Managerial Misdirection? How Firms Use Confounding Events to Obfuscate the Impact of Data Breaches[1]

Ruoyu Gao [a], Ronan Powell [a, 2], and Jared Stanfield [b]

[a] *UCD Michael Smurfit Graduate School of Business, University College Dublin, Ireland*

[b] *Price College of Business, University of Oklahoma, USA*

Draft: August 2025

Abstract

We examine the determinants and outcomes of firms using confounding strategies, such as announcing unrelated events or information, in response to experiencing severe data breaches. Using a comprehensive sample of data breach events from 2005 to 2022 in the U.S., we find that severe breaches significantly reduce firm value, but that confounding strategies by managers help to attenuate this effect in the short run. Further, we do not find strong evidence of a reversal of this effect in the year following the confounded breach using a calendar-time portfolio approach. We also find that managers are more likely to adopt confounding strategies when the breach is severe, and that characteristics such as managerial overconfidence, CEO equity holdings, CEO tenure, and prior misstatements by the firm increase the likelihood of confounding. Our study has important implications for understanding the impact of data breaches on firm value and in predicting managerial behavioral responses to breaches when severity and agency costs are high.

Keywords: Data breach; confounding events; managerial information asymmetry; agency costs

JEL: G10; G32.

[1] We are grateful to seminar participants at the UCD Doctoral Centre Spring Workshops & Research Symposium (Dublin), the Vietnam Symposium in Banking and Finance (VSBF) (Hanoi), the 2024 Annual Boca-ECGI Corporate Finance and Governance Conference (Madrid), the Irish Academy of Finance (IAF) PhD Conference (Cork), and the visit of the Scientific Advisory Board (SAB) of the Financial Mathematics and Computation (FMC) cluster to UCD (Dublin). We are also grateful to Cal Muckley, Kevin Gam, Siqi Liu, Hamid Yahyaei, Tommaso Tamburelli, Oliver Browne, Liam Gallagher, Matthew Spiegel, Elena Carletti, and Feng Zhang for their insightful comments. The authors gratefully acknowledge funding support from the UCD Michael Smurfit Graduate School of Business. Ruoyu Gao gratefully acknowledges funding support from the UCD China Scholarship Scheme (CSC).

[2] Corresponding author: ronan.powell@ucd.ie

1 Introduction

Evidence suggests corporations strategically disseminate information to the market.[3] Further, investors may underreact to relevant firm information or news due to limited attention and distraction by extraneous events (e.g., Hirshleifer, Lim, and Teoh, 2009). While regulation mandates that firms must communicate material corporate news, it provides firms with some flexibility on how and when they communicate this information.[4] Therefore, firms may use this flexibility to mitigate the market impact of bad news with the announcement of other positive corporate events. In this paper, we study whether firms obfuscate the release of negative corporate news by releasing the information near another corporate event.

Data breaches are incidents that expose confidential or sensitive information to unauthorized parties and can impose large and economically meaningful direct (e.g., legal liabilities, fines, and operational disruptions) and indirect costs (e.g., reputational damage and loss of customers).[5]

[3] For example, firms may strategically cast earnings calls by calling on bullish analysts (Cohen, Lou, and Malloy, 2020), disclose bad news during times of low investor attention (deHaan, Shevlin, and Thornock, 2015), purposely obfuscate filings using circuitous language (Guest and Yan, 2025), and make fewer social media posts (Jung et al., 2018) to mitigate the effects of revealing bad news to market participants.

[4] Prior to the SEC's cybersecurity disclosure rule (adopted July 2023, effective December 2023 for most firms; see SEC Final Rule, Release No. 33-11216), breach disclosures were governed primarily by state laws. The SEC had only issued non-binding guidance in 2011 and 2018, which did not preempt state-level rules (SEC CF Disclosure Guidance: Topic 2, 2011; Commission Statement and Guidance on Cybersecurity Disclosures, 2018). As a result, firms during our sample period (2005–2022) largely followed state-level requirements—many of which were relatively lenient in terms of disclosure timing—without additional binding constraints from the SEC, giving managers considerable discretion over when and how to disclose breaches.

[5] Data breaches can occur through various channels, including data loss, theft, and the actions of malicious insiders or outsiders, including state espionage. These breaches often involve identity data, account access, and financial information (Stiennon, 2013). According to IBM, the average cost of a data breach in 2022 was $4.24 million, the highest in 17 years. In 2023, the Identity Theft Resource Center (ITRC) recorded an unprecedented 3,205 U.S. data

Despite the frequency and economic severity of data breaches, the academic evidence on the impact of data breaches on shareholders is mixed.[6] Consistent with the predictions of Bloomfield (2002), we predict that firms will attempt to obfuscate the impact of data breaches to shareholders by releasing information on the breach following other corporate events. Further, we predict that this misdirection is, at least, partially responsible for the mixed evidence on the value impact of data breaches found in the literature.

In this paper, we first utilize a comprehensive sample of data breaches in the United States to examine the preponderance, determinants, and outcomes of breach announcements occurring near confounding events.[7] Consistent with our predictions, we document that firms confound a significant proportion of data breach announcements with other firm news announcements. Specifically, we find that confounding events occur the day of or up to 7 days prior to a breach announcement in nearly half (47%) of breach announcements. Additionally, managers confound

breaches, a 78% surge from the previous year. Statista reports that a total of 641 million data records were leaked in data breach events between 2020 and 2023, affecting millions of firms and individuals.

[6] Some studies find that data breaches have a negative impact on stock returns (e.g., Cummins et al., 2023; Kamiya et al., 2021; Martin, Borah, and Palmatier, 2017; Rasoulian et al., 2023). However, other studies either do not find evidence of a significant impact on stock returns (e.g., Kannan, Rees, and Sridhar, 2007; Acquisti, Friedman, and Telang, 2006; Cavusoglu, Mishra, and Raghunathan, 2004) or document a short-term positive impact (e.g., Rosati et al., 2017; Wang et al., 2023).

[7] This paper focuses on U.S. public firms, where breach disclosure rules are set at the state—not federal—level, leading to significant variation in timing and flexibility. Many states allow delays during law enforcement investigations (e.g., Louisiana, Texas), while others, like Michigan, require disclosure "without unreasonable delay" but still leave room for managerial discretion. In contrast, the EU's GDPR enforces a strict 72-hour rule (Article 33), and China's Personal Information Protection Law (PIPL, Article 57) mandates prompt notification to regulators and affected individuals, though without a fixed deadline. These institutional differences highlight that our findings—particularly regarding timing discretion and potential confounders—are specific to the U.S. and may not generalize to more prescriptive regimes. (See Appendix Table A1 for selected U.S. state-level disclosure rules.)

the large majority (78%) of breach announcements using positive news events such as ratings upgrades, acquisitions of other firms, positive earnings announcements, and active investments. These rates align with other bad-news settings in which managers shape what accompanies the focal disclosure or when it is released. In product recalls, about 73% of announcements cluster over roughly one month, and the leader faces up to 67% larger stock-price penalties than followers (Mukherjee et al., 2022). In 8-K reporting, roughly 33% of filings are paired with a same-day separate press release, most commonly around negative items (Rawson, Twedt, and Watkins, 2023). In downsizing, 39.3% of events include at least one positive press release within the three-day window (Brauer and Vandepoele, 2024). Together, these patterns support treating confounding as a general disclosure device and motivate early identification of confounded breach cases in our analysis.

Despite the preponderance and potential importance of confounding news events near data breach announcements, the existing literature tends to remove confounded observations from their analysis (e.g., Kannan, Rees, and Sridhar, 2007; Gwebu, Wang, and Wang, 2018; Kamiya et al., 2021). Here, “confounding” refers to other firm-specific disclosures within the event window that can independently affect stock prices, such as earnings releases, guidance revisions, ratings changes, financing announcements, M&A transactions, or senior management turnover. For example, Campbell, Gordon, Loeb, and Zhou (2003) began with 84 candidate breaches and removed two that coincided with M&A transactions, which in our setting constitute clear confounding events. While most breach event studies apply comparable screening rules, few disclose the proportion of cases affected, making it difficult to assess the potential scale of selection effects. Although such exclusions sharpen identification by removing contemporaneous shocks, they can bias estimates when the incidence of confounding news is non-random[8] or when firms retain some control over whether confounding occurs. In our sample, overlap is concentrated at the breach announcement date and, to a lesser extent, in the pre-announcement window (Figure 1). Confounding types also differ in timing controllability: rating actions are externally timed; earnings announcements are pre-scheduled with limited flexibility; many other firm-initiated items afford greater discretionary. Taken together, these patterns are more consistent with targeted managerial timing than with chance. Under these conditions, omitting confounded events would therefore induce sample-selection endogeneity, offering one explanation for the mixed results reported in this literature.

By including confounded breach events in our analysis, we can determine whether managers can effectively obfuscate bad news with other corporate events and determine whether specific breach, manager, or firm characteristics increase the likelihood of a confounding strategy. For example, the degree of information asymmetry between managers and shareholders (Jensen and Meckling, 1976) can affect the quality and clarity of firm disclosures. Thus, we focus on several agency-based managerial characteristics (e.g., Jensen and Meckling, 1976; Shleifer and Vishny,

[8] Figure 1 plots the kernel density of distance = breach announcement date − nearest confounding firm-specific announcement within ±30 calendar days (Epanechnikov kernel; bandwidth 1.3849). The density shows a sharp mode at day 0 and elevated mass within ±3 days, evidencing clustering around the breach date. This preview is confirmed later: the timing analysis shows effects are strongest for same-day, weaker for anticipatory, and weakest for reactive confounds (Table 11).

1989; Malmendier and Tate, 2008) that predict the characteristics of managers who are more likely to confound a breach to avoid a fall in private benefits or compensation. Some of these characteristics include 'overconfident' managers (whose stock options are already 'in-the-money'), board and compensation characteristics (CEO tenure, board independence, CEO equity holdings), and firm reporting practices, including prior misstatements.

Our final sample includes 675 data breach incidents from 2005 to 2022 that meet our data requirements, and we also create control samples of 675 non-breached firms based on the Hoberg and Phillips (2016) pairwise similarity index and propensity-score matching (PSM). Using this comprehensive sample, we show that more severe data breaches (breaches in the upper-quartile of severity) generate negative and significant cumulative abnormal returns (CARs) and are associated with economically large dollar losses.

Regressions controlling for firm characteristics and year, industry, and state fixed effects confirm that more severe breaches generate significantly lower CARs of between -0.396% and -0.447% based on 3- and 5-day windows, respectively. Further, these results are not driven by breach announcements occurring on busier news days (as documented by Foerderer and Schuetz, 2022), as we explicitly control for the amount of news on a breach day in all of our regression analyses. The results are also economically large and suggest that a one standard deviation increase in breach severity results in a 0.858% (-0.443% × 1.937) decrease in 5-day CARs, holding other regression factors constant. Economically, the fall in market value is equivalent to a $365 million reduction in firm value for the average breached firm (0.858% × $42.5 billion). The findings are similar to other studies that report negative CARs for specific breach types (e.g., financial) including Kamiya et al. (2021) or based on studies using a similar breach severity score as in Cummins et al. (2023).

We next examine the determinants of confounded breaches (representing 47% of the sample), which we define as breaches that involves additional firm news announcements (e.g., earnings, investments, ratings) on the day of the breach and 7-days prior to the breach.[9] Consistent with our predictions, we find that more severe breaches are significantly more likely to be confounded. Further, certain managerial characteristics increase the likelihood of firms confounding breach announcements. Specifically, we find that managers who are overconfident, and CEOs with greater equity holdings and longer tenure are more likely to confound breach announcements. In addition, we find that firms who have previously experienced a financial misstatement are also more likely to confound breach announcements. These findings suggest that firms with potentially acute agency issues as documented by the empirical literature significantly predict potential confounders. The impact of confounding is also economically large. For example, an increase of one standard deviation in our managerial overconfidence measure increases the likelihood of confounding by 107% ($\exp(0.477 \times 1.525) - 1$), holding other factors constant. In addition, we find that overconfident managers and managers with longer tenure tend to use positive news events to offset the negative effects of breach announcements. This behavior can be seen as a strategic move where

[9] We also obtain confounded breaches based on other confounded periods, such as 30-days prior to the breach and 60-days around the breach announcement date (-30, +30). The results are reported in the supplementary on-line Appendix B (Tables A11, A12, A13, A15, and A16).

management attempts to paint their firm in the most positive light possible, a concept that is central to the impression management theory (Goffman, 2016). A one standard deviation increase in managerial overconfidence and CEO tenure boosts their likelihood of using positive confounding events by 203% ($\exp(0.477 \times 2.328) - 1$) and 278% ($\exp(7.041 \times 0.189) - 1$), respectively. However, managers also adopt a 'big bath' strategy (Fedyk and Khimich, 2018), releasing more negative news, such as ratings downgrades, business losses, negative earnings announcements, corporate restructuring, and disinvestment.[10] Such concurrent announcements may attenuate the market impact of each individual item, either because investor attention is dispersed across multiple negative signals or because investors anticipate that firms tend to bundle adverse announcements, thereby muting the incremental market reaction to subsequent disclosures.

Finally, we examine whether confounding strategies are successful. Specifically, we test whether they reduce the adverse impact of the breach on firm value in the short- and long-term. First, we find that managers who confound breach events tend to ameliorate the negative impact of the breach announcement on firm value. For example, firms with a confounding event near their breach announcement earn significantly higher 5-day CARs (1.798%). This positive impact tends to ameliorate the significantly lower abnormal returns for confounding firms around severe breaches (-0.544%). To assess longer-horizon performance, we follow Fama (1998) and Mitchell and Stafford (2000) and employ a calendar-time portfolio (CTP) strategy, which avoids the

[10] Despite the limited negative and neutral events in the sample, we show a significant negative correlation for negative news and CEO tenure suggesting that negative events are primarily released by newer CEOs or those with shorter tenures. Long-tenured CEOs tend to avoid negative or neutral news, favoring positive confounding events instead.

empirical issues of long-run buy and hold abnormal return strategies following significant corporate events documented by Bessembinder and Zhang (2013) and Bessembinder, Cooper, and Zhang (2019). Markets are unlikely to be instantaneously informationally perfect, and managers may anticipate that investors do not fully process all signals in real time. Under these conditions, confounding, by mixing the breach with other firm news and diverting investor attention, can temporarily attenuate event-time price reactions. As information is disentangled and incorporated, such attenuation should dissipate in calendar time. We find that the positive short-run response for high-severity confounded breaches fades over subsequent months. Over the 12 months following the breach announcement, risk-adjusted alphas from our calendar-time portfolios are generally indistinguishable from zero, with one notable exception: a modest, marginally significant negative alpha for the equal-weighted portfolio restricted to the upper-quartile severity subsample. Except for this narrow case, confounding firms do not systematically underperform unconfounded breach firms over the subsequent year. Taken together, the evidence for confounding firms is consistent with a temporary, attention-driven underreaction that dissipates, rather than a broad long-run reversal of the short-run positive effect or a persistent, systematic underperformance.

Our study adds to the existing literature in several ways. First, our findings fill a gap in both data breach and strategic timing literatures by investigating the effects of firm-related confounding events. We demonstrate a novel contribution of our in-depth study of ‘confounding events’, which differs markedly from the existing data breach studies that exclude confounded breaches (e.g., Kannan, Rees, and Sridhar, 2007; Gwebu, Wang, and Wang, 2018; Kamiya et al., 2021). We show that ignoring these events could lead to an incomplete understanding of the phenomenon of data breaches, so we make this the key contribution of this paper. Further, while some studies consider strategic timing, they diverge from our focus on firm-related confounding events. For example, some studies examine timing strategies influenced by other factors such as news pressures (e.g., Foerderer and Schuetz, 2022), the interval between data breach occurrence and announcement, the content of the announcement, including customer remedies or firm-specific responses, and major unrelated events like terrorist attacks or COVID-19 that can shift investor attention (e.g., Kannan, Rees, and Sridhar, 2007; Durante and Zhuravskaya, 2018).

Second, our study provides a fresh explanation that links attenuated market reactions to managers' confounding strategies. While conflicting evidence regarding market reactions to data breaches may be attributed to several other factors identified in previous studies (Kannan, Rees, and Sridhar, 2007; Martin, Borah, and Palmatier, 2017), such as event window size, type of breach, sample size, and the type of event study methodology used, it appears that the presence of a confound can have a large impact on returns and firm market value.

Third, our study contributes to the prediction of whether firms strategically time their data breach announcements. While several studies research the incidence and impacts of data breaches (e.g., Eling and Loperfido, 2017; Sun, Xu, and Zhao, 2021), we study the determinants of confounding these breaches. We offer evidence that firms' propensity to confound varies based on breach severity and information asymmetry. Specifically, we provide a more comprehensive analysis by covering multiple key factors, including breach severity and firm and managerial behavioral characteristics. We show that when the breach is severe and when managers have certain characteristics related to agency issues, including overconfidence, longer tenure, and larger equity holdings, severe breaches are more likely to be confounded.

Fourth, our findings contribute novel insights to the literature on firm behavior during and after data breaches. Existing research primarily focuses on firms' recovery behaviors, such as public apologies and consumer compensation campaigns (Gwebu, Wang, and Wang, 2018). Our research introduces 'confounding' as a new managerial response to data breaches, influencing announcement outcomes.

The rest of this paper will be presented as follows: Section 2 examines the background literature and sets out some testable hypotheses. Section 3 reports on sample construction, variable definitions, and methodology. Section 4 discusses the empirical results, while Section 5 conducts an analysis of endogeneity and additional robustness tests. The conclusions and discussion are presented in Section 6.

2 Hypotheses Development

2.1 Breach severity and confounding

A large body of research has examined the impact of data breach announcements on firm abnormal stock returns. In a meta-analysis of 64 studies, Ebrahimi and Eshghi (2022) find that most studies show a significant negative relationship exists between data breach announcements and stock returns, on average. The reaction or signal from a breach varies based on contributing factors (Yayla and Hu, 2011), including whether information is confidential or financial (Yayla and Hu, 2011; Campbell et al., 2003; Acquisti, Friedman, and Telang, 2006), data volume (Garg, Curtis, and Halper, 2003), and whether it is a hacker intrusion (Rasoulian et al., 2023).

Breach severity is the degree of loss or threat that the data breach events cause to the affected firms, which can usually be measured by indicators such as the data type, amount, source, sensitivity, harm, and value (Aivazpour, Valecha, and Chakraborty, 2018). Breach severity not only affects the direct costs of the affected firms (e.g., legal fees, compensation fees, repair fees), but also affects their indirect costs, such as reputation loss, customer loss, and competitiveness decline (Martin, Borah, and Pakmatier, 2017).

Managers may have incentives to obscure and blur real information when the firm's performance is poor (Bloomfield, 2002; Li, 2008). Firms with data breaches experience subsequent negative performance due to reputation damage (Gatzlaff and McCullough, 2010; Akey et al., 2024), decreased customer spending (Janakiraman, Lim, and Rishika, 2018), revenue loss (D'Arcy et al., 2020), lower profits (Campbell et al., 2003), reduced firm productivity (Makridis and Dean, 2018), and even a loss of business (Lending, Minnick, and Schorno, 2018). Managers may have incentives to avoid these outcomes. Given that information with higher processing costs might not be fully and quickly incorporated into market prices (Grossman and Stiglitz, 1980), management may strategically present the most positive version of their firm and may use its information advantage to release vague or misleading information to conceal the true nature of the data breaches (Goffman, 2016). Further, since management has legal flexibility on the exact timing that data breaches are publicly disclosed, they can utilize pre-planned or flexible firm-specific confounding events to obfuscate data breach announcements. Specifically, we predict managers will announce data breaches near events that may offset or take focus away from the negative impacts of a data breach, leveraging lower investor attention to reduce potential market penalties (DellaVigna and Pollet, 2009; Michaely, Rubin, and Vedrashko, 2014; deHaan, Shevlin, and Thornock, 2015). These can include information on ratings upgrades, proactive acquisitions of other firms, positive increased earnings announcements, and active investments. Therefore, we predict:

H_1: Firms are more likely to announce data breaches near confounding events when breach severity is high.

2.2 Managerial information asymmetry and confounding

Beyer et al. (2010) argue that the corporate information environment is a function of the dynamic interactions resulting from information asymmetries and agency problems between investors, firms, and managers. Therefore, a firm's corporate information environment is shaped by manager reporting and disclosure, mandated reporting and disclosure regulations, and analysts' expectations.

As managers risk job loss as a result of poor stock and earnings performance (Palepu, 1986; Pound, 1988; Warner et al., 1988; Weisbach, 1988; Morck et al., 1990), Healy and Palepu (2001) predict that managers use corporate disclosures to reduce the likelihood of undervaluation and to explain away poor earnings performance. However, the degree of information asymmetry and the potential for moral hazard (Jensen and Meckling, 1976) will likely affect the quality and transparency of information disclosure strategies. Management may opt to sacrifice the firm's long-term interests for the sake of job security and stability, adopting confounding strategies to smooth the company's negative performance. Further, firms with poor corporate governance may exacerbate these incentives through factors such as overconfidence and insensitivity to potential fallout (Doukas and Petmezas, 2007; Schrand and Zechman, 2012; Ahmed and Duellman, 2013), CEO equity tied to wealth loss (Janakiraman, Lim, and Rishika, 2018; Benischke, Martin, and Glaser, 2019), and prolonged CEO tenure, which may foster board conflicts (McClelland, Liang, and Barker, 2010; Wen, Rwegasira, and Bilderbeek, 2002; Evans III, Nagarajan, and Schloetzer, 2010; Kaczmarek, Kimino, and Pye, 2012; Darouichi et al., 2021). Weak governance signaled by misstatements also erodes reputation (Palmrose, Richardson, and Scholz, 2004; Agrawal and Chadha, 2005) and managerial entrenchment further leads to value-decreasing decisions (Harford, Humphery-Jenner, and Powell, 2012; Masulis, Wang, and Xie, 2007). Thus, we examine the following hypothesis:

H_2: Firms with greater managerial information asymmetry and poorer governance will be more likely to announce data breaches near confounding events.

2.5 Confounding and economic outcomes

Bloomfield (2002) posits that data that are more costly to extract from public disclosures are less fully reflected in market prices. Therefore, confounding may lead to a weakened negative market response to data breaches. Further, managers will be most incentivized to confound when data breaches are more severe.

However, existing empirical studies present mixed evidence of the effect of breach severity on the stock returns of listed firms. Several studies find that data breach events have a negative impact on stock returns (Campbell et al., 2003; Garg, Curtis, and Halper, 2003; Acquisti, Friedman, and Telang, 2006; Yayla and Hu, 2011; Martin, Borah, and Palmatier, 2017; Kamiya et al., 2021; Cummins et al., 2023; Rasoulian et al., 2023). Based on a systematic review of event studies from 1998 to 2015 studying the impact of various IT security breaches on stock market valuations, Spanos and Angelis (2016) report that only 71% of studies confirm a significant negative impact. Many studies document null effects or limited evidence (e.g., Hovav and D'Arcy, 2003; Kannan, Rees, and Sridhar, 2007; Acquisti, Friedman, and Telang, 2006; Goldstein, Chernobai, and Benaroch, 2011; Hilary, Segal, and Zhang, 2016). Meanwhile, some studies even report a short-term positive impact (Wang et al., 2023; Rosati et al., 2017). Additionally, investments in corporate social responsibility tend to increase in the years following a breach (Akey et al., 2024; Lending, Minnick, and Schorno, 2018).

If managers are effective in mitigating the negative impact of data breaches by announcing the breach near another corporate event, part of this mixed evidence may be due to the obfuscation for several reasons. First, the manager may successfully offset the negative impact of the data breach announcement and obfuscate the empirical impact on returns. Second, empirical research often excludes data breach observations where confounding effects occur (e.g., Kamiya et al., 2021;

Gwebu, Wang, and Wang, 2018; Rosati et al., 2017). In our sample, confounding becomes more common as severity increases—47.7% overall (322/675), 52.2% among above-median cases (169/324), and 67.7% in the upper quartile (111/164)—suggesting a potential positive association between severity and the likelihood of confounding. This severity–confounding link may prevent empirical researchers from observing a significant negative effect, even if the confounding effort is not successful.

Finally, data breaches may create a lemons problem (Akerlof, 1970). For example, consider a situation where half of the firm's data breaches are 'confounded' (i.e., 'lemons') and the other half are 'unconfounded'. Both investors and managers are rational, and value investments based on their own information. While the presence of a confounding announcement is observable, whether it reflects strategic timing—and the breach's true severity—are not. Given this ambiguity, rational investors cannot perfectly separate types and will price breaches toward an average, which incentivizes managers to time confounding news to mimic lower-severity events. Consequently, some unconfounded breaches are undervalued and some confounded breaches are overvalued relative to the information available to managers. Therefore, based on the discussion above, we propose the following hypotheses:

H_3: Breach severity has a significant negative impact on firm stock returns.

H_4: Firms confounding a breach announcement will alleviate the short-term stock market decline from data breaches.

3 Sample, Variables and Methodology

3.1 Sample selection and variable definitions

The data breach samples are extracted from the Privacy Rights Clearinghouse (PRC) Database. Due to the data collection limitations of this database, the sample period ranges from 2005 to 2022. We find the highest pairwise similarity scores from text analysis of firm 10-K product descriptions for each breached firm in the Hoberg and Phillips Data Library and use them as comparable matches for breached firms.[11]

We obtain the end-of-day prices of all listed NYSE, Amex, and NASDAQ common stocks along with basic market indices from the CRSP U.S. Stock Database and select the listed firms for our study. We set the data breach announcement (reported) date as the event date (t=0) and employ two primary event windows around this date: 3-day (-1,+1) and 5-day (-2,+2). With an estimation window of (-230, -31), we calculate cumulative abnormal returns (CARs) for both the market-adjusted model (MAM) and the Fama-French 5-factor model (FF5), based on stock returns and the CRSP U.S. value-weighted market index, and five-factor data from Kenneth French’s Data Library. We test the robustness of our results using the CARs derived from FF5. The financial variables are extracted from the merged data of CRSP and Compustat databases.

The data provided by the PRC does not contain common firm identifiers. Therefore, we manually rectify the different display modes of the same firm name and match the firm’s financial

[11] We also examine propensity score matching (PSM) to select comparable firms as matches for our breached sample. See the supplementary on-line Appendix B (Table A6 and A10).

data with the fiscal year of the data breach announcement date. This process results in a sample of 1,351 breached firms. We then eliminate duplicate samples where a single breach event is announced multiple times and represent these duplicate observations with a dummy variable, *MultiAnn*. After calculating all the variables and screening out firms with missing data, we have 675 data breaches and match these with 675 closest-comparable (using Hoberg and Phillips pair-wise similarity score) non-breached firms from 2005 to 2022.

Table 1 presents the composition of both breached and non-breached samples, categorized by industry and year. It encompasses 10 major categories and 58 distinct industries, as per the SIC level 2 classification. The most prevalent major category is Finance, Insurance, Real Estate (#60-#67), accounting for 243 breaches and 238 non-breaches. Conversely, the least common category is Agriculture, Forestry, Fishing (#01-#09), with zero breaches and one non-breached observation.

The year 2019 witnessed the highest frequency of data breaches, with a total of 220, while 2005 recorded the lowest, with only 3 breaches.

<Table 1 here>

To create our confounding dummy we use the event announcement information of listed firms from the LexisNexis. Following Rosati et al. (2017), we identify the following announcements as confounding events: earnings announcements, merger and acquisition announcements, capital increases, investment or disinvestment announcements, public offerings, rating actions, and restructuring announcements. We construct an indicator variable, *7-Day Confounding,* to capture whether there are any confounding events near the breach announcement. *7-Day Confounding* is equal to 1 if there are any confounding events within 7 days on and before the data breach announcement date, and 0 otherwise. In robustness tests, we also define two alternative indicators, *30* (*60*)-*Day Confounding*, which are equal to 1 if there are any confounding events within 30 (60) days on and before the data breach announcement date, and 0 otherwise.

We use the Breach Level Index (BLI), a simple tool jointly developed by IT-Harvest and SafeNet (see Stiennon, 2013), to calculate the breach severity based on publicly disclosed information. The formula for calculating BLI is as follows:

$$BLI = \log_{10}(N \times t \times s \times A) \tag{1}$$

Where *N* represents the number of records breached, such as personal identity information, credit card numbers, and email addresses. The greater the number of records, the greater the impact of the breach. *t* represents the data type. This is the sensitivity and value of the data that has been breached, such as medical records, financial information, and passwords. The more sensitive the data type, the more severe the damage caused by the breach. *s* denotes the source, which refers to the kind of agent that perpetrated the breach, ranging from hackers and insiders to government agencies and others. The risk of the breach escalates with the malice of the source. *A* signifies the impact, which measures the extent to which the breached data has been exploited for nefarious ends, such as fraud, identity theft, or extortion. The consequences of the breach worsen as the impact intensifies.

The BLI is a logarithmic index (base 10) with no upper limit, but the highest score for a breach event in our sample is less than 10. We utilize two methods for accounting for missing values of *N* in the data.[12] First, we replace the missing values of records affected with a value of one, following a methodology similar to Tukey's started-log transformation (e.g., Tukey, 1977, p. 552). Second, we replace the missing values of records affected with the average number of records for that type of breach. From these assumptions, we obtain two different breach severity values, denoted as *BS1* and *BS2.*[13] As documented in Table 2, *BS1* has an average of 2.745, median of 2.204, 75th percentile of 4.000, a standard deviation of 1.937, and displays a right-skewed distribution, suggesting that while most events are less severe, a few outliers cause significant impact. Conversely, *BS2* has an average of 3.158, median of 2.992, 75th percentile of 4.030, and standard deviation of 1.694, exhibits a nearly symmetrical distribution, indicating a majority of events cluster around the average severity level. We utilize *BS1* as our primary measure but demonstrate the robustness of our results using *BS2*.[14]

<Table 2 here>

Building upon the foundation of agency theory (Jensen and Meckling, 1976), we utilize eight managerial variables from the existing literature that are highly correlated with agency costs and

[12] Missing values in PRC's data breach records can be attributed to (a) Data Collection Challenges: Certain breach details may be undisclosed or indeterminable; (b) Data Sensitivity: To safeguard privacy, sensitive data (e.g., Social Security numbers, financial account numbers, driver's license numbers) might be omitted; (c) Reporting Inconsistency: Variations in reporting standards across firms can result in inconsistencies, thereby impacting data completeness.

[13] Since we are using a logarithmic function, defining the breach severity value as 0 reduces the sample size due to the logarithmic function being undefined at 0.

[14] Results calculated based on BS2 can be found in the supplementary on-line Appendix B (Table A5). The sample size for BS2 is consistent with BS1, and most of its results support the conclusions obtained using BS1.

information asymmetry. Further, while equity incentives may be designed to resolve the principal-agent problem, they may not effectively address, and may even exacerbate agency costs (Healy, 1985; Gaver, Gaver, and Austin, 1995) and incentivize managers to manipulate their performance to increase pay.

Therefore, we include the following variables from the existing literature: first, we include the proportion of shares held by the CEO (*CEOEq*, calculated by dividing the number of shares held by the CEO by the total number of shares of the firm) (e.g., Bryan, Hwang, and Lilien , 2000); second, the proportion of CEO's equity incentives (*CEOEqComp*, calculated by dividing the CEO's equity incentive income by total income) (e.g., Song and Wan, 2019); third, the tenure of the CEO (*CEOTen*, calculated by subtracting the year of the CEO's first appointment from the year of the current appointment) (e.g., Casamatta and Guembel, 2010; Faulkender and Yang, 2010; Taylor, 2010; Luo, Kanuri, and Andrews, 2014; Souder, Simsek, and Johnson, 2012). Fourth, we include the independence of the firm's board of directors (*BoardInd*, calculated by the proportion of non-executive directors to the total number of directors) (e.g., Coles, Daniel, and Naveen, 2008; Knyazeva, Knyazeva, and Masulis, 2013). Fifth, we study the level of earnings management of the firm (*EarnMan*, which estimates discretionary current accruals lagged by total assets) (e.g., Kothari, Leone, and Wasley, 2005). Sixth, we study the degree of management entrenchment of the firm (*ManEnt*, which is an antitakeover protection index that attaches to each state a score from 0 to 5 that is equal to the number of 5 key antitakeover statutes that it has) (e.g., Bebchuk, Cohen, and Ferrell, 2009; Karpoff and Wittry, 2018).[15] Seventh, we study whether the firm has committed financial fraud in the past (*PriorMis*, represented by 1 or 0; one if a firm-year's financial statement was reported in Accounting and Auditing Enforcement Releases (AAER) before the breach announcement, and zero otherwise) (e.g., Lo, Ramos, and Rogo, 2017). Finally, we include whether the CEO is overconfident (*OverCon*, represented by 1 or 0; one if a CEO holds vested options that are at least 67% in the money once, and zero otherwise) (Malmendier and Tate, 2008; Hirshleifer, Low, and Teoh, 2012).

We include a number of firm- and industry-level controls common to the literature to control for factors that may affect the impact of data breaches on stock returns. Variables at the firm level include the size of the firm (*Size*, calculated by the logarithm of the firm's total assets), investment opportunities (*TobinQ*, calculated by the ratio of the firm's market value to book value of assets), financial leverage (*Leverage*, calculated by the ratio of the firm's total liabilities to total assets), profitability (*ROA*, calculated by the ratio of the firm's operating income before depreciation to total assets) (e.g., Jewell and Mankin, 2011), liquidity (*Cash*, calculated by the ratio of the firm's cash and cash equivalents to total assets), and research and development investment (*R&D*, calculated by the ratio of the firm's research and development expenses to total assets). To reduce the impact of extreme values or outliers, we also winsorize continuous variables at the 1% level.

[15] We use state-based antitakeover provisions rather than firm-level provisions to address possible endogeneity of firm-level provisions (see, e.g., Karpoff and Wittry 2018; Guernsey, Sepe, and Serfling, 2022).

At the industry level, variables include whether the firm belongs to the high-tech industry (*HighTech*, represented by 1 or 0), the 10-K products text-based firm-specific industry concentration calculated using firm sales data (*HHI*, 10-K TNIC industry concentration data from the Hoberg-Phillips Data Library), the fluidity of the firm's products (*Fluid*, 10-K based product market fluidity data from the Hoberg-Phillips Data Library), and the similarity between the firm's products and those of its competitors (*Similar*, 10-K TNIC total similarity data from the Hoberg-Phillips Data Library). *HHI*, *Fluid,* and *Similar* capture product market competition (PMC) and are included as they might determine how investors react to a data breach. For example, if the breach occurs for a highly concentrated firm with large economic rents investors might price in a loss of these rents if customer switching occurs and similar comparable firms gain. Finally, following Foerderer and Schuetz (2022), we control for news pressure (*NewsPr*, defined as ln (1 + news pressure), where news pressure is the daily count of WSJ articles under 'Financial and Commodities Markets News' on the breach day) throughout our analysis. All the variable descriptions can be found in Appendix A.

3.2 Empirical models

To examine whether breach severity exacerbates the possible negative effects of data breaches on firm returns (CARs), and whether breach severity is one of the main motives for managers to confound to reduce the damage to firm returns, we estimate the following model:

$$\begin{aligned} CAR_i = a_0 + a_1 BS1_i + a_2 Firm\ controls_i + a_3 Industry\ controls_i + a_4 NewsPr_t \\ + \delta_{Year} + \delta_{Industry} + \delta_{State} + \varepsilon_i \end{aligned} \tag{2}$$

Where *BS1* is one of the indicators of breach severity based on replacing missing records with a value of 1. Firm and industry-level control variables are measured at the fiscal year before the data breach occurred. *NewsPr* is measured as the natural logarithm of (1 + news pressure), with news pressure measured by the daily count of WSJ and WSJ Online articles under 'Financial and Commodities Markets News' on the breach announcement day, sourced from Factiva. δ denotes fixed effects, with Year, Industry (SIC level 2), and State FEs accounting for year, industry, and state-level factors where the breach occurred. State-level fixed effects capture any differences in data breach announcement-related regulations in the state where the data breach occurred (Foerderer and Schuetz, 2022).[16] We estimate the above model for 11 different windows of CARs, and since the same firm may have multiple data breach observations in different years, we use firm-level robust clustered standard errors.

When firms face a severe breach, some may delay disclosure or attempt to confound. We use the following logistic model with the aim of showing if certain managerial characteristics help predict firms that are likely to confound. We also aim to identify the types of confounding they exhibit, while controlling for other characteristics:

[16] Each state tailors its data breach laws to better protect its citizens. These laws define what is considered personally identifiable information. For example, the definition of a breach, the parties that must be notified in the event of a breach, and various exemptions often vary from state to state when assessing the damage caused by a data breach. Based on the information published on the official websites of each state's Attorney General, we compile the existing data breach laws by state in Table A1 of the supplementary on-line Appendix B.

$$
\begin{aligned}
Confound_i = \beta_0 + \beta_1 BS1_i + \beta_2 Managerial\ characteristics_i \\
+ \beta_3 Firm\ controls_i + \beta_4 Industry\ controls_i + \beta_5 NewsPr_t \\
+ \delta_{Year} + \delta_{Industry} + \delta_{State} + \varepsilon_i
\end{aligned}
\tag{3}
$$

The model variables follow eq. (2), with firm-level robust clustered standard errors. Consistent with earlier evidence, overlap between breaches and other firm-specific confounding announcements clusters around the breach announcement date (Figure 1) rather than occurring at random. Because such overlaps co-varies with returns and key covariates, excluding confounded cases conditions on managerial timing and can induce sample-selection endogeneity. Event-time returns also differ systematically depending on how these overlapping announcements are treated in sample construction. For example, estimates based solely on unconfounded breaches can diverge form those based on the unconfounded subset within a mixed sample, highlighting that sample composition can materially influence estimated breach effects. In our data, the *7-Day Confounding* indicator averages 0.477 (Table 3), so excluding these cases would remove nearly half the sample and risk skewing the estimates. We therefore retain them in the baseline analysis and report pooled and stratified estimates in parallel. As a robustness check, we use eq. (3) as the selection equation of the Heckman two-step method to predict whether the firm will confound the data breach event, and then use the inverse-mills ratio (*IMR*) obtained from this model as an additional explanatory variable in the second-step CARs regression models to control for the impact of sample selection bias.[17]

To verify the potential motives and consequences of confounding events, we compare the average stock returns of confounded breaches and unconfounded breaches, and we estimate eq. (4) for the breached sample:

$$\begin{aligned} CAR_i = \varphi_0 + \varphi_1 BS1_i \times Confound_i + \varphi_2 BS1_I + \varphi_3 Confound_i \\ + \varphi_4 Firm\ controls_i + \varphi_5 Industry\ controls_i + \varphi_6 NewsPr_t \\ + \delta_{Year} + \delta_{Industry} + \delta_{State} + \varepsilon_i \end{aligned} \tag{4}$$

These models have important theoretical and practical implications for understanding the economic impact of data breaches and the role of managerial information asymmetry, as well as how to better manage and disclose data breach events.

4 Results

4.1 Descriptive statistics

Table 3 shows the descriptive statistics of the main and the control variables for the breached and non-breached firms. Significant differences exist between both samples on several variables,

[17] In our first-step model, we discuss and test three potential instrumental variables (IVs) to determine if they meet the relevance criterion and exclusion restriction (see Section 5). Specifically, these IVs should be related to predicting a confounding event but are unrelated to short-term CARs in the second-stage regression.

reflecting the firms' stock returns, risk exposure, market value, innovation capabilities, decision-making style, and incentive mechanisms.

Specifically, we calculate the CARs of the breached firms and their closest competitors in different event windows to test if breaches earn negative CARs. Table 3 displays the CARs of the breached firms and their closest competitors under the MAM model. The 5-day CAR (5) of the breached firms is significantly negative and lower than that of their closest competitors, indicating the immediate negative impact of data breaches on firm returns. The 3-day CAR (3) of the closest competitors is positive, although not significant, suggesting that some competitors may benefit from customer migration from the breached firms in the short run.

The breached firms have a significantly larger average firm size (8.827) than the non-breached firms (8.195), which may suggest that they possess more sensitive data information and are more susceptible to data breach risks (Kamiya et al., 2021). The breached firms have a significantly higher average Tobin's q (2.241) than the non-breached firms (1.985), indicating that they have a higher market value relative to assets and growth options, possibly because the market perceives them as having stronger innovation capabilities and core competencies. They also have significantly higher cash holdings (0.098) relative to non-breached (0.027) and are more likely to be High-Tech with larger R&D expenditures than non-breached. Breached firms, with a significantly longer CEO tenure (8.892) than non-breached firms (7.656), validate the potential for greater confidence, as corroborated by our OverCon measure. The noticeably larger CEO equity holdings and CEO equity-based compensation in breached firms suggest their CEOs bear a larger risk exposure. This increased stake prompts riskier decisions, thus making these firms more susceptible to data breaches.

<Table 3 here>

4.2 Correlation analysis: managerial characteristics and breaches

According to the correlation analysis (see Table A3 in the supplementary on-line Appendix B), the longer the CEO's tenure, the more likely they are to be overconfident, with a correlation coefficient of 0.206. In addition, CEOs with longer tenures often have a record of data breaches, and there is a positive correlation between overconfidence and past data breaches, with a significant correlation coefficient of 0.289. This suggests that overconfidence is more common among CEOs of companies with a history of data breaches. At the same time, overconfident managers are more likely to issue multiple announcements for a single data breach event, which may reflect that information asymmetry among managers is a driving factor for managerial confusion. These results provide a deeper understanding of the relationship between CEO behavior and company data breach events, which helps us better understand the managerial behavioral links and data breach events.

4.3 Empirical regression results

4.3.1 The short-term impact of data breach events on CARs

We first examine the univariate impact of data breach severity on firm value. We divide the sample into three groups based on the number of records affected: high, medium, and low. Table 4 (Panel A and B) show the CARs of the high severity group (above median) and the highest severity group (upper quartile). We find that the CARs of the breached firms in the high and highest severity groups are significantly lower than those of their closest non-breached rivals, and the difference decreases as the event window lengthens from CAR (5). Interestingly, when breach severity increases investors appear to respond better to comparable firms as the CARs are less negative and are not significant. This suggests that data breach severity exacerbates the negative impact of breaches on firm returns, and that this impact can be attenuated. On the other hand, we also find that the differences in short-run CARs of breached and non-breached firms are significantly negative, indicating that they may benefit from customer loss from the breached firms. This result is consistent with our previous analysis that data breaches lead to market share redistribution, reducing the competitiveness of the breached firms and increasing that of their rivals.

<Table 4 here>

4.3.2 Breach severity and firm returns

This section examines the impact of breach severity on the short-term market returns using OLS regressions. Table 5 Panel A shows that the regression coefficient of breach severity on CARs is negative and highly significant for both the 3-day and 5-day CARs, consistent with H_3. Our results are also consistent with previous studies (e.g., Martin, Borah, and Palmatier, 2017), indicating that breach severity is one of the important factors affecting the market returns of breached firms. We do not find evidence that any other control variable significantly impacts returns using these windows.

In Panel B, we include several alternative windows, including those with longer horizons. Specifically, we continue to find a significant negative effect for 7, 9, 11, and 13-day CARs. However, as the time window expands, the regression coefficient for breach severity becomes less significant. This may suggest that any potential confounding measures taken by management have had some effect and help to attenuate breach severity.

<Table 5 here>

4.3.3 Breach severity, managerial information asymmetry, and confounding

In this section, we test our first two hypotheses, namely, whether firms confound breach announcements by announcing them near firm events (H1) and whether this behavior is more likely when asymmetric information and agency issues are more severe (H2). We use a logit model to first regress the confounding dummy on breach severity and our standard control variables. Next, we study whether confounding is more likely using various managerial information asymmetry variables, such as overconfidence, prior earnings management, CEO equity, CEO tenure, CEO equity-based compensation, board independence, managerial entrenchment, and prior financial misstatements. These variables reflect potential manager agency issues and the information asymmetry environment of the firm. This could impact the quality and timeliness of information disclosures, and potentially explain the occurrence and impact of confounding (Foerderer and Schuetz, 2022).

The results are shown in Table 6 Panel A. We include each variable separately in the logit regression in columns (1) through (10), and then combine them in column (11). Column (1) shows that more severe breaches are significantly more likely to be confounded, confirming H_1. Economically, a one standard deviation increase in breach severity (1.937) is associated with a 35% increase in the likelihood of the firm confounding the breach announcement ($0.35 = \exp(1.937 \times 0.153) - 1$) from column (1). We also find evidence in support of H_2 in columns (2) through (11). Specifically, we find that firms are more likely to confound when the CEO is more overconfident, when the CEO has longer tenure, when the board of directors is less independent, or when firms have prior misstatements. These results hold after controlling for within year, industry, and state fixed effects. We find that larger, higher *Q,* high-tech firms in industries with less competition tend to confound breach announcements more. However, even when controlling for these effects we find support for our first two hypotheses. The results are also not driven by other firm characteristics or events that are announced on days with high news pressure. These effects are also economically significant, as a one standard deviation increase in CEO overconfidence (0.477) more than doubles the likelihood of confounding a breach announcement with a firm event ($1.07 = \exp(0.477 \times 1.525) - 1$) from column (2). Taken together, firms experiencing higher breach severity or with worse governance or asymmetric information problems are significantly more likely to confound a breach event by announcing it near another firm event.

Firms may confound breach announcements with positive or negative events. As seen in Figure 2, rating actions and earnings announcements are the most prevalent breach confounders, accounting for 51% and 29% of confounding events, respectively. Within these two categories, positive events are particularly prevalent, constituting 70% and 89% of their respective totals. In Panel B of Table 6, we next turn to study the subset of breaches confounded by only positive events. As can be seen in Panel B, we continue to find consistent results supporting H_1 and H_2. These results are intuitive - more severe breach events or more poorly governed managers are more likely to confound a negative event such as a breach with a positive event to obfuscate the breach's negative impact.

<Table 6 and Figure 2 here>

4.3.4 Breach severity, confounding events, and economic outcomes

The results of Table 6 show that breach severity may be one of the motives for firms to engage in confounding events, because more severe data breaches lead to greater negative impacts, and firms may try to divert investors' attention or mitigate losses by releasing other information. We use two methods to test the impact of confounding on the short-term returns: one is to compare the CARs between confounded and unconfounded breaches, and the other is to use a multivariate regression model, controlling for other variables, to examine the impact of the interaction term of the breach severity with confounding on the firm's stock returns.

Table 7 reports the CARs of breached firms, split by whether they have confounding events within 7 days before the breach announcement. Compared to Panel A, Panels B and C, representing higher breach severity (above-median and upper-quartile), exhibit more significant differences in CARs. This implies an interaction between breach severity and confounding strategy, with confounding significantly affecting the firm's stock price only when breach severity is high. For instance, in Panel B, the differences between confounded and unconfounded abnormal returns for severe breaches are significant for both CAR (3) and CAR (5) at -2.258% and -2.620%, respectively. We observe similar effects with upper-quartile breaches in Panel C.

To ensure these effects are not driven by industry, year, state, or other firm or competitive factors, we also analyze abnormal returns in a multinomial regression setting in Panel D. The significantly positive regression effect on Confound coupled with the negative regression coefficients of interaction terms of confounding with breach severity demonstrates that the confounding event can mitigate the loss due to the breach. As can be seen in column (1), the positive Confound coefficient (1.389%) coupled with a one standard deviation increase in breach severity with the interacted coefficient (–1.149% = –0.593% × 1.937) yields a net effect of 0.24%. This confirms that managers can effectively mitigate breach severity and mislead investors through confounding activities. This provides evidence consistent with H_4, firms can effectively mitigate the short-term negative impact of a severe data breach by announcing it near a confounding event.

<Table 7 here>

4.3.5 Calendar time portfolio approach

We follow Fama (1998) and Mitchell and Stafford (2000) and employ the calendar time portfolio (CTP) approach to assess the long-term market effects of data breaches and the role of confounding events. This methodology avoids the empirical issues of long-run buy and hold abnormal return strategies following significant corporate events as documented by Bessembinder and Zhang (2013) and Bessembinder, Cooper, and Zhang (2019). Specifically, since corporate events, in our case data breaches and the decision to confound them, may be driven by various risk factors, forming an appropriate benchmark to compute buy and hold abnormal returns may be difficult and result in biased estimates. We construct rolling portfolios using a 12-month window, selecting firms each month based on their breach announcement date. Three equal-weighted and three value-weighted portfolios are formed: (A) a long position in confounded breach firms and a short position in unconfounded breach firms; (B) a long position in confounded breach firms and a short position in unconfounded breach firms with above-the-median breach severity; and (C) a long position in confounded breach firms and a short position in unconfounded breach firms with upper-quartile breach severity. Unlike the event-time CAR approach, which captures firm-level abnormal returns around the breach announcement, the CTP approach aggregates abnormal returns at the portfolio level over time. Portfolio returns are measured as the return differential between the treatment and the control group. This design facilitates an assessment of both the persistence of breach-related underperformance and the extent to which confounding events influence long-term market outcomes.

To account for firm size biases, portfolio returns are computed using both equal-weighted (EW) and value-weighted (VW) methods. Monthly excess returns are estimated using the Fama-French three-factor (FF3) and five-factor (FF5) models (Fama and French, 1993, 2015), where the FF3 model includes market (MKT), size (SMB), and value (HML) factors, and the FF5 model further incorporates profitability (RMW) and investment (CMA) factors. By controlling for these risk factors, the analysis isolates abnormal returns (alpha), capturing the extent to which breaches and confounding strategies lead to systematic deviations from expected performance, while controlling for the risk factors that may be related to the choice to confound itself.

Empirical results are reported in Table 8. Panel A summarizes portfolio return distributions. All portfolios indicate that confounded firms underperform unconfounded firms on average. However, there exists significant heterogeneity across observations, with the median portfolio-month earning weakly positive returns.

Panels B and C report FF3 and FF5 regression estimates. As seen in both the FF3 and FF5 regressions, it appears that risk factors significantly explain return differentials between confounded and unconfounded firms, especially in value-weighted portfolios. After controlling for these risk factors, most of the long-short portfolios do result in significant alphas. Of the six portfolios, we only observe a negative and significant alpha in one: using an equal-weighted long-short portfolio buying firms that confounded a data breach and selling firms that did not confound the most severe data breaches. This portfolio earns negative alphas (using both Fama-French three-factor and five-factor models) of around 43 basis points per month. This implies that confounding breach firms earn roughly -5% relative to non-confounding severe breach firms in the 12 months following the breach. However, given its marginal significance and lack of significant effect across other portfolios, we caution the reader in strongly interpreting this result. Taken together, it appears that while confounded breach firms tend to underperform unconfounded breach firms in the year following a breach, this effect is not largely significant. We take this as evidence that confounded firms do not experience a significant long-run reversal of the short-run return benefit they enjoyed by confounding their breach announcements and are largely able to escape the strong market penalties experienced by unconfounded firms.

<Table 8 here>

5 Robustness Tests

5.1 The Heckman two-stage regression

A substantial number of firms, which did not confound data breach events, also participated in managerial earnings management. They utilized confounding events to intentionally obscure certain significant occurrences, thereby misleading investors. However, these actions were not entirely represented in the sample of data breach announcements that were confounded. Consequently, the sample selection presented in Table 7 is not random, as confirmed by the kernel density in Figure 1, which shows a pronounced day-0 mode and elevated mass within ±3 days. This selection bias may lead us to overestimate the influence of the confounding strategy on CARs. To more effectively isolate the estimation bias induced by the selection of confounded firms, we incorporate the standard Heckman (1979) two-stage method into our model. The regression equations are presented as follows:

First-stage (selection equation):

$$
\begin{aligned}
&Prob(Confound\ dummy = 1) \\
&\qquad = b_0 + b_1 Severity_i + b_2 Data\ breach\ characteristics(IVs)_i \qquad (5) \\
&\qquad + b_3 Controls_i + \varepsilon_i
\end{aligned}
$$

Second-stage (outcome equation):

$$
\begin{aligned}
CAR_i = {} & a_0 + a_1 Confound \times Severity_i + a_2 Confound_i + a_3 Severity_i + a_4 Controls_i \\
& + a_5 IMR + \varepsilon_i \qquad (6)
\end{aligned}
$$

In the first-stage regression, eq. (5) represents a Probit model with *Confound* dummy serving as the dependent variable. This variable is assigned a value of 1 if the data breach event is confounded, and 0 otherwise. The independent variables consist of exogenous factors influencing the firm’s decision to confound, such as managerial and breach characteristics. In the second-stage regression, the dependent variables are *CARs*, with the independent and control variables consistent with those in eq. (4). Additionally, Eq. (6) includes the Inverse Mills ratio (*IMR*) derived from the first-stage regression. This variable can be utilized to estimate whether the impact of the confounding on the firm’s short-term CARs remains significant after controlling for selection bias with the unconfounded sample. We use IVs from the PRC database that reflect data breach characteristics. These include *PrevBreach*, a dummy variable indicating if the firm has previously experienced a data breach (1 for yes, 0 for no); *MultiAnn*, another binary variable indicating if the data breach event has been announced multiple times (1 for yes, 0 for no); and *IndAtt*, a measure of industry attention, calculated as the ratio of a specific industry's data breaches to the total across all SIC2 industries in the prior year. These IVs are defined and summarized in Appendix A and Appendix B (Table A2). They are chosen based on the following motivations.

Firstly, ‘Previous breaches’ can function as an IV. It fulfills the relevance criterion, as firms may adjust their strategies based on past events. For example, a firm with past data breaches may be more vigilant in managing and announcing these events to mitigate further reputational damage. It also satisfies the exclusion restriction, as it is unlikely to be associated with short-term CARs. Gatzlaff and McCullough (2010) argue that there is no difference in market reactions for a repeat occurrence, suggesting that investors have already factored in the risk of future data breaches into the firm’s stock price.

Secondly, ‘Multiple announcements’ can act as an IV. It captures breach cases involving more than one public disclosure, thereby reflecting situations where the severity of the incident or managerial responses may evolve beyond the initial announcement. Management might initially underestimate the breach’s severity or adopt a staged release strategy to mitigate regulatory and public backlash. Such cases fit naturally with the confounding narrative, as later disclosures may adjust or supplement earlier information. At the same time, ‘Multiple announcements’ is plausibly exogenous to short-term returns, as market reactions are typically concentrated on the first disclosure, with subsequent announcements attracting less attention once the event is known.

Thirdly, ‘Industry attention’ can serve as an IV. External factors, such as industry characteristics, influence the motivation of managers to confound. For instance, industries like finance and insurance often face a high frequency of data breaches. In these industries, managers might be more concerned about how to handle and announce these data breach events, needing to protect the firm's reputation and comply with relevant privacy laws. As a result, they may confound or obscure these announcements to mitigate negative public perceptions. On the other hand, 'Industry attention' is exogenous as it does not directly impact investor reactions or the firm's future performance. It merely serves as an indicator reflecting the overall data breach risk level of the industry.

In line with our economic arguments, when we regress the IVs, along with confounding and control variables on short-run performance measures, we find that none of the IVs is significantly related to any of our short-run CARs. In essence, they meet the IV exclusion requirement. We employ a first-step probit model with the Confound dummy as the outcome variable. In line with standard practice, we include all control variables from the second-step equation in the first-step, excluding Confound dummy × Severity and Confound dummy due to their high correlation with the outcome variable in this sample of confounded and unconfounded breaches. As demonstrated in Table 9, our three IVs exhibit a significant and positive association with the confounding indicator. The first-stage models possess high explanatory power, as indicated by pseudo R^2 statistics exceeding 30%.

The second-step regression results show that the IMRs are insignificant, confirming no severe sample selection issue and so unbiased OLS coefficients in Table 7 (Panel D). The joint effect of the confound dummy and breach severity consistently and strongly impacts the firm's short-term returns. When breach severity is high, confounding mitigates the negative impact of the breach announcement.

<Table 9 here>

5.2 Additional tests

5.2.1 Confounding and breaches including financial information

Kamiya et al. (2021) base their analysis on breaches that contain financial information. We predict that severe breaches containing financial information will be more likely to be confounded. We classify breach types involving financial access (e.g., bank account credentials, credit card data), identity theft (information enabling impersonation), existential data (information critical to national security or business continuity), and other financial information breaches as financial information loss. As seen in Table 10, we find that breach severity still significantly increases the likelihood of confounding when controlling for the loss of financial information. However, the loss of financial information only increases the likelihood of confounding when breach severity is high and not independently. We interpret this evidence as consistent with the findings of Table 6 and the predictions of H_1.

<Table 10 here>

5.2.2 Timing of confounding events: anticipatory, same-day, and reactive strategies

Finally, we study whether the negative effects of breach severity and confounding strategies on firm value vary on whether the confounding event occurs in the 30 days prior to the event (anticipatory), the day of the event, or in the 30 days following the event (reactive). We report the results of these tests in Table 11. Interestingly, and consistent with our predictions we find that our results are centered on confounding events that occur on the same day as the breach announcement, and to a lesser extent anticipatory confounding events. This evidence is consistent with firms actively confounding breach events to mitigate their negative announcement impact, rather than randomly choosing breach announcement dates that happen to fall near confounding events.

<Table 11 here>

6 Conclusion

This paper examines the influence of data breaches on stock returns, as well as the role of managerial information asymmetry and confounding strategies. Using 675 data breaches from U.S. listed firms (2005-2022), we constructed a data breach severity score, calculated CARs, conducted an empirical analysis using event study methodology and multiple regressions, and validated our results through robustness and endogeneity tests.

We find evidence that firms often confound negative events, in our setting severe data breaches, by announcing other events near the announcement of the breach. We find that severe breach events significantly reduce firm value. We also find firms with worse managerial information asymmetry and governance are more likely to confound severe breach events, even after controlling for general news pressure on the day of the announcement. In other words, when facing severe data breaches and significant managerial information asymmetry, managers may use positive confounding events to offset data breaches bad news and mislead investors. Importantly, we find evidence that confounding strategies can moderate the negative impact of severe data breaches, and that firms face limited to no long-term return repercussions for this strategy. This provides insight into why managers resort to confounding strategies. These results do not appear to be driven by a series of alternative mechanisms.

Taken together, these findings demonstrate that requirements that give managers flexibility in reporting bad news to shareholders may create incentives to obfuscate or confound this news. Our evidence suggests that managers choose to report bad news near other firm-level news announcements to ameliorate the impact of announcements of negative firm-level information. Additionally, our analysis provides evidence that, since confounding events may be endogenously chosen, empirical researchers should carefully consider removing observations that coincide with confounding events.

Appendix A. Variable Definitions

Variables	Definition
CAR	Cumulative abnormal returns over different event windows, calculated using both the market-adjusted model (MAM) and the Fama-French five-factor model (FF5).
Confound	Equal to 1 if the breached firm confounded 7 days before, 30 days before, or 30 days before and 30 days after the announcement, and 0 otherwise.
Anticipatory	Within the 61-day window around the breach announcement (30 days before and 30 days after), a confounding event announced before the breach is defined as 1, otherwise 0.
Same-day	A confounding event announced on the same day as the breach is defined as 1, otherwise 0.
Reactive	A confounding event announced after the breach within this window is defined as 1, otherwise 0.
Severity	Log (N*t*s*A), N = the total number of records breached, or, in the case of intellectual property loss the equivalent dollar loss, t = the type of data in the records, s = source of the breach, A = whether or not the stolen data has been used to cause harm, be it identity theft, credit application, or bank account withdrawals.
NewsPr	Ln (1 + news pressure). News pressure is the daily count of WSJ and WSJ online articles under 'Financial and Commodities Markets News' for the breach announcement day, sourced from Factiva.
FinInfo	Breaches involving financial access, identity theft, existential data, and other financial data are classified as financial information loss, with FinInfo set to 1 for these cases and 0 otherwise (Kamiya et al., 2021).
Managerial characteristics	
CEOEq	CEO equity holdings, defined as the fraction of the firm’s shares owned by the CEO.
CEOTen	The duration the CEO has been operating in the CEO position to the year under study.
CEOEqComp	Equity-based compensation, defined as (stock options granted + restricted stocks granted) / total compensation (Song and Wan, 2019).
BoardInd	Non-exec board independence (%), defined as the proportion of non-executive directors (BoardEX dataset).
EarnMan	Earnings management, defined as estimates of discretionary current accruals (lagged by total assets) (Raman and Shahrur, 2008).
ManEnt	Managerial entrenchment, defined as antitakeover protection index attached to each state with a score from 0 to 5 that is equal to the number of 5 key antitakeover statutes that it has (Bebchuk and Cohen, 2005; Karpoff and Wittry, 2018).
PriorMis	Prior misstatement. Equal to 1 if a firm-year’s financial statement was reported in Accounting and Auditing Enforcement Releases (AAER) before the breach announcement, and 0 otherwise (Lo, Ramos, and Rogo, 2017).
OverCon	Equal to 1 if a CEO holds vested options that are at least 67% in the money once, and 0 otherwise (Hirshleifer, Low, and Teoh, 2012).
RiskComm	Risk committee. Equal to 1 if a board committee name includes the word “risk,” and 0 otherwise (Kamiya et al., 2021; BoardEx dataset).
Firm-level variables	
Size	Ln (market capitalization), measured at year t-1 prior to the data breach.
TobinQ	(Total assets – common/ordinary equity + market value of equity) / total assets.
Leverage	Total debt / total assets.
ROA	Operating income before depreciation / total assets.
Cash	Cash / total assets.
R&D	R&D expenditures / total assets.

Industry-level features	
HighTech	Equal to 1 if the breached firm belongs to high-tech industries in manufacturing (NAICS 3254, 3341, 3342, 3344, 3345, and 3364), information (5112, 5161, 5179, 5181, and 5182), and services (5413, 5415, and 5417), and 0 otherwise (Hecker, 2005).
HHI	10-K TNIC industry concentration data from the Hoberg-Phillips Data Library (Hoberg and Phillips, 2010, 2016).
Fluid	10-K based product market fluidity data from the Hoberg-Phillips Data Library (Hoberg, Phillips, and Prabhala, 2014).
ProdSim	10-K TNIC total similarity data from the Hoberg-Phillips Data Library (Hoberg and Phillips, 2016).
Data breach characteristics used as IVs	
PrevBreach	Equal to 1 if the breached firm had a breach before the current announcement, and 0 otherwise.
MultiAnn	Equal to 1 if the breached firm made multiple announcements about the same breach, and 0 otherwise.
IndAtt	Industry attention, defined as the ratio of the number of breach announcements in the firm's industry to the total number of breach announcements in all industries in the year before the firm's breach announcement.
Matching variables for PSM	
Age	Ln (the current year - the year of listing).
ROE	Net income / total equity.
Tangi	(Total property, plant, and equipment + inventories) / total assets
Div	The amount of dividends payable / net profit
Size	As previously defined.
TobinQ	As previously defined.
Leverage	As previously defined.
Cash	As previously defined.

Note: This table serves as a comprehensive reference guide, detailing the variables used in the study, their definitions, and calculation methods. All variables (unless stated otherwise) are measured at year t-1 prior to the data breach. The table provides definitions for several categories of variables: key variables of this study, managerial characteristics, firm-level characteristics, industry-level characteristics, data breach characteristics used as IVs, and matching variables needed for PSM. It also provides references to the sources or methods used to calculate or derive each variable, enhancing the transparency and replicability of the study. This table is an essential resource for understanding the data and methodology used in the study, providing the necessary context for interpreting the study's findings, and contributing to the robustness and credibility of the research.

References

Acquisti, Alesandro, Allan Friedman, and Rahul Telang, 2006, Is there a cost to privacy breaches? An event study, *Proceedings of the Twenty-Seventh International Conference on Information Systems,* Milwaukee, WI.

Agrawal, Anup and Sahibi Chadha, 2005, Corporate governance and accounting scandals, *Journal of Law and Economics* 48(2), 371-406.

Ahmed, Anwer S., and Scott Duellman, 2013, Managerial overconfidence and accounting conservatism, *Journal of Accounting Research* 51(1), 1-30.

Aivazpour, Zahra, Rohit Valecha, and Rajarshi Chakraborty, 2018, The impact of data breach severity on post-breach online shopping intention, *International Conference on Interaction Sciences.*

Akerlof, George A, 1970, The market for "Lemons": Quality uncertainty and the market mechanism, *The Quarterly Journal of Economics* 84(3), 488-500.

Akey, Pat, Stefan Lewellen, Inessa Liskovich, and Christoph Schiller, 2024, Hacking corporate reputations, European Corporate Governance Institute Working Paper.

Bebchuk, Lucian A., and Alma Cohen, 2005, The costs of entrenched boards, *Journal of Financial Economics* 78(2), 409-433.

Bebchuk, Lucian A., Alma Cohen, and Allen Ferrell, 2009, What matters in corporate governance?, *Review of Financial Studies* 22(2), 783-827.

Benischke, Mirko H., Geoffrey P. Martin, and Lotte Glaser, 2019, CEO equity risk bearing and strategic risk taking: The moderating effect of CEO personality, *Strategic Management Journal* 40(1), 153-177.

Bessembinder, Hendrick, Michael J. Cooper, and Feng Zhang, 2019, Characteristic-based benchmark returns and corporate events, *Review of Financial Studies* 32(1), 75-125.

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

Beyer, Anne, Daniel A. Cohen, Thomas Z. Lys, and Beverly R. Walther, 2010, The financial reporting environment: Review of the recent literature, *Journal of Accounting and Economics* 50(2-3), 296-343.

Bloomfield, Robert J., 2002, The “incomplete revelation hypothesis” and financial reporting, *Accounting Horizons* 16(3), 233-243.

Brauer, Matthias, and Louis Vandepoele. 2024, Managing the “downside” of downsizing: Firms’ impression offsetting around downsizing announcements. *Journal of Management Studies* 61(8), 3684-3716.

Bryan, Stephen, LeeSeok Hwang, and Steven Lilien, 2000, CEO stock-based compensation: An empirical analysis of incentive-intensity, relative mix, and economic determinants, *The Journal of Business* 73(4), 661-693.

Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan, 2004, The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers, *International Journal of Electronic Commerce* 9(1), 70-104.

Campbell, Katherine, Lawrence A. Gordon, Martin P. Loeb, and Lei Zhou, 2003, The economic cost of publicly announced information security breaches: empirical evidence from the stock market, *Journal of Computer Security* 11(3), 431-448.

Casamatta, Catherine, and Alexander Guembel, 2010, Managerial legacies, entrenchment, and strategic inertia, *The Journal of Finance* 65(6), 2403-2436.

Cohen, Lauren, Dong Lou, and Christopher J. Malloy, 2020, Casting conference calls, *Management Science* 66(11), 4921-5484.

Coles, Jeffrey L., Naveen D. Daniel, and Lalitha Naveen, 2008, Boards: Does one size fit all? *Journal of Financial Economics* 87(2), 329-356.

Cummins, Mark, Bruce Grundy, Ronan Powell, and Pierangelo Rosati, 2023, The Effect of Data Breaches on Breached Firms and their Competitors, Working paper.

D'Arcy, John, Idris Adjerid, Corey M. Angst, and Ante Glavas, 2020, Too good to be true: Firm social performance and the risk of data breach, *Information Systems Research* 31(4), 1200-1223.

deHaan, Ed, Terry Shevlin, and Jacob Thornock, 2015, Market (in)attention and the strategic scheduling and timing of earnings announcements, *Journal of Accounting and Economics* 60(1), 36-55.

DellaVigna, Stefano, and Joshua M. Pollet, 2009, Investor inattention and Friday earnings announcements, *The Journal of Finance* 64(2), 709-749.

Doukas, John A., and Dimitris Petmezas, 2007, Acquisitions, overconfident managers and self-attribution bias, *European Financial Management* 13(3), 531-577.

Durante, Ruben, and Ekaterina Zhuravskaya, 2018, Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict, *Journal of Political Economy* 126(3), 1085-1133.

Ebrahimi, Sepideh, and Kamran Eshghi, 2022, A meta-analysis of the factors influencing the impact of security breach announcements on stock returns of firms, *Electronic Markets* 32(4), 2357-2380.

Eling, Martin, and Nicola Loperfido, 2017, Data breaches: Goodness of fit, pricing, and risk measurement, *Insurance: Mathematics and Economics* 75, 126-136.

Evans III, John Harry, Nandu J. Nagarajan, and Jason D. Schloetzer, 2010, CEO turnover and retention light: Retaining former CEOs on the board, *Journal of Accounting Research* 48(5), 1015-1047.

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

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

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

Faulkender, Michael, and Jun Yang, 2010, Inside the black box: The role and composition of compensation peer groups, *Journal of Financial Economics* 96(2), 257-270.

Fedyk, Tatiana, and Natalya V. Khimich, 2018, Hidden in the woodshed: Big bath herding, *Available at SSRN 3224962.*

Foerderer, Jens, and Sebastian W. Schuetz, 2022, Data breach announcements and stock market reactions: A matter of timing? *Management Science* 68(10), 7298-7322.

Gatzlaff, Kevin M., and Kathleen A. McCullough, 2010, The effect of data breaches on shareholder wealth, *Risk Management and Insurance Review* 13(1), 61-83.

Garg, Ashish, Jeffrey Curtis, and Hilary Halper, 2003, Quantifying the financial impact of IT security breaches, *Information Management & Computer Security* 11(2), 74-83.

Gaver, Jennifer J., Kenneth M. Gaver, and Jeffrey R. Austin, 1995, Additional evidence on bonus plans and income management, *Journal of Accounting and Economics* 19(1), 3-28.

Goffman, Erving, 2016, The presentation of self in everyday life, *Social Theory Re-wired*, 482-493.

Goldstein, James, Anna Chernobai, and Michel Benaroch, 2011, An event study analysis of the economic impact of IT operational risk and its subcategories, *Journal of the Association for Information Systems* 12(9), 606-631.

Grossman, Sanford J., Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70(3), 393–408.

Guernsey, Scott, Simone M. Sepe, and Matthew Serfling, 2022, Blood in the water: The value of antitakeover provisions during market shocks, *Journal of Financial Economics* 143(3), 1070-1096.

Guest, Nicholas M., and Jiawen Yan, 2025, Circuitousness in disclosure narratives, Cornell University Working Paper.

Gwebu, Kholekile L., Jing Wang, and Li Wang, 2018, The role of corporate reputation and crisis response strategies in data breach management, *Journal of Management Information Systems* 35(2), 683–714.

Harford, Jarrad, Mark Humphery-Jenner, and Ronan Powell, 2012, The sources of value destruction in acquisitions by entrenched managers, *Journal of Financial Economics* 106(2), 247-261.

Healy, Paul M., 1985, The effect of bonus schemes on accounting decisions, *Journal of Accounting and Economics* 7(1-3), 85-107.

Healy, Paul M., and Krishna G. Palepu, 2001, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* 31(1-3), 405-440.

Heckman, James J., 1979, Sample selection bias as a specification error, *Econometrica: Journal of the econometric society* 47(1), 153-161.

Hilary, Gilles, Benjamin Segal, and May H. Zhang, 2016, Cyber-risk disclosure: Who cares? Georgetown University Working Paper.

Hirshleifer, David, Sonya S. Lim, and Siew Hong Teoh, 2009, Driven to distraction: Extraneous events and underreaction to earnings news, *The Journal of Finance* 64(5), 2289-2325.

Hirshleifer, David, Angie Low, and Siew Hong Teoh, 2012, Are overconfident CEOs better innovators? *The Journal of Finance* 67(4), 1457-1498.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *Review of Financial Studies* 23(10), 3773-3811.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124(5), 1423-1465.

Hoberg, Gerard, Gordon Phillips, and Nagpurnanand Pranhala, 2014, Product market threats, payouts, and financial flexibility, *The Journal of Finance* 69(1), 293-324.

Hovav, Anat, and John D'Arcy, 2003, The impact of denial-of-service attack announcements on the market value of firms, *Risk Management and Insurance Review* 6(2), 97-121.

Janakiraman, Ramkumar, Joon Ho Lim, and Rishika Rishika, 2018, The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer, *Journal of Marketing* 82(2), 85-105.

Jensen, Michael C., and William H. Meckling, 1976, Theory of the firm: Managerial behavior, agency costs and ownership structure, *Corporate Governance*, 77-132.

Jewell, Jeffrey J., and Jeffrey A. Mankin, 2011, What is your ROA? An investigation of the many formulas for calculating return on assets, *Academy of Educational Leadership Journal* 15(2), 79-91.

Jung, Michael J., James P. Naughton, Ahmed Tahoun, and Clare Wang, 2018, Do firms strategically disseminate? Evidence from corporate use of social media, *The Accounting Review* 93(4), 225-252.

Kaczmarek, Szymon, Satomi Kimino, and Annie Pye, 2012, Board task-related faultlines and firm performance: A decade of evidence, *Corporate Governance: An International Review* 20(4), 337-351.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz, 2021, Risk management, firm reputation, and the impact of successful cyberattacks on target firms, *Journal of Financial Economics* 139(3), 719-749.

Kannan, Karthik, Jackie Rees, and Sanjay Sridhar, 2007, Market reactions to information security breach announcements: An empirical analysis, *International Journal of Electronic Commerce* 12(1), 69-91.

Karpoff, Jonathan M., and Michael D. Wittry, 2018, Institutional and legal context in natural experiments: The case of state antitakeover laws, *Journal of Finance* 73(2), 657-714.

Kothari, Sagar P., Andrew J. Leone, and Charles E. Wasley, 2005, Performance matched discretionary accrual measures, *Journal of Accounting and Economics* 39(1), 163-197.

Knyazeva, Anzhela, Diana Knyazeva, and Ronald W. Masulis, 2013, The supply of corporate directors and board independence, Review of Financial Studies 26(6), 1561-1605.

Lending, Claire, Kristina Minnick, and Patrick J. Schorno, 2018, Corporate governance, social responsibility, and data breaches, *Financial Review* 53(2) 413-455.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45(2-3), 221-247.

Lo, Kin, Felipe Ramos, and Rafael Rogo, 2017, Earnings management and annual report readability, *Journal of Accounting and Economics* 63(1), 1-25.

Luo, Xueming, Vamsi K. Kanuri, and Michelle Andrews, 2014, How does CEO tenure matter? The mediating role of firm-employee and firm-customer relationships, *Strategic Management Journal* 35(4), 492-511.

Makridis, Christos, and Benjamin Dean, 2018, Measuring the economic effects of data breaches on firm outcomes: Challenges and opportunities, *Journal of Economic and Social Measurement* 43(1-2): 59-83.

Malmendier, Ulrike, and Geoffrey Tate, 2008, Who makes acquisitions? CEO overconfidence and the market's reaction, *Journal of Financial Economics* 89, 20-43.

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier, 2017, Data privacy: Effects on customer and firm performance, *Journal of Marketing* 81(1), 36-58.

Masulis, Ronald W., Cong Wang, and Fei Xie, 2007, Corporate governance and acquirer returns, *The Journal of Finance* 62(4), 1851-1889.

McClelland, Patrick L., Xin Liang, and Vincent L. Barker, 2010, CEO commitment to the Marketing status quo: Replication and extension using content analysis. *Journal of Management* 36(5), 1251-1277.

Michaely, Roni, Amir Rubin, and Alexander Vedrashko, 2014, Corporate governance and the timing of earnings announcements, *Review of Finance* 18(6), 2003-2044.

Mitchell, Mark L., and Erik Stafford, 2000, Managerial decisions and long-term stock price performance, *Journal of Business* 73(3), 287-329.

Morck, Randall, Andrei Shleifer, and Robert W. Vishny, 1990, Do managerial objectives drive bad acquisitions? *Journal of Finance* 45(1), 31-48.

Mukherjee, Ujjal K., George P. Ball, Kaitlin D. Wowak, Karthik V. Natarajan, and Jason W. Miller, 2022, Hiding in the herd: The product recall clustering phenomenon. *Manufacturing and Service Operations Management* 24(1), 392-410.

Palepu, Krishna G., 1986, Predicting takeover targets: A methodological and empirical analysis, *Journal of Accounting and Economics* 8(1), 3-35.

Palmrose, Zoe Vonna, Vernon J. Richardson, and Susan Scholz, 2004, Determinants of market reactions to restatement announcements, *Journal of Accounting and Economics* 37(1), 59-89.

Pound, John, 1988, Proxy contests and the efficiency of shareholder oversight, *Journal of Financial Economics* 20, 237-265.

Privacy Rights Clearinghouse (PRC), 2021, Chronology of Data Breaches: FAQ. Available at: https://privacyrights.org/data-breaches

Rasoulian, Shahin, Yany Grégoire, Renaud Legoux, and Sylvain Sénécal, 2023, The effects of service crises and recovery resources on market reactions: An event study analysis on data breach announcements, *Journal of Service Research* 26(1), 44-63.

Rawson, Caleb, Brady J. Twedt, and Jessica C. Watkins. 2023, Managers' strategic use of concurrent disclosure: Evidence from 8-K filings and press releases. *The Accounting Review* 98(4), 345-371.

Ray, Soumya, Terence Ow, and Sung S. Kim, 2011, Security assurance: How online service providers can influence security control perceptions and gain trust, *Decision Sciences* 42(2), 391-412.

Richardson, Vernon J., 2000, Information asymmetry and earnings management: Some evidence. *Review of Quantitative Finance and Accounting* 15, 325-347.

Rosati, Pierangelo, Mark Cummins, Peter Deeney, Fabian Gogolin, Lisa van der Werff, and Theo Lynn, 2017, The effect of data breach announcements beyond the stock price: Empirical evidence on market activity, *International Review of Financial Analysis* 49(C), 146-154.

Schrand, Catherine M., and Sarah LC Zechman, 2012, Executive overconfidence and the slippery slope to financial misreporting, *Journal of Accounting and Economics* 53(1-2), 311-329.

Shleifer, Andrei, and Robert W. Vishny, 1989, Management entrenchment: The case of manager-specific investments, *Journal of Financial Economics* 25(1), 123-139.

Song, Wei-Ling, and Kam-Ming Wan, 2019, Does CEO compensation reflect managerial ability or managerial power? Evidence from the compensation of powerful CEOs, *Journal of Corporate Finance* 56(C), 1-14.

Souder, David, Zeki Simsek, and Scott G. Johnson, 2012, The differing effects of agent and founder CEOs on the firm's market expansion, *Strategic Management Journal* 33(1), 23-41.

Stiennon, Richard, 2013, Categorizing data breach severity with a breach level index. URL: https://breachlevelindex. com/pdf/Breach-Level-Index-WP. pdf.

Sun, Hong, Maochao Xu, and Peng Zhao, 2021, Modeling malicious hacking data breach risks, *North American Actuarial Journal* 25(4), 484-502.

Taylor, Lucian A, 2010, Why are CEOs rarely fired? Evidence from structural estimation, *The Journal of Finance* 65(6), 2051-2087.

Tukey, John W., 1977, Exploratory data analysis, Springer.

Wang, Qian, Eric W. T. Ngai, Daniel Pienta, and Jason Bennett Thatcher, 2023, Information technology innovativeness and data-breach risk: A longitudinal study, *Journal of Management Information Systems* 40(4), 1139-1170.

Warfield, Terry D., John J. Wild, and Kenneth L. Wild, 1995, Managerial ownership, accounting choices, and informativeness of earnings, *Journal of Accounting and Economics* 20(1), 61-91.

Warner, Jerold B., Ross L. Watts, and Karen H. Wruck, 1988, Stock prices and top management changes, *Journal of Financial Economics* 20, 461–492.

Weisbach, Michael S., 1988, Outside directors and CEO turnover, *Journal of Financial Economics* 20, 431–460.

Wen, Yu, Kami Rwegasira, and Jan Bilderbeek, 2002, Corporate governance and capital structure decisions of the Chinese listed firms, *Corporate Governance: An International Review* 10(2), 75-83.

Yayla, Ali Alper, and Qing Hu, 2011, The impact of information security events on the stock value of firms: The effect of contingency factors, *Journal of Information Technology* 26(1), 60-77.

Table 1 Composition of Breached and Non-breached Sample

Industry & Year of breach		Breached sample	Non-breached sample
Standard Industry Classification Code (SIC2)	01–09 Agriculture, Forestry, Fishing	0	1
	10–14 Mining	3	5
	15-17 Construction	3	5
	20-39 Manufacturing	126	120
	40-49 Transportation & Public Utilities	45	46
	50-51 Wholesale Trade	19	12
	52-59 Retail Trade	93	94
	60-67 Finance, Insurance, Real Estate	243	238
	70-89 Services	138	152
	91-99 Public Administration	5	2
	Total	675	675
Years	2005	3	3
	2006	38	38
	2007	40	40
	2008	19	19
	2009	6	6
	2010	38	38
	2011	29	29
	2012	27	27
	2013	19	19
	2014	38	38
	2015	20	20
	2016	23	23
	2017	24	24
	2018	24	24
	2019	220	220
	2020	60	60
	2021	45	45
	2022	2	2
	Total	675	675

Note: This table offers a detailed view of the distribution and frequency of data breaches across different industries and years. It shows the composition of both breached and non-breached samples matched based on the highest Hoberg Phillips 10-K products text-based pairwise similarity scores, sorted by the industry of the breached firm and the year of the breach. The industry classification used is the Standard Industry Classification Code (SIC2), which includes 10 major categories and 58 distinct industries. The data breaches span from 2005 to 2022. For each year within this period, the number of breached samples is equated with the number of non-breached samples.

Table 2 Summary Statistics of Breach Severity

Variable: Breach severity	N	Mean	Median	75th Percentile	Std. Dev.	Min	Max
BS1	675	2.745	2.204	4.000	1.937	0.301	9.193
BS2	675	3.158	2.992	4.030	1.694	0.301	9.193

Note: This table shows summary statistics for two data breach severity variables, including observations (N), mean, median, 75th percentile, standard deviation (Std. Dev.), minimum value (Min), and maximum value (Max). The severity of a data breach is calculated based on the Breach Level Index (BLI) (refer to Stiennon, 2013). The formula for BLI is: BLI = lg (N*t*s*A), where N is the total number of records breached or the equivalent dollar loss in the event of intellectual property loss, t is the type of data in the records, s is the source of the breach, and A is whether the stolen data has been used to cause harm such as identity theft, credit application, or bank account withdraws. Due to missing values in the data, we use two methods to calculate the breach severity: (1) Replace missing values with 1 (*BS1*); (2) Replace missing values with the average value of the breach type (*BS2*).

Table 3 Summary Statistics of Variables for Breached and Non-breached Firms

	Breached firms		Non-breached firms		Difference of breached and non-breached firms
Variable	Mean	Std. Dev.	Mean	Std. Dev.	
%CAR (3)	-0.066%*	0.033	0.100%	0.442	-0.166%
%CAR (5)	-0.352%*	0.049	-0.116%	0.051	-0.236%*
Market Capitalization ($million)	42,491	89,828	10,458	26,001	32,033***
7-Day Confounding	0.477	0.500			
30-Day Confounding	0.573	0.495			
60-Day Confounding	0.664	0.473			
Size	8.827	2.217	8.195	2.261	0.632***
TobinQ	2.241	2.249	1.985	1.728	0.256***
Leverage	0.278	0.250	0.256	0.209	0.022***
ROA	0.037	0.130	0.017	0.136	0.02**
Cash	0.098	0.115	0.027	0.078	0.071***
R&D	0.024	0.065	0.011	0.062	0.013****
HighTech	0.159	0.365	0.103	0.427	0.056***
HHI	0.204	0.195	0.156	0.143	0.048***
Fluid	6.599	3.414	7.608	3.527	-1.009***
ProdSim	8.013	14.001	13.493	20.367	-5.480**
CEOEq	1.699	4.991	1.684	4.960	0.015***
CEOTen	8.892	7.041	7.656	7.713	1.236***
CEOEqComp	0.023	0.102	0.018	0.163	0.005**
BoardInd	0.109	0.159	0.125	0.173	-0.016***
EarnMan	-0.001	0.048	-0.006	0.031	0.005
ManEnt	2.987	1.935	2.736	1.749	0.251***
PriorMis	0.053	0.225	0.048	0.313	0.005*
OverCon	0.350	0.477	0.318	0.412	0.032**

Note: This table presents the descriptive statistics of the main variables in this paper for both breached and non-breached samples, including count (N), mean, and standard deviation (Std. Dev.). It also displays the mean difference between the groups. *P-values*, indicated by asterisks, are only shown for *CARs* and mean differences (*** for 1% significance level, ** for 5% significance level, and * for 10% significance level). *CARs* around the data breach announcement are calculated using the market-adjusted model (MAM). Additional *CARs* calculated through other windows or using the Fama-French 5-Factor Model (FF5) can be found in the Appendix B (Table A2). The table shows the average *CARs* of 675 breached firms and their closest non-breached competitors around the announcement dates. It also presents the *market value* a year before the data breach. We define confounding variables based on different assumptions for different periods. Firm-level variables include *Size*, *TobinQ*, *Leverage*, *ROA*, *Cash*, and *R&D*. Industry-level control variables include *HighTech* (firm's industry), *HHI* (TNIC3-defined industry concentration), *Fluid* (Hoberg, Phillips, and Prabhala (2014) fluidity measure of product market innovation), and *ProdSim* (total product similarity between firm's products and those of competitors in its TNIC3-defined industry). Variables capturing management information asymmetry include *CEOEq* (CEO's shareholding proportion), *CEOEqComp* (CEO's equity incentives), *CEOTen* (CEO's tenure), *BoardInd* (board of directors' independence), *EarnMan* (firm's earnings management level), *ManEnt* (firm's management entrenchment degree), and *PriorMis* (firm's past financial fraud). The definitions and calculation methods of each variable can be seen in the Appendix A.

Table 4 %CARs of Breached Firms and Closest-Comparable Rivals

	CAR (3)	CAR (5)
Panel A: Average of 324 breached and their closest-comparable %CAR and $CAR for breaches in the above-median of Severity		
Breached %CAR	-0.493%***	-0.822%***
Breached $CAR ('000)	-67,548	-102,332
Closest-comparable %CAR	0.008%	-0.158%
Closest-comparable $CAR ('000)	4,772	21,828
Difference in %CARs of breached and non-breached firms	-0.511%***	-0.664%***
Difference in $CARs of breached and non-breached firms	-72,320	-124,160
Panel B: Average of 164 breached and closest-comparable %CAR and $CAR for breaches in the upper-quartile of Severity		
Breached %CAR	-0.798%***	-1.302%***
Breached $CAR ('000)	-202,866**	-170,087
Closest-comparable %CAR	-0.247%	-0.395%
Closest-comparable $CAR ('000)	-57,536	-35,206
Difference in %CARs of breached and non-breached firms	-0.551%***	-0.907%***
Difference in $CARs of breached and non-breached firms	-145,330**	-134,881

Note: Closest-comparable firms are as defined in Table 1. *%CAR(t)* is the cumulative abnormal return relative to the value-weighted market over t days centered on the breach, with t = 3 and 5. *$CAR* is the product of *%CAR* and the market value of equity five days before the breach, expressed in thousands. The table also presents the difference in *%CARs* and *$CARs* between breached and non-breached firms, along with their *t-test* statistical significance. Panel A shows the average *%CAR* and *$CAR* of the 324 breached firms above the median breach severity and their 324 closest non-breached rivals. Panel B shows the average *%CAR* and *$CAR* of the 164 breached firms in the upper quartile of breach severity and their 164 closest non-breached rivals. The default breach severity uses *BS1*. Standard errors are robust clustered at the firm level. Standard errors are shown in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels respectively.

Table 5 Regression of %CARs of Breached Firms on Breach Severity

Panel A

	CAR (3)	CAR (5)
Severity	-0.396***	-0.443***
	(0.099)	(0.133)
Size	-0.103	0.117
	(0.137)	(0.195)
TobinQ	-0.039	0.044
	(0.118)	(0.180)
Leverage	0.048	0.439
	(0.860)	(1.134)
ROA	3.158	1.551
	(3.264)	(5.226)
Cash	-0.367	-2.838
	(1.970)	(2.560)
R&D	0.000	0.000
	(0.000)	(0.000)
HighTech	-0.055	-0.082
	(0.628)	(0.898)
HHI	-0.777	-0.161
	(0.883)	(1.358)
Fluid	-0.026	0.020
	(0.070)	(0.104)
ProdSim	0.033	0.025
	(0.032)	(0.040)
NewsPr	0.498	1.049*
	(0.385)	(0.542)
Constant	3.613	0.195
	(2.632)	(3.814)
Year, Ind, State FE	Yes	Yes
Observations	526	526
R^2	0.276	0.327

Panel B					
	CAR (7)	CAR (9)	CAR (11)	CAR (13)	CAR (15)
Severity	-0.517***	-0.468***	-0.441**	-0.369*	-0.286
	(0.160)	(0.172)	(0.179)	(0.200)	(0.207)
Size	0.162	0.232	0.233	0.442	0.613
	(0.248)	(0.275)	(0.301)	(0.355)	(0.410)
TobinQ	-0.011	0.074	0.165	-0.022	-0.114
	(0.209)	(0.242)	(0.278)	(0.307)	(0.299)
Leverage	-0.241	1.321	2.071	0.259	-0.934
	(1.504)	(1.705)	(1.798)	(2.115)	(2.260)
ROA	3.794	2.937	1.501	-0.391	-1.408
	(5.794)	(6.179)	(6.343)	(7.494)	(8.198)
Cash	-1.208	-2.054	-1.321	-1.468	-0.370
	(3.613)	(3.979)	(4.241)	(4.849)	(4.725)
R&D	0.000	-0.000	-0.000	-0.000	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
HighTech	0.130	-0.159	-0.627	-0.996	-1.140
	(1.005)	(1.129)	(1.159)	(1.376)	(1.523)
HHI	-0.166	-1.217	-0.681	-0.956	-1.091
	(1.666)	(2.051)	(2.397)	(2.513)	(2.347)
Fluid	0.018	-0.037	0.075	-0.033	-0.019
	(0.131)	(0.152)	(0.165)	(0.183)	(0.192)
ProdSim	0.012	0.034	-0.003	-0.002	-0.034
	(0.055)	(0.089)	(0.084)	(0.068)	(0.079)
NewsPr	1.181*	0.952	1.057	0.792	0.725
	(0.656)	(0.721)	(0.806)	(0.956)	(1.085)
Constant	0.828	-1.367	-1.577	-4.892	-7.729
	(4.846)	(4.998)	(6.140)	(7.450)	(8.048)
Year, Ind, State FE	Yes	Yes	Yes	Yes	Yes
Observations	526	526	526	526	526
R^2	0.330	0.303	0.304	0.277	0.312

Note: This table reports regressions of the *%CARs* of breached firms on breach severity, year dummies, industry dummies, and state dummies. The default is to use the CAR calculated by the market-adjusted model (MAM). The typical shorter-run CAR includes *CAR (3)*, *CAR (5), CAR (7)*, *CAR (9)*, *CAR (11)*, *CAR (13)*, and *CAR (15)*. The default breach severity uses *BS1*, and the regression results of *BS2* are seen in the Appendix B (Table A7) for robustness checks. The regression results of CARs calculated by the Fama-French 5-Factor Model (FF5) can be found in the Appendix B (Table A8). Standard errors are robust clustered at the firm level and are shown in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels respectively.

Table 6 The Predictive Model of Managerial Influence on Confounding

Panel A: Predictive Model of Confounding

	Confound										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
BS1	0.153**										0.164*
	(0.070)										(0.094)
OverCon		1.525***									1.295**
		(0.366)									(0.512)
EarnMan			-0.319								2.679
			(2.757)								(4.159)
CEOEq				0.046*							0.043
				(0.027)							(0.035)
CEOTen					0.106***						0.053*
					(0.031)						(0.028)
CEOEqComp						2.316					5.647
						(2.259)					(3.552)
BoardInd							-1.433				-2.989**
							(0.902)				(1.433)
ManEnt								0.199			-0.583
								(0.363)			(0.441)
PriorMis									2.161***		1.486**
									(0.619)		(0.750)
RiskComm										0.162	-0.007
										(0.313)	(0.461)
Size	0.124	0.176*	0.137	0.463***	0.196**	0.422***	0.102	0.128	0.096	0.167*	0.432***
	(0.082)	(0.098)	(0.085)	(0.109)	(0.094)	(0.104)	(0.085)	(0.082)	(0.082)	(0.092)	(0.168)
TobinQ	0.134*	0.073	0.127	0.152	0.130*	0.164*	0.139*	0.139*	0.156**	0.172**	0.085
	(0.077)	(0.088)	(0.081)	(0.095)	(0.075)	(0.095)	(0.074)	(0.074)	(0.073)	(0.085)	(0.119)
Leverage	0.421	0.755	0.377	1.386	0.636	1.378	0.417	0.489	0.716	0.554	0.800
	(0.701)	(0.844)	(0.751)	(0.878)	(0.821)	(0.880)	(0.666)	(0.670)	(0.680)	(0.703)	(1.330)
ROA	-0.811	-2.908	-2.283	-1.447	-1.752	-1.146	-1.382	-1.327	-1.294	-2.937	-4.725
	(2.130)	(2.426)	(2.599)	(2.968)	(2.029)	(2.875)	(2.069)	(2.041)	(2.001)	(2.306)	(3.863)
Cash	0.924	1.067	1.665	0.680	1.053	0.479	0.791	0.998	1.175	1.126	0.865
	(1.511)	(1.670)	(1.601)	(1.741)	(1.575)	(1.757)	(1.522)	(1.502)	(1.557)	(1.760)	(2.166)
R&D	-0.000	-0.000	-0.000	-0.000*	-0.000*	-0.000**	-0.000	-0.000	-0.000	-0.000	-0.001*

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HighTech	1.769***	1.976***	1.751***	2.008***	1.776***	2.010***	1.829***	1.831***	1.784***	1.772***	1.254
	(0.449)	(0.596)	(0.464)	(0.596)	(0.475)	(0.595)	(0.440)	(0.436)	(0.453)	(0.454)	(0.851)
HHI	1.599**	1.756*	1.359*	2.277***	1.392*	2.662***	1.669**	1.539**	1.687**	2.052**	2.495**
	(0.723)	(0.915)	(0.753)	(0.864)	(0.784)	(0.808)	(0.702)	(0.702)	(0.716)	(0.815)	(1.250)
Fluid	0.018	0.007	0.022	0.010	0.035	0.010	0.018	0.013	0.013	0.032	0.065
	(0.055)	(0.066)	(0.059)	(0.069)	(0.067)	(0.064)	(0.052)	(0.054)	(0.056)	(0.061)	(0.080)
ProdSim	0.019	0.018	0.017	0.008	0.007	0.023	0.016	0.017	0.025	0.009	0.032
	(0.026)	(0.027)	(0.029)	(0.035)	(0.023)	(0.036)	(0.025)	(0.025)	(0.025)	(0.029)	(0.038)
NewsPr	0.689	0.702	0.970	0.327	0.581	0.319	0.680	0.689	0.866	0.482	0.732
	(0.581)	(0.560)	(0.658)	(0.628)	(0.588)	(0.620)	(0.575)	(0.583)	(0.695)	(0.597)	(0.732)
Constant	-4.031	-3.951	-4.116	-4.545	-4.998*	-4.366	-3.095	-3.940	-4.144	-3.652	-4.469
	(2.995)	(2.946)	(3.138)	(2.772)	(2.709)	(2.775)	(2.875)	(2.856)	(2.941)	(2.965)	(3.404)
Year, Ind, State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	476	444	443	427	476	426	476	476	476	435	356

Panel B: Predictive Model of Positive Confounding Events

	Positive Confound										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
BS1	0.155**										0.286***
	(0.069)										(0.096)
OverCon		2.328***									2.442***
		(0.387)									(0.634)
EarnMan			0.305								2.411
			(2.688)								(4.740)
CEOEq				0.039							-0.031
				(0.024)							(0.035)
CEOTen					0.189***						0.164***
					(0.039)						(0.051)
CEOEqComp						0.492					4.093
						(2.273)					(3.602)
BoardInd							-0.755				-2.290
							(0.868)				(1.642)
ManEnt								-0.319			-0.804*

								(0.352)			(0.472)
PriorMis									1.978***		2.031**
									(0.676)		(0.933)
RiskComm										0.212	0.576
										(0.346)	(0.540)
Size	0.068	0.072	0.067	0.358***	0.199*	0.316***	0.056	0.072	0.039	0.071	0.040
	(0.083)	(0.107)	(0.084)	(0.104)	(0.113)	(0.100)	(0.085)	(0.082)	(0.083)	(0.091)	(0.196)
TobinQ	0.125*	0.046	0.120	0.075	0.135*	0.087	0.134*	0.133*	0.145**	0.170**	0.168
	(0.073)	(0.089)	(0.076)	(0.093)	(0.076)	(0.091)	(0.072)	(0.072)	(0.071)	(0.083)	(0.133)
Leverage	-0.007	0.023	-0.015	0.430	0.333	0.381	0.030	0.064	0.226	-0.134	-0.754
	(0.657)	(0.816)	(0.705)	(0.778)	(0.830)	(0.767)	(0.639)	(0.638)	(0.657)	(0.691)	(1.547)
ROA	-0.124	-2.755	-1.416	0.239	-1.897	0.688	-0.785	-0.748	-0.569	-1.658	-3.729
	(2.071)	(2.591)	(2.267)	(2.849)	(2.179)	(2.743)	(1.990)	(1.973)	(1.968)	(2.170)	(4.153)
Cash	1.303	1.534	1.064	0.971	1.804	0.587	1.261	1.367	1.614	1.186	-1.752
	(1.445)	(1.749)	(1.486)	(1.712)	(1.677)	(1.727)	(1.459)	(1.443)	(1.495)	(1.660)	(2.929)
R&D	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HighTech	1.380***	1.801***	1.339***	1.502***	1.534***	1.520***	1.443***	1.445***	1.399***	1.527***	1.560*
	(0.425)	(0.603)	(0.431)	(0.494)	(0.437)	(0.488)	(0.424)	(0.424)	(0.443)	(0.424)	(0.810)
HHI	1.084	1.038	0.927	1.419	0.673	1.739**	1.095	1.031	1.113	1.194	0.707
	(0.743)	(0.936)	(0.785)	(0.928)	(0.903)	(0.878)	(0.752)	(0.753)	(0.776)	(0.813)	(1.462)
Fluid	0.001	-0.020	0.011	0.015	0.012	0.015	-0.001	-0.004	-0.003	-0.018	0.004
	(0.056)	(0.070)	(0.058)	(0.068)	(0.077)	(0.066)	(0.054)	(0.055)	(0.057)	(0.060)	(0.088)
ProdSim	0.025	0.023	0.019	0.020	0.006	0.034	0.022	0.022	0.030	0.020	0.039
	(0.026)	(0.031)	(0.027)	(0.035)	(0.024)	(0.034)	(0.025)	(0.025)	(0.024)	(0.029)	(0.039)
NewsPr	0.404	0.484	0.589	0.171	0.189	0.145	0.423	0.419	0.568	0.270	0.777
	(0.588)	(0.619)	(0.660)	(0.599)	(0.616)	(0.595)	(0.591)	(0.591)	(0.645)	(0.624)	(0.779)
Constant	-1.444	-0.408	-1.698	-2.319	-3.488	-2.191	-0.874	-1.450	-1.488	-0.335	1.842
	(2.957)	(2.846)	(2.959)	(2.977)	(2.942)	(3.060)	(2.848)	(2.798)	(2.854)	(2.777)	(3.250)
Year, Ind, State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	476	441	444	426	476	425	476	476	476	435	354

Note: In Panel A, the confounding dummy variable is '*7-Day Confounding*', which refers to confounding events that occur 7 days prior to the data breach announcement. As positive confounding events make up the majority of confounding events (78%), we create a *Positive Confound* dummy to identify positive confounds in Panel B. If the confounding events utilized by the firm's management are positive, it is assigned a value of 1, otherwise, it is assigned a value of 0. The other terms are consistent with Panel A. The default breach severity uses *BS1*. The table reports the raw results of the logit model, with robust clustered standard errors shown in parentheses. Significance levels are denoted as follows: *** for 1% significance level, ** for 5% significance level, * for 10% significance level.

Table 7 The Impact of Confounding on %CARs

	CAR (3)	CAR (5)
Panel A: Average %CARs of 675 breached firms		
322 Confounded breaches	-0.185%	-0.544%*
	(0.208)	(0.308)
353 Unconfounded breaches	0.043%	-0.176%
	(0.147)	(0.225)
Difference in CARs of confounded and unconfounded breaches	-0.277%	-0.368%
	(0.251)	(0.377)
Panel B: Average %CARs of 324 breached firms with an above-median Severity breach		
169 Confounded breaches	-1.573%***	-2.076%***
	(0.262)	(0.421)
155 Unconfounded breaches	0.685%**	0.544%
	(0.241)	(0.363)
Difference in CARs of confounded and unconfounded breaches	-2.258%***	-2.620%***
	(0.358)	(0.560)
Panel C: Average %CARs of 164 breached firms with an upper-quartile Severity breach		
111 Confounded breaches	-1.597%***	-2.080%***
	(0.246)	(0.370)
53 Unconfounded breaches	0.877%*	0.327%
	(0.467)	(0.486)
Difference in CARs of confounded and unconfounded breaches	-2.474%***	-2.407%***
	(0.480)	(0.632)
Panel D: Regression of %CARs on Confound dummy with Severity		
Confound * Severity	-0.593***	-0.976***
	(0.205)	(0.306)
Confound	1.389**	1.798**
	(0.604)	(0.824)
Severity	-0.067	0.176
	(0.172)	(0.259)
Constant	2.846	-3.691
	(2.891)	(4.418)
Controls	Yes	Yes
Year, Ind, State FE	Yes	Yes
Observations	526	526
R^2	0.277	0.328

Note: Panel A depicts the average *%CAR(t)* of all 675 breaches of public companies over t days surrounding the breach with t = 3 and 5. Panel B depicts the average *%CAR* of breaches of above-median *Severity*. Panel C depicts the average *%CAR* of breaches of upper-quartile *Severity*. The default severity still uses *BS1*. Standard errors of observing an average *%CAR* or a difference in the average *%CAR* of equal or greater absolute value to that documented are shown in parentheses. Panel D reports regressions of the *%CARs* of breached firms on the interaction term of confound dummy with breach severity. The default confound dummy uses *7-Day Confounding*, and the default breach severity uses *BS1*. Standard errors are robust clustered at the firm level and are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels.

Table 8 Long-run Abnormal Returns

Panel A: Monthly portfolio returns

	N	Mean	Median	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Min	Max
Portfolio A: Breached confounded (long) – breached unconfounded (short)										
Equal-weighted portfolio return	209	-0.371	0.047	-5.874	-2.101	0.047	1.620	4.433	-15.373	10.743
Value-weighted portfolio return	209	-0.097	-0.058	-4.857	-2.050	-0.058	1.895	5.422	-12.956	10.392
Portfolio B: Breached confounded (long) – breached unconfounded with above-median severity (short)										
Equal-weighted portfolio return	209	-0.348	-0.083	-6.204	-2.073	-0.083	1.525	4.515	-15.743	11.218
Value-weighted portfolio return	209	-0.020	0.006	-4.870	-1.876	0.006	1.667	5.817	-11.956	9.487
Portfolio C: Breached confounded (long) – breached unconfounded with upper-quartile severity (short)										
Equal-weighted portfolio return	209	-0.385	0.028	-6.217	-2.026	0.028	1.545	4.809	-14.325	11.003
Value-weighted portfolio return	209	-0.046	-0.008	-4.783	-1.959	-0.008	1.741	5.437	-11.182	9.334

Panel B Fama-French three-factor alpha

	Excess Portfolio returns					
	Portfolio A		Portfolio B		Portfolio C	
	EW	VW	EW	VW	EW	VW
MKT	-0.078	-0.287***	-0.068	-0.237***	-0.066	-0.227***
	(0.056)	(0.059)	(0.055)	(0.057)	(0.057)	(0.058)
SMB	-0.061	0.255**	-0.029	0.276***	-0.013	0.291***
	(0.118)	(0.105)	(0.119)	(0.099)	(0.120)	(0.097)
HML	-0.081	-0.088	-0.060	-0.080	-0.013	-0.016
	(0.085)	(0.082)	(0.086)	(0.082)	(0.088)	(0.083)
Alphas	-0.406	0.016	-0.391	0.055	-0.427*	0.024
	(0.246)	(0.220)	(0.247)	(0.213)	(0.249)	(0.214)
Observations	209	209	209	209	209	209
R^2	0.023	0.153	0.014	0.124	0.008	0.109

Panel C Fama-French five-factor alpha

	Excess portfolio returns					
	Portfolio A		Portfolio B		Portfolio C	
	EW	VW	EW	VW	EW	VW
MKT	-0.077	-0.271***	-0.066	-0.227***	-0.067	-0.221***
	(0.056)	(0.058)	(0.056)	(0.056)	(0.058)	(0.057)
SMB	-0.052	0.347***	-0.028	0.356***	-0.004	0.366***
	(0.141)	(0.126)	(0.143)	(0.121)	(0.142)	(0.118)
HML	-0.085	-0.240**	-0.064	-0.207*	-0.018	-0.134
	(0.115)	(0.112)	(0.116)	(0.112)	(0.117)	(0.112)
RMW	-0.013	0.204	-0.021	0.193	0.003	0.166
	(0.171)	(0.163)	(0.172)	(0.153)	(0.170)	(0.153)
CMA	0.033	0.241	0.025	0.171	0.014	0.142
	(0.179)	(0.191)	(0.180)	(0.183)	(0.178)	(0.184)
Alphas	-0.408	-0.112	-0.389	-0.053	-0.430*	-0.067
	(0.252)	(0.231)	(0.252)	(0.224)	(0.254)	(0.226)
Observations	209	209	209	209	209	209
R^2	0.023	0.182	0.014	0.147	0.008	0.130

Note: Each month from January 2005 to December 2022, we construct three portfolios based on firms that reported data breaches in the past 12 months comparing the returns of confounded vs. unconfounded breach firms. (A) contrasts confounded vs. unconfounded breaches, (B) restricts unconfounded breaches to those with above-median severity, and (C) to those with upper-quartile severity. Panel A reports monthly portfolio returns, while Panels B and C present OLS regression results, with equal-weighted (EW) and value-weighted (VW) portfolio returns in excess of the risk-free rate as the dependent variables. Independent variables include MKT, SMB, and HML from Fama and French (1993) and the extended five-factor model (MKT, SMB, HML, RMW, and CMA) from Fama and French (2015). Robust standard errors clustered at the calendar month level are reported in parentheses. q***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively

Table 9 Heckman Two Stage Model

	Step 1	Step 2	
	Confound	CAR (3)	CAR (5)
Confound * BS1		-0.562***	-0.938***
		(0.170)	(0.216)
Confound		1.643***	2.679***
		(0.535)	(0.738)
BS1	0.007	-0.024	0.115
	(0.013)	(0.145)	(0.174)
PrevBreach	0.142**		
	(0.060)		
MultiAnn	0.214***		
	(0.059)		
IndAtt	0.395*		
	(0.208)		
IMR		0.522	0.425
		(0.326)	(0.416)
Constant		-0.839	-3.260
		(3.232)	(4.337)
Controls	Yes	Yes	Yes
Year, Ind, State FE	Yes	Yes	Yes
Observations	380	380	380
R^2/Pseudo R^2	0.325	0.345	0.423

Note: This table displays the regression results of the standard Heckman (1979) two-stage model. This method is used to effectively isolate the estimation bias induced by the selection of confounded firms. The first-stage selection equation is a Probit model with the *Confound* dummy as the explanatory variable, assigned a value of 1 if the data breach event is confounded, and 0 otherwise. The explanatory variables include exogenous factors such as managerial and data breach characteristics. The instrumental variables (IVs) we use, including Previous breaches (*PrevBreach*), Multiple announcements (*MultiAnn*), and Industry attention (*IndAtt)*, satisfy the relevance criterion and exclusion restriction. The definitions of these IVs can be found in the Appendix A. In the second-stage regression, the explained variables are *CAR (3)* and *CAR (5)*). The Inverse Mills ratio (*IMR*) derived from the first-stage regression is also included as an explanatory variable. The default confound dummy uses *7-Day Confounding*, and the default breach severity is *BS1*. Standard errors are robust clustered at the firm level and are shown in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 10 The Predictive Model of Financial Information Breaches on Confounding

	Confound		
FinInfo * Severity			0.998***
			(0.180)
FinInfo	-0.107	-0.223	-2.919***
	(0.271)	(0.272)	(0.556)
BS1		0.134*	-0.367***
		(0.071)	(0.129)
Size	0.071	0.068	0.054
	(0.081)	(0.082)	(0.087)
TobinQ	0.133*	0.129*	0.100
	(0.073)	(0.075)	(0.084)
Leverage	0.476	0.434	0.373
	(0.671)	(0.689)	(0.698)
ROA	-0.676	-0.303	-0.488
	(1.960)	(2.016)	(1.963)
Cash	1.060	1.066	1.317
	(1.518)	(1.535)	(1.769)
R&D	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
HighTech	1.930***	1.877***	1.761***
	(0.445)	(0.453)	(0.537)
HHI	1.081	1.176	1.350*
	(0.724)	(0.736)	(0.813)
Fluid	0.011	0.012	0.067
	(0.053)	(0.054)	(0.056)
ProdSim	0.017	0.020	0.016
	(0.023)	(0.024)	(0.024)
NewsPr	0.121	0.113	0.066
	(0.221)	(0.220)	(0.233)
Constant	-2.823	-2.874	-3.299
	(2.859)	(2.954)	(3.047)
Year, Ind, State FE	Yes	Yes	Yes
Observations	460	460	460

Note: This table discusses the impact of data breach information types on managerial confounding decisions and the moderating role of breach severity. We classify breach types involving financial access (e.g., bank account credentials, credit card data), identity theft (information enabling impersonation), existential data (information critical to national security or business continuity), and other financial information breaches as financial information loss. The dummy variable FinInfo is set to 1 for these cases and 0 otherwise (Kamiya et al., 2021). The confounding dummy variable is '*7-Day Confounding*', which refers to confounding events that occur 7 days prior to the data breach announcement. The default breach severity uses *BS1*. The table reports the raw results of the logit model, with robust clustered standard errors shown in parentheses. Significance levels are denoted as follows: *** for 1% significance level, ** for 5% significance level, * for 10% significance level.

Table 11 Regression of %CARs on Confounding Types of Managerial Response and Breach Severity

	CAR (3)			CAR (5)		
Anticipatory * Severity	-0.280			-0.609**		
	(0.180)			(0.273)		
Anticipatory	0.374			0.921		
	(0.709)			(0.989)		
Same-day * Severity		-0.675**			-0.842**	
		(0.298)			(0.416)	
Same-day		2.466**			2.179	
		(1.218)			(1.601)	
Reactive * Severity			0.846			2.216
			(0.543)			(1.351)
Reactive			-1.380			-2.841
			(1.372)			(2.510)
BS1	-0.289*	-0.358***	-0.488***	-0.122	-0.351**	-0.563***
	(0.150)	(0.109)	(0.108)	(0.229)	(0.166)	(0.154)
Size	-0.059	-0.103	-0.095	0.334	0.274	0.243
	(0.166)	(0.165)	(0.163)	(0.243)	(0.241)	(0.232)
TobinQ	0.030	0.051	0.022	0.082	0.136	0.066
	(0.117)	(0.119)	(0.120)	(0.212)	(0.201)	(0.213)
Leverage	-0.101	-0.183	-0.161	0.548	0.458	0.396
	(0.799)	(0.809)	(0.799)	(1.127)	(1.148)	(1.079)
ROA	-0.890	-0.592	-0.781	-4.088	-3.876	-3.756
	(2.670)	(2.781)	(2.711)	(5.227)	(5.246)	(5.396)
Cash	-0.952	-1.333	-0.943	-3.291	-4.066	-3.317
	(2.017)	(2.000)	(2.061)	(3.388)	(3.343)	(3.525)
R&D	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HighTech	0.095	0.042	0.141	-0.071	-0.048	0.129
	(0.752)	(0.710)	(0.750)	(1.090)	(1.046)	(1.101)
HHI	-0.723	-0.776	-0.816	-0.006	-0.006	-0.093
	(0.973)	(0.970)	(0.986)	(1.471)	(1.489)	(1.512)
Fluid	-0.029	-0.055	-0.023	-0.002	-0.029	0.014
	(0.089)	(0.096)	(0.091)	(0.134)	(0.142)	(0.137)
ProdSim	0.036	0.044	0.036	-0.019	-0.011	-0.014
	(0.040)	(0.040)	(0.040)	(0.075)	(0.075)	(0.076)
NewsPr	1.087	0.863	0.888	3.952**	3.559**	3.437**
	(1.089)	(1.105)	(1.087)	(1.742)	(1.721)	(1.661)
Constant	3.047	4.178	4.064	-3.484	-1.787	-0.730
	(2.835)	(2.916)	(2.772)	(4.486)	(4.445)	(4.126)
Year, Ind, State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	526	526	526	526	526	526
R^2	0.264	0.272	0.266	0.315	0.311	0.332

Note: This table reports the effect of different confounding types on firms' short-term returns, moderated by breach severity, with cumulative abnormal returns assessed over 3-day and 5-day windows. Confounding types are classified by managerial response timing: within the 61-day window around the breach announcement (30 days before and after), a confounding event announced before the breach is classified as anticipatory impression management (IM), indicating management's advance preparation; an event on the same day as the breach is classified as *Same-day*; and an event announced after the breach is reactive IM, indicating management's response post-announcement. Default breach severity is *BS1*. Robust standard errors are clustered at the firm level and shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

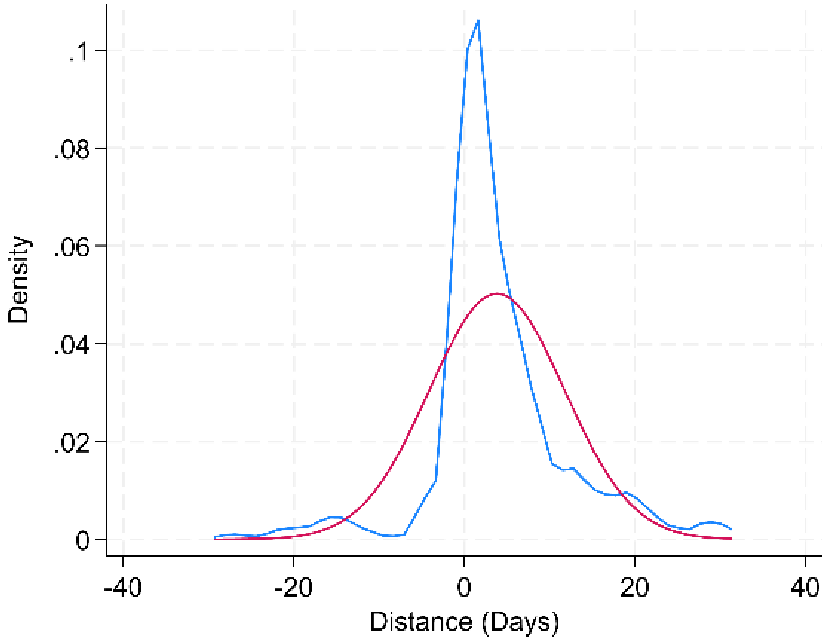

Figure 1 Kernel Density of Breach–Confound Distance (Days)

Note: Distance = breach announcement date − nearest confounding firm-specific announcement (±30 days). Blue = kernel density estimate (Epanechnikov, bandwidth 1.3849); red = normal density. The density peaks at day 0 and concentrates within ±3 days, indicating non-random clustering around the breach announcement date.

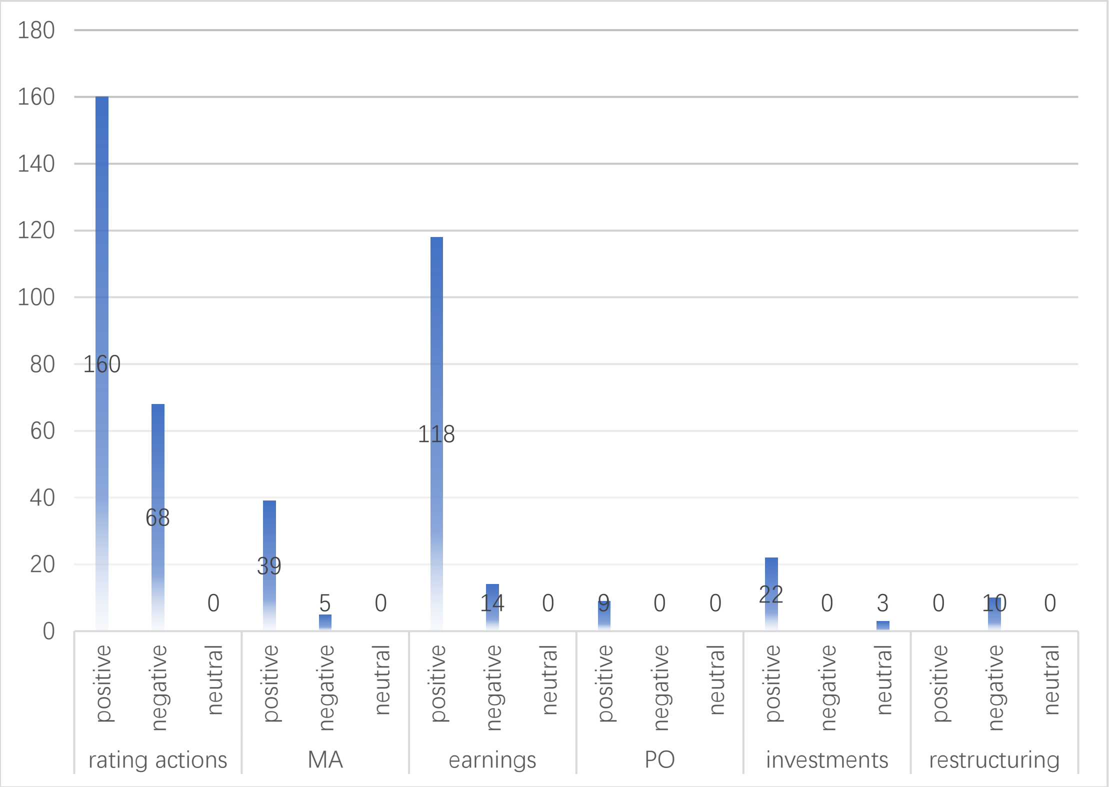

Figure 2 60-day Confounding Events: Positive Offset and 'Big Bath'