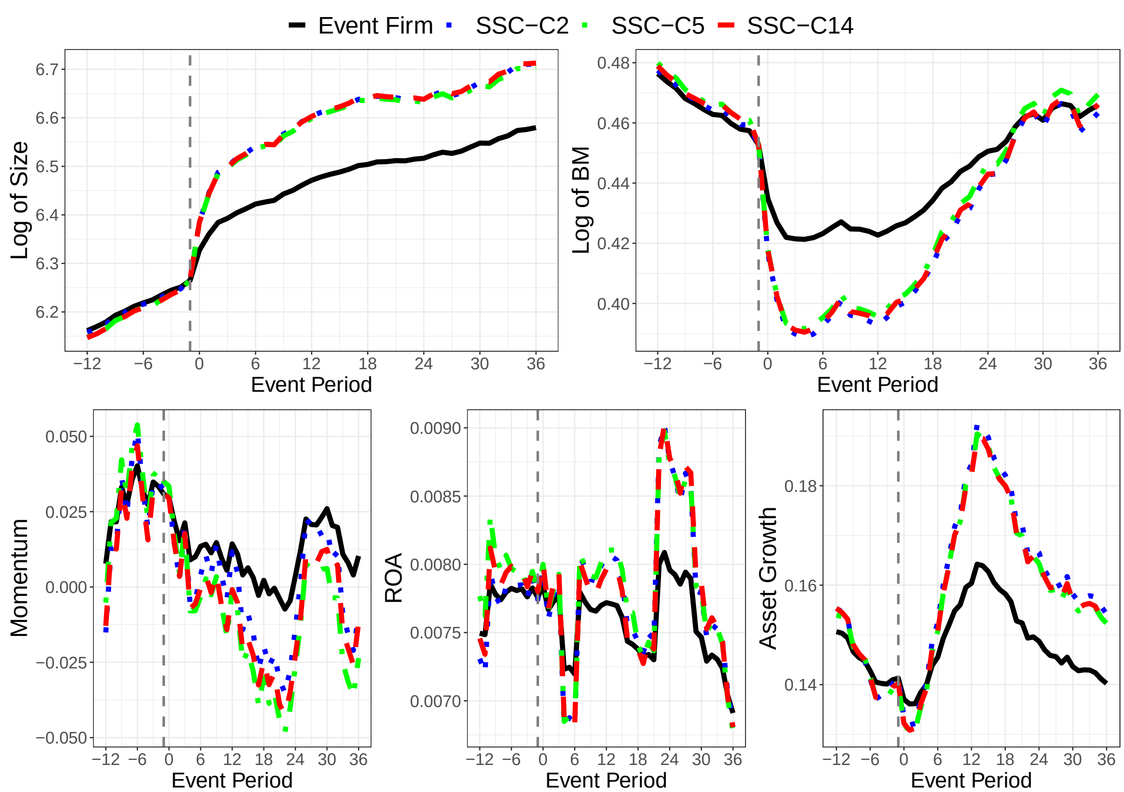


Figure A2: C5 variables and stacked synthetic control methods

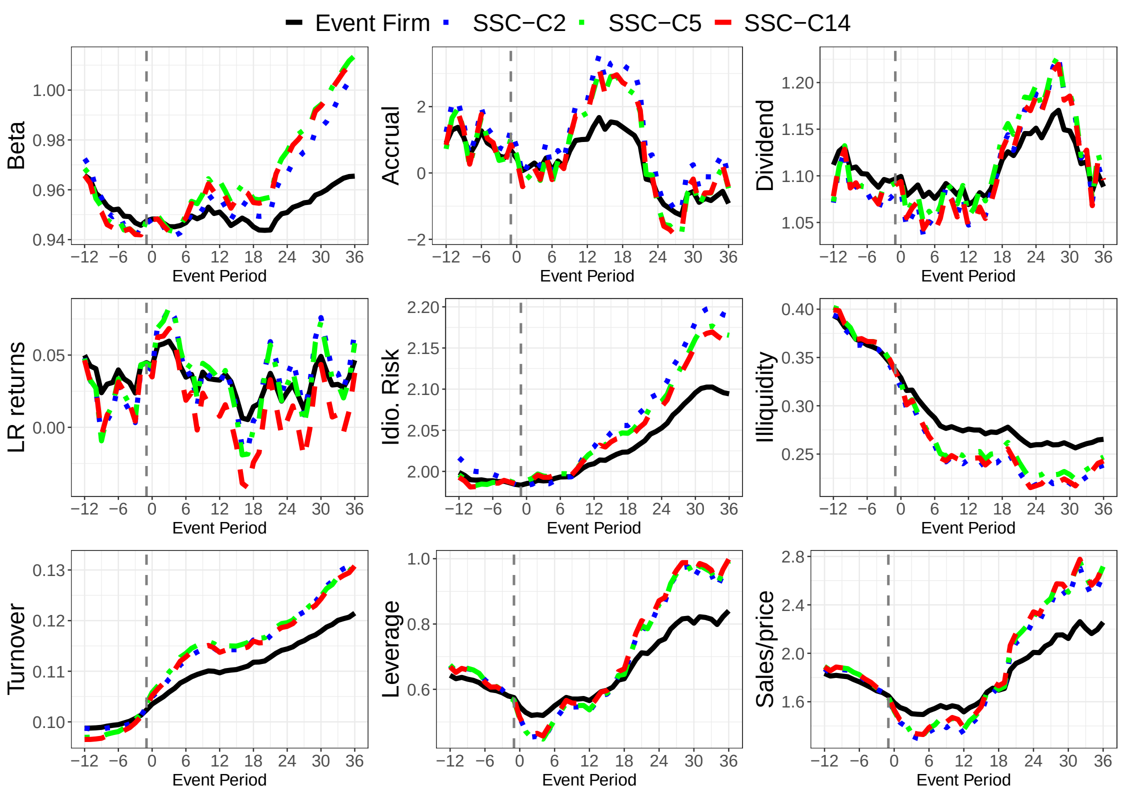


Averaged firm characteristics before 12 months and after the event for 36 months. SSC-C2 stands for stacked synthetic control with size and book-to-market variables. SSC-C5 stands for the stacked synthetic control method using C5 variables, whereas SSC-C14 uses the C14 variables.

Table A3: Test for parallel trend assumption by Roth (2022)

	Classical			TSCS-C5			SSC-C5		
Variable	Power	Bayes factor	Likelihood Ratio	Power	Bayes factor	Likelihood Ratio	Power	Bayes factor	Likelihood Ratio
C5 variables									
Log of BM	0.4865	0.9027	0.0082	0.4867	0.9024	0.0000	0.4999	0.8792	0.5576
Log of size	0.5018	0.8758	0.0000	0.5018	0.8758	0.0000	0.4999	0.8792	0.7743
Momentum	0.5018	0.8758	0.0383	0.5071	0.8665	1.0000	0.4998	0.8793	0.5378
ROA	0.5019	0.8757	0.0001	0.4312	1.0000	1.0000	0.5023	0.8750	0.5738
Asset growth	0.5051	0.8700	0.5686	0.4867	0.9024	0.0000	0.4988	0.8811	0.5600
Additional 9 variables for C14									
Beta	0.5018	0.8758	5.8602	0.5018	0.8758	0.0000	0.4999	0.8793	0.6115
Accrual	0.4625	0.9449	0.9035	0.5068	0.8671	0.0638	0.4987	0.8813	0.5925
Dividend	0.4312	1.0000	1.0000	0.4312	1.0000	1.0000	0.5027	0.8742	0.5830
Log of long run return	0.5018	0.8758	0.1514	0.5018	0.8758	97.1573	0.4999	0.8793	0.5878
Idiosyncratic risk	0.4312	1.0000	1.0000	0.4312	1.0000	0.0000	0.5028	0.8740	0.5782
Illiquidity	0.5018	0.8758	0.2421	0.5018	0.8758	0.0773	0.4999	0.8793	0.6528
Turnover	0.4880	0.9001	0.2210	0.4588	0.9514	0.0000	0.5028	0.8741	0.8331
Leverage	0.5018	0.8758	0.9315	0.5018	0.8758	0.0000	0.5001	0.8789	0.4631
Sales / price	0.4999	0.8791	11.4022	0.4999	0.8791	0.0000	0.4998	0.8794	0.4466
Returns	0.4899	0.8967	0.1386	0.4628	0.9443	0.0146	0.5001	0.8789	0.5935

Figure A3: Remaining C14 variables and stacked synthetic control methods



Averaged firm characteristics before 12 months and after the event for 36 months. SSC-C2 stands for stacked synthetic control with size and book-to-market variables. SSC-C5 stands for the stacked synthetic control method using C5 variables, whereas SSC-C14 uses the C14 variables.