Inside Out:
Who Trade Before the *Start* of Cyber Attacks?*

Xi Dong†

Edward Xuejun Li‡

Xintian Lin§

Xin Yuan**

First Version: May 1, 2025
This Version: March 15, 2026

* We would like to thank Musaib Ashraf (discussant), Arthur Beddock (discussant), Frederico Belo, Svetlana Bryzgalova, Thomas Ernst (discussant), Eliezer Fich, In-Mu Haw, Grace Xing Hu, Amy Hutton, Wei Jiang, Jonathan Karpoff, Timothy Loughran, Charles Martineau, Andriy Shkilko (discussant), Joel Peress, and seminar participants at the 2025 JFQA & Future of Financial Information Conference, 35th Annual Conference on Financial Economics and Accounting (CFEA), 2025 SFS Cavalcade Asia-Pacific, 2026 Midwest Finance Annual Meeting, INSEAD, City University of New York-Baruch College, Dongbei University of Finance and Economics, Hong Kong Baptist University, and Southwestern University of Finance and Economics for comments and suggestions.
† Associate Professor, Zicklin School of Business, Baruch College. E-mail: Xi.Dong@baruch.cuny.edu.
‡ Professor, Baruch College and Columbia University. E-mail: Edward.Li@baruch.cuny.edu.
§ Assistant Professor, Central University of Finance and Economics. E-mail: lxt@cufe.edu.cn.
** Assistant Professor, Dongbei University of Finance and Economics. E-mail: xinyuan@dufe.edu.cn.

Inside Out:
Who Trade Before the *Start* of Cyber Attacks?

Abstract

Using a unique dataset, we provide the first comprehensive evidence of informed trading before successful cyberattacks. Although only hackers should expect an imminent breach, short selling in victim firms intensifies a few weeks earlier, whereas insider and institutional trades remain flat. Retail investors, ostensibly the least informed, likewise presciently divest/short these stocks, coinciding with spikes in "<company-name>+hacking" Google queries and rising trading costs. Post-attack, victim firms experience negative returns, suggesting a wealth transfer of a magnitude exceeding widely-publicized ransom demands. Short selling also rises ahead of some notable non-cyber outsider-initiated attacks. Collectively, our findings suggest that cyberattacks—the tip of the iceberg of a broader class of outsider-initiated events—challenge the traditional information-economics paradigm in which outsiders merely learn about exogenous states of the world, rather than also being entangled with the formation of price-relevant states that generate trading profits.

JEL Classification: G12, G14, K24

Keywords: Data Breach, Cyberattack, Outsider-Generated Information, Information Asymmetry, Short Selling, Retail

1. Introduction

Information in financial markets is traditionally modeled as concerning price-relevant states that already exist but are observed earlier by some agents—typically insiders—than by others. In canonical frameworks such as Grossman and Stiglitz (1980), Kyle (195), and Glosten and Milgrom (1985), informed traders possess superior information about exogenous fundamentals, while the rest of the market learns from prices, disclosures, or order flow. Prices adjust as private information is incorporated into the market, and adverse selection arises because market makers face the risk of transacting against better-informed traders. This traditional information structure underlies central results on information production, market efficiency, disclosure design, and welfare.

In this paper, we document a reversal of this traditional informational logic in a specific but increasingly relevant setting. We characterize this as an "inside-out" information structure: outsiders, rather than insiders within the firm, possess and trade on private information tied to events that alter firm fundamentals before the firm insiders fully recognize the change. This structure departs from canonical models in a deeper sense as well. In Grossman-Stiglitz type environments, informed traders acquire signals about exogenous states of the world, and the resulting price discovery is typically viewed as socially valuable because it improves allocative efficiency and risk sharing. In our setting, by contrast, private information is entangled with outsider-initiated actions that can damage firm value, as in cyberattacks. As a result, greater price efficiency need not reflect socially beneficial information production; it may instead arise alongside the change of firm fundamentals themselves. More broadly, our setting sits outside the scope of standard information-based frameworks, which typically model informed agents as learning about exogenous states of the world rather than facing environments in which outsider actions may be entangled with the emergence of price-relevant states.

Our setting also complicates the intuitions underlying both canonical trading models and prompt-disclosure regulation. In Kyle-type frameworks, informed trading is smoothed over time, so adverse selection appears as a relatively continuous cost of supplying liquidity. A similar logic underlies disclosure policy: faster public release of material information is expected to reduce persistent informational advantages and curb informed rents. In outsider-initiated settings such as cyberattacks, however, the relevant information event is privately timed by outsiders yet largely unanticipated by the market. This creates incentives for concentrated pre-event trading rather than gradual information incorporation. Prompt disclosure may then have a more ambiguous effect. Although it reduces ex post trading on stale information, it can also raise the ex ante payoff to early information by allowing outsiders to monetize the induced price decline more quickly and with less unrelated market risk.[2] More broadly, our setting highlights a distinct class of informational frictions not clearly contemplated by existing theory or regulatory frameworks, calling for a rethinking of what counts as material information, who controls it, and how regulation should adapt when private information is tied to outsider-initiated changes in fundamentals rather than merely to the gradual incorporation of exogenous fundamental information.

While our conceptual framework has broader relevance, we focus on corporate cyberattacks as our primary empirical setting. Cyberattacks provide a clean and economically important setting in which to isolate an "inside-out" information structure.[4] The direction of information flow is

[2]As anecdotal evidence suggests, ransomware gangs have reportedly filed SEC whistleblower complaints against victim firms for delayed disclosure, implicitly threatening further market damage unless paid (see, e.g., the ALPHV/BlackCat ransomware operation).

[4] The focus on cyberattacks is motivated by three factors. First, the economic magnitude is substantial: the International Monetary Fund (IMF) identifies cyberattacks as a systemic threat to financial stability (IMF, Finance & Development, "The Global Cyber Threat to Financial Systems"), and the World Economic Forum (Global Risks Report 2023) estimates annual cyber-related losses exceeding $10 trillion. Second, regulatory scrutiny has intensified, exemplified by the New York Department of Financial Services (NYDFS) Cybersecurity Regulation and the European Union's General Data Protection Regulation (GDPR). Third, technological advancements—particularly in artificial intelligence and cloud computing—structurally elevate these risks by migrating critical data outside traditional firm boundaries, thereby expanding the surface area for outside attacks.

unambiguous: attacks originate outside the firm and impair real fundamentals, yet management often detects them only with a significant delay (Kamiya, Kang, Kim, Milidonis, and Stulz, 2021; Huang and Wang, 2021). This sharply departs from standard information-based models. Under canonical assumptions, informed trading should not arise before the onset of an attack, because insiders cannot trade on a breach of which they are unaware, despite the fact that the breach will later generate a price-relevant shock, while outsiders are not typically modeled as possessing private information tied to outsider-initiated changes in firm value. Although cyberattacks on public firms are themselves of central importance in an increasingly digitalized economy, we also supplement our main analysis with suggestive case studies of notable non-cyber outsider-initiated events. Taken together, our evidence suggests that cyberattacks may be only the tip of the iceberg of a broader inside-out information structure.

To test this prediction, we hand-collect a comprehensive sample of 397 cyber breaches at U.S. public firms spanning 2007–2018. For each breach, we identify three distinct dates: (i) the attack start date, when unauthorized access begins; (ii) the detection date, when the firm first becomes aware; and (iii) the public disclosure date. This structure allows us to isolate trading that occurs before any plausible insider knowledge. We analyze trading by four distinct groups: short sellers, retail investors, institutional investors, and corporate insiders.

Our evidence suggests substantial informed trading before attack onset. Short interest in victim firms begins to rise approximately six weeks before the breach, peaking in the one- to three-week window preceding the successful attack. These increases are concentrated among firms with more diverse share-lender bases and are absent in counterfactual settings—including industry peers with comparable *ex ante* cyber risk, minor breaches, and placebo time windows. Decomposing short interest reveals a sharp increase in short-sale volume—a more direct proxy for information-

motivated shorting (Blocher, Dong, Ringgenberg, and Savor, 2023)—that is strictly confined to the pre-attack window and dissipate immediately afterward.

Retail investors exhibit similar anticipatory behavior. Despite being typically characterized as uninformed noise traders (Barber and Odean, 2000), retail investors reduce long positions and increase shorting activity before attacks. Using the algorithm of Boehmer and Song (2020), we find significant increases in retail short-sale volume, consistent with information motivation rather than hedging. These patterns coincide with spikes in online search intensity combining victim-firm names with cybersecurity-related terms (Da, Engelberg, and Gao, 2011), suggesting that some outsiders receive early private signals and seek corroboration through digital channels. In contrast, corporate insiders (Cohen, Malloy, and Pomorski, 2012) and institutional investors (Bushee, Matsumoto, and Miller, 2003; Solomon and Soltes, 2015) exhibit no abnormal pre-attack informed trading. This null result suggests that the informational advantage lies entirely outside traditional corporate or institutional boundaries.

The economic implications are significant. Large breaches are associated with cumulative abnormal returns of approximately -3% in 30 days following the attack onset (consistent with Kamiya et al., 2021; Michel, Oded, and Shaked, 2020). Combining these price declines with the pre-attack buildup in short interest implies a substantial wealth transfer from uninformed investors to those establishing short positions prior to the attack. Across the 109 large breaches in our sample, the estimated aggregate wealth transfer approaches $324 million. While this represents a conservative lower bound, it exceeds the widely publicized ransom payments (typically $1–2 million per victim firm) by orders of magnitude. For the subset of incidents with the more precise pre-attack shorting, the average transfer per event exceeds $12 million. Furthermore, market quality deteriorates markedly around attack onset: relative spreads, shorting costs, and price

impacts all rise. This pattern indicates that liquidity providers bear elevated adverse-selection costs around the event window even in the absence of insider-originated information. In this setting, counterparties are exposed not only to the subsequent value destruction associated with the outsider-initiated shock itself, but also to short-run losses from trading against better-informed outsiders whose orders are concentrated in a narrow pre-event window rather than smoothed over time.

Although our empirical identification relies on corporate cyberattacks, the underlying economic mechanism—an "inside-out" information structure—may apply more broadly to settings in which outsider actions are entangled with the formation of price-relevant states that generate trading opportunities. These settings include operational disruptions, reputational crises, externally generated fundamental shocks, and even events whose realization can be traded on in prediction markets.[5] The common intuition across them is that once outside traders can profit from an outcome, society must confront the possibility that incentives need not operate solely through forecasting exogenous states, but may also become entangled with actions that contribute to the realization of those states. To illustrate this broader relevance, we present three high-profile case studies. First, in operational disruptions, the March 2024 shutdown of Tesla's German Gigafactory by environmental activists erased over $120 billion in market capitalization. Second, in legal and reputational shocks, the April 2019 civil suit against JD.com's founder removed approximately $2.8 billion in shareholder value. Third, in externally generated fundamental shocks, DeepSeek's AI releases in December 2024 and January 2025 preceded a 17 percent decline in Nvidia's stock and over $580 billion in lost market value. In each case, the shock originated outside the focal firm, materially affected valuation, and was preceded by rising short interest in the prior weeks.

[5] Recent controversy over growing prediction markets tied to war or catastrophic geopolitical events share the same underlying economic logic.

Taken together, our evidence suggests that informational hierarchies in modern markets need not follow the conventional insider-centric paradigm. In settings where price-relevant shocks originate outside the firm, private information may arise and be traded on before either management or insiders are fully aware of the change. Our findings suggest a structural gap in market design and standard frameworks for disclosure, market monitoring, and informed trading. In particular, the notion of "material" information often presupposes an internal source, disclosure duties become harder to assign when firms themselves are uninformed, and monitoring mechanisms calibrated to corporate insiders may miss the actual source of informed trading. In this sense, cyberattacks provide proof-of-concept evidence that market-integrity frameworks may require recalibration in environments shaped by outsider-initiated, price-relevant shocks.

Literature and Contribution

We contribute to the nascent literature on the evolving firm information environment in the digital age. Recent work shows that externally generated "big data" sources erode insiders' informational edge (Cao, Jiang, Wang, and Yang 2024). This outsider-beats-insider evidence extends the earlier feedback literature where managers glean decision-making cues from stock prices (Chen, Goldstein, and Jiang, 2007; Edmans, Goldstein, and Jiang, 2015). We push this inversion of information asymmetry further by introducing a new setting: not only are outsiders discovering data, but some of them are actively creating shocks that change firm real value—e.g., events like material cash-flow damage from cyberattacks. The perpetrators or others in their informational orbit (e.g., via networks, social media, or the Dark Web) can profit by front-running the markets. Insiders, by contrast, lack prior knowledge of these outsider-generated shocks. Unlike prior studies on cyberattacks that focus on announcement-period reactions (see footnote 5), we depart from this conventional path by being the first to analyze pre-attack trading. Demonstrating that the

informational center of gravity can shift outside the firm has important implications for the information economics literature in several ways discussed below.

First, our evidence calls for a framework beyond the canonical Grossman–Stiglitz (1980) model. In a GS-type model, information acquisition is costly but purely extractive—it never changes the firm's cash flows. The persistence of "noise" in prices ensures a sufficient incentive to produce socially valuable information. In contrast, outsider attacks such as cyberattacks simultaneously generate private information and alter (often damage) real firm fundamentals. This change upends the GS logic. Here, outsiders' incentives are aligned with both information precision and real impact—larger breaches yield clearer signals and greater profits. Information production is intertwined with real value change (e.g., destruction), so a partial-equilibrium GS setup no longer suffices. The marginal dollars spent on hacking not only sharpen private signals but also worsen fundamentals, exacerbating adverse selection. The resulting arms race yields no welfare gain from sharper pricing; it is pure rent-seeking that transfers wealth away from the uninformed, especially when firm value destruction is the largest—precisely the massive breaches with big ex-post realized returns that we document. Ironically, these are also the hardest cases for prosecutors: Large-cap stocks—often the target of these attacks—mask malicious trading amid heavy volume, making detection and prosecution exceptionally difficult, particularly when share lending is diffuse.

Second, the logic of disclosure—for regulators and for firms—changes once outsider-generated attacks become more important. For regulators, rules that prioritize rapid, detailed disclosure are designed to protect uninformed outsiders, yet they may unintentionally help attacker–traders by giving them a clean exit and encouraging future hacks. A staggered approach—e.g., releasing partial, redacted bulletins—could blunt that monetization. Early but vague

disclosure turns a hacker's pure-alpha signal into a noisy, risk-bearing one: prices drift, spreads widen, and volatility spikes. The attacker must now bear the mark-to-market noise and higher trading costs, cutting the expected payoff and thus the incentive to attack. As shown by Hu, Pan, and Wang (2017), tiered information releases can shrink, rather than widen, high-speed traders' informational edge. For firms, the classic voluntary disclosure theory (e.g., Verrecchia, 1983) suggests that managers accelerate good news and delay bad news. With an inverted information hierarchy, the intuition flips: insiders might instead disclose *expected* weaknesses *before attacks materialize*—"pre-emptive confession"—to neutralize outsiders' trading advantage, effectively turning disclosure policies into a form of information armor.

Third, our findings provide insights for the market microstructure theory to evolve beyond its insider-centric foundations. In seminal frameworks (Glosten-Milgrom, 1985; Kyle, 1985), dealers set spreads to offset the risk of trading against a gradual and/or relatively predictable set of insiders. Regulatory tools—10b5-1 plans, blackout periods, and Form 4 filings—further constrain the speed and size of insider trades. In contrast, outsider attacks like cyberattacks introduce an exogenous, jumpy source of informed order flow. The informational edge now lies with loosely coordinated outsider networks, potentially numbering in the hundreds, which act en masse based on real-time digital cues (Brugler and Comerton-Forde, 2025). These trades arrive in bursts, often at random times, and are concentrated in left-tail events, altering both the magnitude and distribution of dealer risk. The traditional spread-setting logic, featuring smoothed and/or predictable informed order flows, fails to accommodate this new environment. Our evidence supports this updated view. Abnormal short selling and retail activity spikes are confined to the event window but absent on pseudo-attack dates. This timing suggests that trades are driven by the attacks themselves, rather than broader cybersecurity vulnerability concerns. Moreover, the downside tail risk borne by

market makers is more severe when trading is concentrated around large, value-destroying events identified *ex-post* by examining realized outcomes after the attacks. These characteristics call for revised microstructure models that accommodate large, sudden, unnecessarily endogenous, and externally driven shocks as well as broader sets of informed actors. Market intermediaries need to adapt not only to new types of information but also to the new mechanisms of information production and diffusion.

Fourth, our findings offer a micro-foundation for the emerging evidence on cyber-risk premium. Florackis et al. (2023) document that firms with weaker cyber defenses post higher beta-adjusted returns even if they have never been hacked. The classic asset-pricing theory (e.g., Easley and O'Hara, 2004) links expected returns to cash-flow volatility and insider informational advantage. Our findings suggest a related, third wedge—adverse selection arising from informed outsider trading. Markets thus price not only higher β-risk but also fatter left tails (skewness and jump risk) as investors brace for sudden, value-destroying shocks that arise externally. A single high-profile, outsider-informed attack can trigger repricing cyber risk across the sector, elevating required returns on all cyber-vulnerable firms, regardless of their breach history.

Finally, we recast the debate on the role of short sellers in society by showing they trade on material, outsider-generated fundamentals—not ephemeral rumors. The literature is split between two archetypes. The price-discovery view casts shorts as information hunters who align prices with pre-existing fundamentals (Miller 1977; Diamond and Verrecchia, 1987; Christophe et al. 2004; Diether et al. 2009; Karpoff and Lou 2010), whereas the predatory view depicts them as rumor-mongers who inject noise and volatility (Allen and Gale, 1992; Goldstein and Guembel, 2008; Brunnermeier and Oehmke, 2014). Both literatures assume the news they exploit is either pre-

existing or already known to insiders.[6] Our evidence reveals a third channel: short sellers swarm before cyber-breach disclosures, betting on real value destruction that insiders have not yet seen. The burst of shorting is tightly confined to the days just prior to the attack and disappears immediately afterward, ruling out long-term fundamental pessimism or short-term self-reversing rumor trades. Hence these traders are neither mere traders of pre-existing fundamental information nor noise creators. Instead, they are early movers on outsider-created, value-destroying shocks that even insiders have yet to know. This bridges the two camps and redirects attention to a new class of policy-relevant information that is outsider-engineered and triggers real—rather than purely noisy—volatility.

Moreover, our analysis offers a unified view of trading behaviors across key market participants—short sellers, insiders, institutions, and retail investors. Examining all four sharpens the narrative by highlighting distinct informational dynamics. While we observe pronounced pre-attack activities among short sellers and retail investors, we find no comparable trading patterns among insiders or institutions. The asymmetry underscores the external nature of the information: it is neither possessed by insiders nor readily inferred from public signals by sophisticated institutions. To sharpen the analysis, we refine the identification of retail short selling using Boehmer and Song's (2020) methodology, revealing that retail traders—not traditionally viewed as informed—may respond to private cues circulating through unconventional channels such as social media or personal networks.

These findings are particularly surprising given the prevailing academic view of retail investors as less sophisticated (e.g., Barber and Odean, 2000). Our evidence suggests a more nuanced interpretation. In contexts where insiders lack an informational advantage—such as outsider

[6] Engelberg, Reed, and Ringgenberg (2012) show that a substantial portion of short sellers' return predictability comes from their ability to analyze publicly available information such as corporate disclosures.

attacks—retail traders may be better positioned to act on emergent, outsider-generated signals. This challenges assumptions about the fixed informational hierarchy in financial markets and opens new avenues for research into how digital ecosystems distribute value-relevant information outside traditional corporate or institutional channels.

2. Data and Variable Definitions

We hand-collect data from multiple sources to construct a comprehensive dataset of cybersecurity breaches and associated market behavior. Our primary breach incident data are drawn from the California Attorney General Website (Cal), the Privacy Rights Clearinghouse (PRC),[7] and the Audit Analytics Cybersecurity Dataset (AA).[8] To verify and supplement these datasets, we conduct multi-step cleaning procedures to manually search and verify key information for each incident, including event dates, attack type, and breach severity, by triangulating from Google, Factiva, and SEC filings, breach notice letters, legal complaints, and other publicly accessible sources. Importantly, we cross-check incidents across databases, consolidate duplicate cases, and align major event dates. The process is essential because the three databases can have different date definitions or incident descriptions and a firm may experience multiple cyberattacks.[11] Through the hand-collection process, we identify unique cyberattacks affecting the U.S. public

[7] For our use of the PRC database, we focus primarily on breaches classified as "HACK" and involving business types labeled "BSF," "BSO," "BSR," or "MED," which are most indicative of cyberattacks targeting firms or business organizations. These classification criteria align with prior empirical studies in this area.

[8] The California law requires that when a breach notice is issued to more than 500 residents, a copy must also be submitted to the California Attorney General. As a result, this database is likely to capture many major cyber incidents, given that many large and industry-leading firms are headquartered in California and that California residents account for a nontrivial share of the U.S. population. Besides, the Privacy Rights Clearinghouse (PRC) complements the California database by compiling publicly reported data breaches experienced by a variety of organizations across the U.S. By contrast, the Audit Analytics (AA) dataset focuses on the cyber incidents which involve public firms. Therefore, the combination of the three databases summarizes the universe of major data breaches in the U.S.

[11] For example, although the AA dataset provides a "Date of Breach," the field contains missing data, especially in the early years. The California database also reports a breach date, but since the date is self-reported by firms, it can be either the breach start date or breach disclosure date. The PRC database faces the similar issue on date definitions.

firms and construct a comprehensive database that contains major event dates.[12] For trading behavior and market structure variables, we use daily short position and borrowing cost data from the Markit Securities Finance Buyside Analytics (Markit), transaction-level short trading data from the Financial Industry Regulatory Authority (FINRA) Short Sale Volume dataset, high-frequency investor trading data from TAQ, insider trading data from LSEG, and institutional holding and insider holding metrics from prior literature (i.e., Lee and Radhakrishna, 2000; Barber et al., 2024). Market microstructure data (e.g., relative spread, price impact), stock transaction information, and firm financial data are obtained from WRDS Millisecond Intraday Indicators (IID) database, CRSP, and Compustat.

2.1 Sample Selection

We focus on cyberattacks occurring between 2007 and 2018, a period aligning with the span of our proprietary Markit data. Considering the substantial differences among cyber events, we focus on the cybersecurity data breaches that result in an unauthorized access to employee or customer data with the selection criteria listed in the Internet Appendix. Identifying the cyber incidents of similar legal and economic consequences, the selection criteria can create a sample that more cleanly represents the insider-out information structure of cyberattacks. Based on Cal, AA, and PRC, our initial sample includes 527 unique data breach events involving public firms. We retain only events involving data exfiltration,[16] eliminate 130 incidents lacking meaningful data loss, and

[12] We restrict our sample to entities that are publicly traded firms—or subsidiaries or local operating branches of such firms—at the time the cyberattack starts. To identify qualifying incidents, we conduct manual searches using Factiva, Google, and other publicly accessible media sources, querying each victim firm's name alongside the disclosure date and cybersecurity-related keywords. These keywords include "attack," "breach," "hack," "intrude," "compromise," "virus," "malware," "security," and "unauthorized access." The keyword selection is informed by Gordon, Loeb, and Sohail (2010) and a review of Item 1A "Risk Factors" sections in 10-K filings of several affected firms.

[16] We collect information from multiple sources, including incident-related news reports, press releases, and regulatory filings (e.g., SEC submissions, breach notification letters, and legal complaints), to construct a comprehensive context for each event. Based on this information, we classify the cybersecurity incidents into eleven categories, detailed in Internet Appendix IA2. We exclude events that are not attributable to unauthorized cyber intrusions or involve attacks

drop additional 33 breaches due to incomplete coverage in Compustat, CRSP, Markit or LSEG databases. For each remaining breach, we hand-collect and cross-check the accuracy of three key dates: the breach start (when the system is compromised), detection (when the firm identifies it), and disclosure (when the firm publicly discloses it).[17] After removing incidents without verifiable breach start/disclosure dates or with missing trading data around the event window (±45 days), we arrive at a final sample of 262 breaches across 210 public firms.[18]

We define a subsample of 109 "massive" breaches based on attack type (e.g., phishing, central system intrusion) and the absence of earnings announcements in the 30 days post-breach. These cyber events are massive data breaches ("massive data breach sample" or "attack sample") which presumably result in substantial economic losses for the victim firms. Massive breaches affect 89 firms and form the primary focus of our analyses. As shown in Table 1, the full and massive breach samples together represent 28% and 14% of the contemporaneous S&P 500 market capitalization, respectively. Consistent with the prior literature, our samples highlight the economic significance of cyber-related data breaches.

2.2 Data Breach Characteristics

We hand-collect several breach-specific characteristics for use in our empirical analysis. The variable *Breach Duration* captures the number of days between the breach start and firm detection, reflecting the potential window of undetected data exposure. [19] We also construct two return-based

on non-public entities (Types 8–11). We also remove incidents where no customer or employee data were compromised (Types 4–7), resulting in the exclusion of 130 events.

[17] When firms disclose an exact start date for a breach, we use that date directly. If only a breach period is provided, we approximate the start date—e.g., assigning April 1, 2013 when the disclosure indicates "early April 2013"—provided the period does not exceed one month. When neither firm disclosures nor public sources specify a breach date, we rely on the "Date(s) of Breach" variable from the California (Cal) database, which is subject to regulatory oversight. Additional details are provided in Internet Appendix IA3.

[18] We allow incidents to have missing value for Firm Detection Date but require non-missing value for Breach Disclosure Date in order to recognize the significant effects of major disclosure events on cyber-related trading activities (Akey, Grégoire, and Martineau, 2022).

[19] For incidents with missing Firm Detection Date, we use Breach Disclosure Date to calculate Breach Duration.

variables—*Breach Start BHAR* and *Breach Disclosure BHAR*—which measure buy-and-hold abnormal returns over the [0, +30] day window surrounding the breach start and disclosure date, respectively. Abnormal returns are calculated as the DGTW-adjusted buy-and-hold returns by following Daniel, Grinblatt, Titman, and Wermers (1997).[20]

2.3 Short Selling Variables

To measure short-selling activity, we follow the prior research (i.e. Geczy, Musto, and Reed, 2002; Chen, Da, and Huang, 2022) and define short interest as the ratio of shares sold short to total shares outstanding (in percentage terms), using daily equity lending data from Markit. Since we attempt to investigate abnormal short selling around data breaches, we compute abnormal short interest in three steps. First, consistent with previous work (Chen et al., 2018; Wang, Yan, and Zheng, 2020), we detrend the daily short interest by subtracting the 90-day pre-event moving average. The calculation of detrended short interest using a 90-day benchmark window (Day -90 to -1) is shown below.

$$\text{D}etrended\ SI_t = SI_t - \frac{SI_{t-1} + SI_{t-2} + \cdots + SI_{t-90}}{90}$$

Second, to control for firm characteristics, we compute daily-level abnormal short interest with the regression-based approach suggested by Karpoff and Lou (2010) (hereafter KL). Following the KL methodology, we regress the detrended short interests of all the non-victim firms on size, book-to-market, and momentum indicators along with the industry dummies to control for industry-specific factors. With the estimated KL coefficients, we calculate the abnormal short interest of each victim firm (daily ABSI) as the difference between the actual detrended short

[20] Specifically, for each victim firm, we identify a benchmark portfolio based on the cross-sectional quintiles of the firm's market capitalization, book-to-market ratio, and prior 12-month compound returns. Using the daily value-weighted average return of the benchmark portfolio to compute the buy-and-hold benchmark return over the event window, we calculate breach-specific BHAR as the compound return of the victim firm net of the benchmark return over the same period.

interest on trading day *d* and the predicted value from the model. Compared to the KL model, an alternative high-dimensional industry-time fixed effect model would be less efficient because our sample on average has only 2 incidents for each Fama-French 48 industry over the entire sample period (12 years). Therefore, an industry-year panel regression would not have strong power to identify coefficients. A local peer-comparison KL model with firm fundamental and industry indicators is more appropriate. As a robustness check, we use the same procedure to generate two alternative abnormal short interest measures, which differ only in the set of variables included in the regression model (daily ABSI(2) and ABSI(3)).

Finally, with the daily-level measures, we calculate weekly abnormal short interest ABSI, ABSI(2), and ABSI(3) as the weekly average of the daily abnormal short interest to attenuate the volatility of daily measures. In our analyses, the event window covers the 19 weeks around a breach event date, spanning from event week -9 to event week 9.[21] Thus, we construct variables *ΔABSI_PRE*, *ΔABSI(2)_PRE*, and *ΔABSI(3)_PRE* to capture the change in abnormal short interest during the pre-start period, which is the abnormal short interest in event week 0 net of the abnormal short interest in event week -9.

To further distinguish information-driven short selling from non-informational motives, we draw on Blocher et al. (2023), who find that trading to close prior short positions, i.e., short covering, primarily reflects limits-to-arbitrage considerations, whereas new short positions opened, i.e., short volume, are more likely information-motivated. Using FINRA's millisecond-level intraday short transaction data, we compute daily short volume as the total volume of executed short trades, including both regular and aftermarket hours. We then apply the same detrending and

[21] Event week 0 spans the period from event day 0 to event day 4. Event day 0 is either the breach event date or the first trading day after the breach event date.

regression-based adjustment used to construct ABSI to estimate weekly abnormal short volume (ABSV), which better isolates information-driven trades.

In our main analysis, we focus on the baseline KL model (ABSI), as the other versions (ABSI(2) and ABSI(3)) yield consistent results. However, our main results are robust to several ways of constructing abnormal holding or trading measures, which include removing the KL model adjustments. We select and report some robustness test results in Internet Appendix IA4.

We also use variable *Lender Concentration Ratio* to capture the distribution of borrowed shares across lenders. Calculated as value-based concentration at the beginning of the pre-attack period, the variable ranges from zero (fully dispersed lending) to one (fully concentrated lending). We use the measure to assess whether dispersed lending facilitates harder-to-detect informed trading.

2.4 Retail Holding and Retail Short Selling Variables

We identify retail trades using the algorithm of Barber et al. (2024), who leverage sub-penny price improvements to identify retail trades, excluding trades within 40%-60% of the bid-ask spread. Buys and sells are determined by comparing trade prices to the NBBO midpoint. Daily retail net buying is calculated as the difference between retail buys and sells, using NYSE Trade and Quote (TAQ) tick-by-tick trade and quote data. We compute detrended retail holdings as a weighted sum of past retail net buys using a 90-day decay function: [22]

$$\text{D}etrended\ RH_t = RNB_t + \frac{89}{90} \times RNB_{t-1} + \frac{88}{90} \times RNB_{t-2} + \cdots + \frac{1}{90} \times RNB_{t-89},$$

where RH_t and RNB_t represent retail holdings and retail net buys in day t, respectively. We then construct the weekly abnormal retail holdings (ABRH) with the detrended daily measure by using KL's regression approach, which is analogous to the construction of abnormal short interest.

[22] See the Internet Appendix IA1 for a detailed derivation of the measure.

To more sharply identify information-motivated short selling, we attempt to compute measures that capture retail short selling. Our conversation with the NYSE research director suggests that 2/3 of FINRA short volume comes heavily (not 100%) from retail investors, while the remaining 1/3 is attributed to institutional liquidity-motivated trades. Therefore, the information-motivated component in FINRA short volume is more likely to originate from retail short selling already than the Markit data. We further identify the trades likely initiated by retail short sellers with the algorithm proposed by Boehmer and Song (2020), which classifies trades priced just above a round penny (fraction in (0, 0.4) cents) as retail short sells. We aggregate these trades for each trading day and adopt the identical procedures as those applied to estimating weekly abnormal short volume to construct abnormal retail short volume (ABRSV).

2.5 Institutional Holding and Insider Holding Variables

We define institutional trades as those exceeding $20,000 in value, following Lee and Radhakrishna (2000), and compute daily institutional net buying as the difference between larger buys and sells. We then construct weekly abnormal institutional holding (ABIT) using the detrending and KL-style regression framework of retail holding, which allows us to capture change in institutional holding not attributable to firm fundamentals.

To measure insider holdings, we rely on daily insider trading records from LSEG (formerly Thomson/Refinitiv) and follow the trader-based classification of Cohen et al. (2012) to identify opportunistic insider trades, which are more likely to reflect genuine insider information. We calculate daily net buys as the difference between buys and sells for opportunistic insider trades. These net buys are aggregated with decaying weights, mirroring our earlier construction for retail and institutional holdings. The resulting series captures the cumulative buildup of insider positions over time, reflecting changing sentiment or information. Applying the KL regression method to

this constructed series, we then compute weekly abnormal insider holding (ABIH) measured by opportunistic insider trades.

The inclusion of insider and institutional activity complements our short selling and retail trading measures. With the variables, we can conduct a comprehensive assessment of how informed trading varies across investor types. This full spectrum analysis provides sharper identification of whether informational advantages lie within or outside the firm in the context of cyberattacks.

2.6 Trading and Short Selling Cost Variables

To capture liquidity-related frictions, we use two measures of trading costs: relative spread and price impact. Relative spread is the bid-ask spread relative to the midpoint of the bid and ask prices, and price impact is estimated via the coefficient from regressing the log change in mid-price on the signed square root of dollar volume imbalances. Both measures are obtained from the Millisecond Intraday Indicators ("IID") database. In addition, we use the indicative fee from Markit as a proxy for the short selling cost, reflecting expected share borrowing costs.

We estimate abnormal trading costs by applying the KL regression framework to remove variation attributable to firm characteristics and industry effects. This yields measures of abnormal relative spread (ABRS), abnormal short cost (ABSC), and abnormal price impact (ABPI). As alternative measures to capture the costs, three additional variables (ABRS(2), ABSC(2), and ABPI(2)) are created by adding the controls of share turnover and institutional ownership to the KL regression model. The calculation procedure is consistent with KL Model 2.

2.7 Control Variables

We include several control variables to account for firm-level heterogeneity. *Log Size* is the natural logarithm of a firm's market capitalization, and *Book-to-Market* is the ratio of book equity to

market equity, both measured at the end of the month preceding the breach start date. *Ret6m* and *Ret30d* are six-month and 30-day DGTW-adjusted buy-and-hold returns, respectively. Both return variables are measured over the pre-start period. *Share Turnover* captures liquidity, and *Institutional Ownership* reflects the percentage of shares held by institutions at the end of the calendar quarter prior to the breach start. We also include *Idiosyncratic Volatility*, measured as the standard deviation of daily idiosyncratic returns, and *Illiquidity*, calculated as the average of daily return-to-dollar-volume ratios. All the variables are winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

3. Short Sellers in Data Breach

We begin by evaluating whether short sellers—who lack corporate insider access—anticipate the start of data breaches. Prior research shows that short sellers improve price efficiency through trading based on both private and public information (Engelberg, Reed, and Ringgenberg, 2012; Wang, Yan, and Zheng, 2020). They make profits by timing trades ahead of corporate events, anticipating the events' information content (Christophe et al., 2004; Christophe, 2010; Massoud, Nandy, Saunders, and Song, 2011). In the context of cyberattacks, short sellers may similarly detect and trade upon the breach risk. However, the unique risks associated with cyber-related speculation could restrain short selling activities (Engelberg, Reed, and Ringgenberg, 2018; Akey, Grégoire, and Martineau, 2022). We therefore test whether these outsiders build positions before the breach starts.

3.1 Short Selling around Breach Start

Using weekly abnormal short interest (ABSI), we trace trading behavior over a [-9, +9] week window centered on breach start, focusing on our 109 massive breach sample.[23] Figure 1 aligns week -9 to zero baseline and plots subsequent changes with 95% confidence bands. Panel A shows a significant increase in ABSI during the weeks -8 to -1, with a 0.44% peak in week –2 (about 0.32 standard deviation of non-event weekly ABSI).[24] Post-breach, ABSI rises modestly and declines to insignificance by week 9. This pattern suggests that short sellers accurately predict breach start despite lacking formal access to firm systems.

The timing pattern is consistent with anecdotal evidence that hackers typically require substantial preparation time to orchestrate cyberattacks. Appendix A Panel A illustrates the point with the case of 2013 Target data breach, in which hackers reportedly spent months probing the firm's systems, whereas breaches unfolded rapidly once executed. Short sellers who are informed about early signals may begin accumulating positions before hackers successfully penetrate the system. Therefore, it is possible that short interest increases several weeks prior to the breach.

Since certain industries can be more susceptible to cyberattacks, we conduct a within-industry comparison to examine whether short sellers target specific industries over individual victim firms (Kamiya et al., 2021). To examine the industry-wide abnormal short selling activity, we estimate the KL regression *without* industry fixed effects. Further, using two-digit SIC codes to identify industry peers, we compute the equal-weighted industry average of the weekly abnormal short interest for victim firms. Figure 1 Panel B shows that victim firms experience a 0.41% pre-breach ABSI rise (about 0.30 standard deviation of the non-event period distribution), whereas peer firms

[23] Since different KL models lead to similar results, our figures mainly illustrate the abnormal measure calculated through Model 1 for simplicity.

[24] The non-event period covers a two-year window surrounding Breach Start Date, excluding the period of the event week window. We use the same definition of non-event period for different variables in this paper.

remain flat. Post-breach, both groups see elevated ABSI, but the pre-breach pattern is unique to attacked firms. We also apply the test to a group of counterfactual peer firms identified through propensity score matching based on firms' cyber risk score, fundamentals, industry and year fixed effects. Internet Appendix IA5 Panel A reports no abnormal short interest around the breach start for the peer firms.

In addition, we examine the short selling activity around the start of smaller, non-massive breaches. Figure 1 Panel C confirms that these breaches do not attract pre-breach shorting. Taken together, the evidence suggests that short sellers identify specific attacks and anticipate the potential economic impact of the events effectively, rather than speculating broadly on cyber risk.

Importantly, short sellers who build positions ahead of an attack need not hold them until the firm's formal detection or disclosure date to make profits, because official detection and disclosure dates are administrative milestones rather than the only economically relevant information dates. Our results suggest the existence of different monetization channels. One possibility is the standard one: traders hold positions until the incident is publicly disclosed and profit from the ensuing price decline. Besides, a second possibility is that traders cover earlier because the market begins to incorporate the breach before the firm formally detects or discloses it. This is plausible because information about major breaches can diffuse rapidly after breach onset through multiple external channels, including government agencies, payment networks, banks, and cybersecurity firms. For example, in the 2013 Target breach, external parties including the U.S. Department of Justice and JPMorgan Chase alerted Target and card issuers, prompting the firm's investigation. Similarly, Home Depot was notified by law enforcement officials and banking contacts in connection with its 2014 breach. Even when firms initially detect an incident themselves, the subsequent

involvement of outside cybersecurity specialists and law enforcement can widen the circle of informed parties before any formal public announcement, as illustrated by Equifax in 2017.

While these possibilities can rationalize the profitability of short selling before breach, several caveats make the exact timing of profit realization difficult to pinpoint. First, the detection date we observe is the firm's formal discovery date, which may lag the moment at which firm outsiders or insiders first suspected a breach. Early signals about possible data breaches could lead to unexpected decreases in stock prices. Second, even hackers themselves face uncertainty about the exact timing of a successful attack, and they are likewise uncertain about whether the discovery of incident will occur immediately or only gradually. This uncertainty causes short covering decisions to vary across cases. Third, as disclosure approaches, a second wave of short sellers may enter after obtaining information through channels closer to the firm, including insiders. In that case, later short sellers may partially replace earlier informed traders, so the aggregate short interest can remain elevated even if the identity of the informed traders changes. Overall, official detection and disclosure dates are the noisy lower-frequency markers of a higher-frequency information process, not allowing us to precisely recover the profit-realization timing of first-moving short sellers.

3.2 Short Selling around Firm Detection and Disclosure

We next examine whether short sellers trade against subsequent firm's detection (i.e., when firm management first becomes aware of the intrusion) and public announcement (i.e. when firms officially confirm breaches to the public) of data breaches. Appendix A Panel B illustrates the timeline of massive data breaches. We find that the median gap between the breach start date and the firm detection date comprises 34 trading days, and the median gap between the start and disclosure dates includes 61 trading days. Since some market participants could be informed about the incident soon after the breach start date, it remains an open question whether short sellers are

willing to take more risks in order to target the detection and disclosure dates. Therefore, we investigate if short selling patterns around the firm detection and breach disclosure are different from the trading before the event onset.

Figure 2 Panel A plots ABSI around the firm detection date. While ABSI rises modestly—by approximately 0.17% from week -9 to week -2—this change is statistically insignificant, suggesting that short sellers do not intensify their trading in response to the subsequent firm detection. In addition, Figure 2 Panel B reports ABSI changes around the breach disclosure date. While a 0.16% pre-event rise is observed, the pattern again lacks statistical significance. These results suggest that short sellers do not specifically target firms' detection or announcement dates. The modest but insignificant change in short interest around firm detection and disclosure may instead reflect the aftermath of trading following breach onset. The results also confirm the distinctiveness of the short-selling pattern around the breach start date, which may imply early acquisition of breach-related information before breach onset, potentially from pre-breach activities or external networks linked to hackers.

3.3 Cross-Sectional Regressions

We conduct pooled cross-sectional regressions across all 262 breach events to test whether short sellers discriminate between high- and low-severity incidents. Specifically, we regress the pre-start change in ABSI on the *Massive Data Breaches* indicator, along with three alternative ABSI measures—Δ*ABSI_PRE*, Δ*ABSI(2)_PRE*, and Δ*ABSI(3)_PRE*—each constructed via a KL-style regression with expanding sets of control variables. The models also include variables *Breach Start BHAR*, *Breach Disclosure BHAR*, and *Breach Duration*, and standard controls in firm size, valuation, momentum, liquidity, and risk.

The regression results (Table 3) suggest that *Massive Data Breaches* predicts greater short selling. Pre-start ABSI is 0.35% higher for massive breaches—equivalent to a 0.24 standard deviation of *ΔABSI_PRE* variables—suggesting that short sellers not only anticipate breach timing but also infer its severity. Moreover, *Breach Start BHAR* is negatively associated with pre-start ABSI, reinforcing the notion that informed short selling predicts market reaction. Other breach characteristics (e.g., *Breach Duration*) are unrelated to short selling, highlighting that trading is concentrated around price-relevant breach events. Importantly, neither short interest nor its changes correlate with returns over the 30 days (*Ret30d*) or the 6 months before the attack (*Ret6m*), ruling out the reverse causality of reactive trading.

To explore whether equity lending conditions mediate informed short selling, we examine the role of lender concentration. We expect that low concentration—i.e., borrowing from a broad base of lenders—facilitates stealthier trades and mitigates recall or fee risks (D'Avolio, 2002; Geczy, 2002). Table 4 supports this view: Massive breaches attract significantly more short interest when lender concentration is lower. A one standard-deviation decline in *Lender Concentration Ratio* amplifies the pre-breach short interest response by 34% (e.g., Column (1): 0.165*1.912/0.918); a shift from the 90th to 10th percentile of *Lender Concentration Ratio* raises the short interest by 69% (e.g., Column (1): 0.330*1.912/0.918). This evidence affirms that market microstructure features can condition the exploitation of informational advantages (Comerton-Forde, Putniņš, and Tang, 2011).

3.4 Short Volume around Breach Start

To better isolate information-motivated trades, we analyze short volume—new positions initiated, rather than total short interest. Blocher et al. (2023) suggest that short volume is more likely to reflect speculative trading on private signals, whereas short covering (existing positions) mainly

reflects non-informational limits-to-arbitrage. Using FINRA's millisecond-level short selling transaction data, we construct weekly abnormal short volume (ABSV) around the breach start. Figure 3 shows that the ABSV increases significantly in weeks -5, -4, -3 and -1, peaking at roughly 0.05%, before subsiding post-breach. Combined with short interest evidence, the trend of abnormal short volume provides suggestive evidence that short sellers access and trade on private information prior to data breaches and bet on the negative price movement of the victim firm's stock. Notwithstanding, the evidence based on FINRA short selling sharpens the linkage of short selling activities to our information-based explanation.

4. Retail Investors in Data Breach

While retail investors are often considered the least sophisticated market participants, unable to consistently outperform benchmarks (Hvidkjaer, 2008), a growing body of research suggests that some subsets possess informational or strategic advantages. For example, Kaniel (2012) documents instances of informed trading among retail investors, and Barber and Odean (2000) and Kelley and Tetlock (2013) find evidence that retail investors can be less susceptible to the hedging motive than general short sellers. Motivated by this literature, we analyze retail holding and retail short selling around the start of breaches.

4.1 General Retail Holding around Breach Start

To assess retail behavior, we compute weekly abnormal retail holdings (ABRH) around the breach start date, identifying retail buys and sells using the methodology of Barber et al. (2024). Figure 4 shows that ABRH decreases by 0.06% during the pre-start period and reaches a statistically significant trough in the week [-3, 1] event window, amounting to 0.18 standard deviation of the weekly ABRH distribution during the non-event period. By the end of the event window, ABRH remains slightly depressed (-0.03%), though the decline is no longer statistically significant.

The timing and direction of these trades suggest that retail investors systematically reduce exposure ahead of cyber incidents. This is notable because prior literature typically portrays retail investors as reactive and misaligned with fundamentals. For example, Dong and Yang (2024) show that retail investors mostly trade on the wrong side as opposed to the stock return predictability inferred from public information. However, our findings suggest that, at least in this setting, retail trading behavior is directionally consistent with that of short sellers, raising the possibility that some retail investors may act on non-public signals related to impending breaches.

4.2 Retail Short Selling around Breach Start

To further examine potential information-based trading among retail participants, we analyze retail short-selling activities. Unlike institutional short sellers, retail short sellers are less likely to hedge exposures and thus may be more motivated by directional information (Boehmer and Song, 2020). We construct weekly abnormal retail short volume (ABRSV) using FINRA Short Sale Volume Data. Trades are classified based on the methodology proposed by Boehmer and Song (2020), which identifies retail short sells by trade pricing patterns near round pennies.

Figure 5 shows that ABRSV rises significantly in the weeks leading up to breach start for massive breaches. The peak occurs in week -4, where ABRSV reaches 0.004%--approximately 0.29 standard deviation of the weekly ABRSV of the non-event period. The elevation persists through the breach start week, with statistical significance at the 10% level over a five-week window preceding the attack. These results echo the abnormal short volume patterns observed in Section 3.4 (Figure 3), reinforcing the view that retail short sellers may act on privileged information ahead of cyber breaches. The rise in information-based retail short selling lends further support to the possibility that short sellers, at least a subset of them, could possess private

information about imminent data breaches, possibly through online forums, dark web intelligence, or proximity to breach perpetrators.

5. Insiders and Institutional Investors in Data Breach

In this section, we investigate whether other market participants, namely, company insiders and institutional investors, adjust their trading behavior in anticipation of cybersecurity breaches. While insiders are generally positioned to trade ahead of material corporate events due to privileged access (Fama, 1970), legal concerns may constrain such trading when it involves sensitive incidents like massive data breaches. Similarly, institutional investors could exhibit trading patterns aligned with insiders due to their ongoing interactions with management, which can result in an information set like that of insiders.

5.1 Insider Trading around Breach Start

Insiders are known to strategically time trades around corporate events and disclosures, often exploiting their informational advantage (Lee, Mikkelson, and Partch, 1992; Huddart, Ke, and Shi, 2007; Dechow, Lawrence, and Ryans, 2016). Thus, if insiders are privy to breach-related information before the breach starts, one would expect changes in their trading activity. These transactions may also serve as public signals for sophisticated outsiders, converting private information into broader market trading.

To assess insider activities, we examine changes in weekly abnormal insider holdings (ABIH) around the breach start date using the 109 massive data breaches. Using trader-based classification, abnormal insider holdings are measured by opportunistic insider trades, which should capture insider private information more accurately (Cohen et al., 2012). Figure 6 shows a modest rise in ABIH from week -5 to week 4 (up 0.005%). The measure then declines to nearly zero between week 5 and week 7. By week 9, insider holdings elevate by 0.002% compared to the week -9

baseline. However, none of these changes attain statistical significance. Overall, our findings suggest that insiders either do not trade around the breach start or trade in the opposite direction against the event, inconsistent with the common assumption that insiders hold informational advantage.

5.2 Institutional Trading around Breach Start

We next examine whether institutional investors adjust their positions around the breach start. The literature provides mixed evidence on the information content of institutional trades. Some studies find that institutions anticipate firm-specific events such as earnings surprises and news releases (Campbell, Ramadorai, and Schwartz, 2009; Yan and Zhang, 2009; Hendershott, Livdan, and Schürhoff, 2015). Others highlight conditions under which short-horizon institutional trades generate negative abnormal returns, raising questions about their predictive power (Chakrabarty, Moulton, and Trzcinka, 2017). Given this ambiguity, institutional behavior around cybersecurity events remains an open empirical question.

We plot weekly abnormal institutional holdings (ABIT) around the start date of massive breaches. Figure 7 shows that institutional holdings remain flat during the pre-start period. Post-breach, they increase by 0.11% from week 0 to week 5—roughly 0.18 standard deviation above the non-event period baseline—before declining by 0.10% from week 5 to week 9. This delayed drop, coupled with a lack of significant pre-event activity, suggests that institutions lack prior knowledge of the massive data breaches. The directionally inconsistent increase in holdings post-breach further weakens the case for informed institutional trading. Although our proxy for institutional trading—transactions exceeding $20,000—is relatively coarse, the absence of anticipatory trading suggests that institutions neither possess nor act upon private breach-related information. This contrasts with prior evidence showing institutions outperform retail investors.

Therefore, the muted institutional trading lends credence to the alternative possibility that the abnormal retail trading activities around the start of the cyber event originate from private information channels, rather than from broadly disseminated signals.

6. Capital Market Consequences

6.1 Market Reaction to Breach Start and Wealth Transfer

Table 3 shows a negative association between the pre-start short interest change and the subsequent market reaction, motivating a more direct analysis of capital market consequences. We first examine the abnormal reactions following the breach start and then estimate the associated financial wealth transfer, particularly from long-positioned investors to short sellers. Given prior evidence that short sellers disproportionately target massive data breaches, we partition our sample into the 109 massive breaches and all others. Within each group, we further separate breaches based on whether firms experience a positive or non-positive change in pre-breach abnormal short interest. For each subgroup, we calculate the equal-weighted buy-and-hold abnormal returns (BHARs) over [0, 30] trading day window using the DGTW-adjusted benchmarks.

Table 5 shows that the subsample of massive data breaches preceded by a positive abnormal short interest change have statistically significant BHARs which are nearly -3.4%, depending on the short interest specification. The return decline is even larger when the return window is extended through the end of week +9, ranging from approximately -4.2% to -4.4%. In contrast, massive breaches without pre-breach increases in short interest exhibit no significant abnormal returns. The return difference between the two subgroups is approximately -5%, showing statistical significance across all specifications. These results suggest that short sellers could potentially earn substantial profits by preemptively short selling stocks subject to cyberattacks.

Based on a representative market capitalization of $3 trillion, a 0.4% pre-breach rise in short interest, and a -2.7% post-breach BHAR across all 109 massive breaches, we estimate the financial wealth transfer to short sellers to be approximately $324 million.[25] Moreover, the precise short selling—positive change in short interest before the attacks—imply even greater profits about $924 million, even though the number of incidents drops from 109 to 72. This estimate, albeit conservative, illustrates the material capital reallocation triggered by informed outsider trading around the onset of data breach.

On the contrary, non-massive breaches yield no meaningful differences in post-event returns between the two short interest-based subgroups. Firms with increased pre-start short interest show statistically insignificant abnormal returns, while those without increased short selling experience mild gains. Thus, informed short selling appears to be concentrated around massive breaches. The pattern highlights the information value embedded in pre-start short interest changes, suggesting that investors correctly differentiate the economic implications of data breaches.

6.2 Trading and Short Selling Costs around Breach Start

We next examine whether increased short selling activities preceding the breach is associated with changes in market quality, focusing on trading and borrowing costs.[26] First, we analyze changes in stock liquidity by testing the weekly abnormal relative spread (ABRS and ABRS(2)) estimated using both the KL baseline model and an expanded specification that includes share turnover and institutional ownership. Figure 8 Panel A shows a statistically significant rise in ABRS beginning in week -2, peaking in week 0 with an increase of approximately 0.08%, or 0.36 standard deviation

[25] When calculating the wealth transfer, we use the same event week window [-9, +9] around the breach start date to be consistent with our main analyses on short selling activities around the breach start.

[26] The sample used to examine the short selling cost comprises 108 massive data breaches, whereas the sample designed to investigate stock liquidity consists of fewer than 100 data breaches. Some incidents are omitted due to missing short selling cost or stock liquidity data. We therefore winsorize these variables at the 2nd and 98th percentiles.

relative to the non-event period average. The elevated spread persists through week 2, with the end-of-window spread still 0.03% higher than the spread in week -9. These results suggest an increase in adverse selection risk around the attacks, consistent with informed outsider trading.

Second, we study the cost of short selling by plotting changes in the weekly abnormal short cost (ABSC) and its extended version ABSC(2) over the event week window [-9, 9]. Figure 8 Panel B shows that ABSC basically remains flat by week 0 but begins to rise since week 3, with a cumulative gain of about 0.20%—roughly 0.17 standard deviation of non-event period. The increase is statistically significant from week 5 to 9. Although we do not document a cost increase during the pre-attack stage, the broader rising trend is consistent with the view that surging demand for stock loans—driven by informed short selling—places upward pressures on borrowing costs, aligning with findings in Kolasinski, Reed, and Ringgenberg (2013).

Third, we analyze abnormal price impact (ABPI and ABPI(2)) as a measure of transitory illiquidity due to informed trading. Figure 8 Panel C shows that ABPI is volatile before week -3 and tends to increase afterwards. Specifically, the price impact rises by 0.17 standard deviation with statistical significance in week 7. At the end of the post-start period, we observe an increase of 0.10 standard deviation in ABPI. Considering the results from spreads, short costs, and price impact, our findings imply a tangible degradation in market quality during cyberattack events. Importantly, this market quality deterioration is unlikely attributable to insider information leakage. As documented earlier, there is no significant trading by insiders or institutional investors prior to the breach, and the observed patterns of short interest, volume, and cost coincide only with outsider activities. The decoupling between firm-internal information and trading outcomes strengthens the paper's core argument: the informed trading around data breaches originates externally to the firm. Further, the adverse consequences observed—increased spreads, rising borrowing costs, and

heightened price impact—represent a form of externality imposed by informed outsiders, reshaping both the firm's capital market environment and the trading conditions for uninformed investors.

7. Additional Analyses

We perform three supplementary analyses to corroborate our main finding—that outsider-generated information drives trading ahead of breaches—and to explore its broader applicability. First, we examine whether Google search activity anticipates breaches, as proxy for information acquisitions by outsiders, especially retail traders. Second, we use a set of pseudo breach start dates to test whether abnormal short interest patterns are indeed associated with impending breaches. Third, we explore whether short sellers target the start of *non-cyber* "outside attacks," such as competitor actions, operational disruption, or reputational events.

7.1 Google Search around Breach Start

Using Google Trends, we track abnormal search activity for keywords like "hack," "cyberattack," or "data breach" in conjunction with breached firm names over a [-45, 45] trading day window around massive breach start. We calculate the forward five-day moving average of each firm's Google Trends indexes and define abnormal Google searches relative to day -45. As reported in Figure 9 Panel A, Google searches spike between weeks -3 and -2, with statistically significant increases in the pre-start period. Figure 9 Panel B replicates this pattern for four high-profile breaches—Yahoo, eBay, Equifax, and Global Payments. In contrast, we do not find abnormal Google searches around the breach start for the counterfactual peer firms, as shown in Internet Appendix IA5 Panel B. The results suggest that search behavior reflects genuine information gathering by outsiders ahead of impending data breaches.

7.2 Pseudo Breach Start

To further verify that elevated short interest reflects informed trading rather than the coincidental market activity, we conduct a simulation using pseudo breach start dates. For each of the 109 massive data breaches, we randomly assign dates from the [-250, -46] and [46, 250] non-event windows, drawing with replacement. We simulate 100, 200, and 300 iterations, each time calculating the change in weekly abnormal short interest *(*Δ*ABSI_PRE*, Δ*ABSI(2)_PRE*, and Δ*ABSI(3)_PRE*) over the [-9, 0] window. A dummy variable, *Actual*, equals one if real breaches and zero for pseudo-events.

Table 6 reports the simulation results. Panel A shows that the increase in short interest before real breach dates significantly exceeds that prior to pseudo-events, with a consistent difference of approximately 0.36%. Given that abnormal short interest increases roughly 0.40% during the pre-start period of actual breaches, the lack of similar patterns in the pseudo-events adds to the notion that trading reflects anticipation of real breaches. Panels B and C confirm this pattern across alternative model specifications, providing further evidence that the significant rise in pre-start short selling is not spurious.

7.3 More General Outsider Attacks

We expand our analysis beyond cybersecurity incidents to assess whether informed short selling occurs around broader outsider-generated shocks. First, we analyze the February 29, 2024 protest and subsequent suspected arson at Tesla's Berlin factory, which we labeled as an "Operational Disruption." Tesla's stock dropped 19% in 10 days following the protest, wiping out $124 billion in Tesla's market capitalization. Figure 10 Panel A shows that the abnormal short interest rose by

0.20% in the six weeks prior to the protest,[28] indicating approximately $1.3 billion in short positions. The implied profit from this episode approaches $250 million, again pointing to the possibility that some traders acted on foreknowledge of this operational shock.

Second, we consider a reputational crisis to JD.com stemming from the civil lawsuit filed against founder Richard (Qiangdong) Liu. The lawsuit itself constituted an externally initiated legal and reputational shock, triggering a roughly 10% stock price decline and a $2.78 billion loss in market capitalization. Figure 10 Panel B shows abnormal short interest increased by 0.17% in the six weeks leading up to the civil suit's April 16, 2019 filing, implying about $47 million in short volume and nearly $5 million in trading profit for a 10% price decline. Panel B also reports a significant rise in the abnormal short interest which exceeds 1% before the lawsuit initiation.

Third, we examine DeepSeek's AI releases and their implications for Nvidia Corporation, which we classify as an externally generated fundamental shock. On December 25, 2024, DeepSeek announced a new AI model widely perceived as rivaling leading global counterparts at far lower cost and with almost comparable capability, with some reports suggesting that part of its performance gains reflected distillation from leading U.S. models such as ChatGPT; on January 20, 2025—the date of the U.S. presidential inauguration—it released DeepSeek-R1, which soon surged in popularity. The timing of these releases, coinciding with both Christmas and a symbolically charged political date in the United States, likely amplified their psychological and market impact. The shock triggered a 17% decline in Nvidia's stock price, erasing roughly $588 billion in market capitalization, and contributed to nearly $1 trillion in losses across the broader U.S. stock market. Figure 10, Panel C, shows that Nvidia's abnormal short interest rose by 0.05%

[28] For Tesla and Nvidia stocks, we calculate their five-day moving average of daily short interest. For JD.Com, we use the raw daily short interest of the ADR stock as Markit categorizes JD.com in a separate dataset called "Other," which may not provide clean data due to JD.Com's cross-listing status.

in the three weeks before December 25, 2024, and by a further 0.02% before January 20, 2025. Given roughly $2.4 billion in shares involved, this pattern would imply profits of about $408 million from a 17% price decline. These findings are consistent with short sellers positioning ahead of an external shock generated by rival innovation rather than by information originating within the firm. Consistent with the unusual timing and magnitude of the move, Bill Ackman publicly called for an investigation into suspicious short selling around the DeepSeek releases.

Overall, these cases suggest that opportunistic short selling is not confined to cybersecurity incidents but extends to a broader set of outsider-generated shocks, raising broader concerns about market vulnerability in an era of greater digital access and reputational interdependence.

8. Conclusion

The increasing frequency and complexity of cybersecurity breaches represent a structural shift in the information dynamics of financial markets. These outsider-initiated events invert the traditional hierarchy of informational asymmetry by granting an informational advantage to those external to the firm, such as hackers, dark-web participants, and informed speculators, while leaving corporate insiders in the dark. Our comprehensive analysis of cybersecurity breaches at U.S. firms from 2007 to 2018 reveals that short sellers and retail investors—both outsiders—consistently anticipate price drops before the formal detection or disclosure of attacks. In contrast, insiders and institutions, historically presumed to be better informed, show no discernible pre-attack trading advantage. These findings cast doubt on foundational models of insider-dominant information environments and suggest that financial markets are increasingly shaped by information flows that originate beyond the corporate boundary.

The rise in abnormal short interest before breaches—absent around pseudo-events or industry peers with comparable ex ante cyber risk—highlights the precision and timing of informed outsider

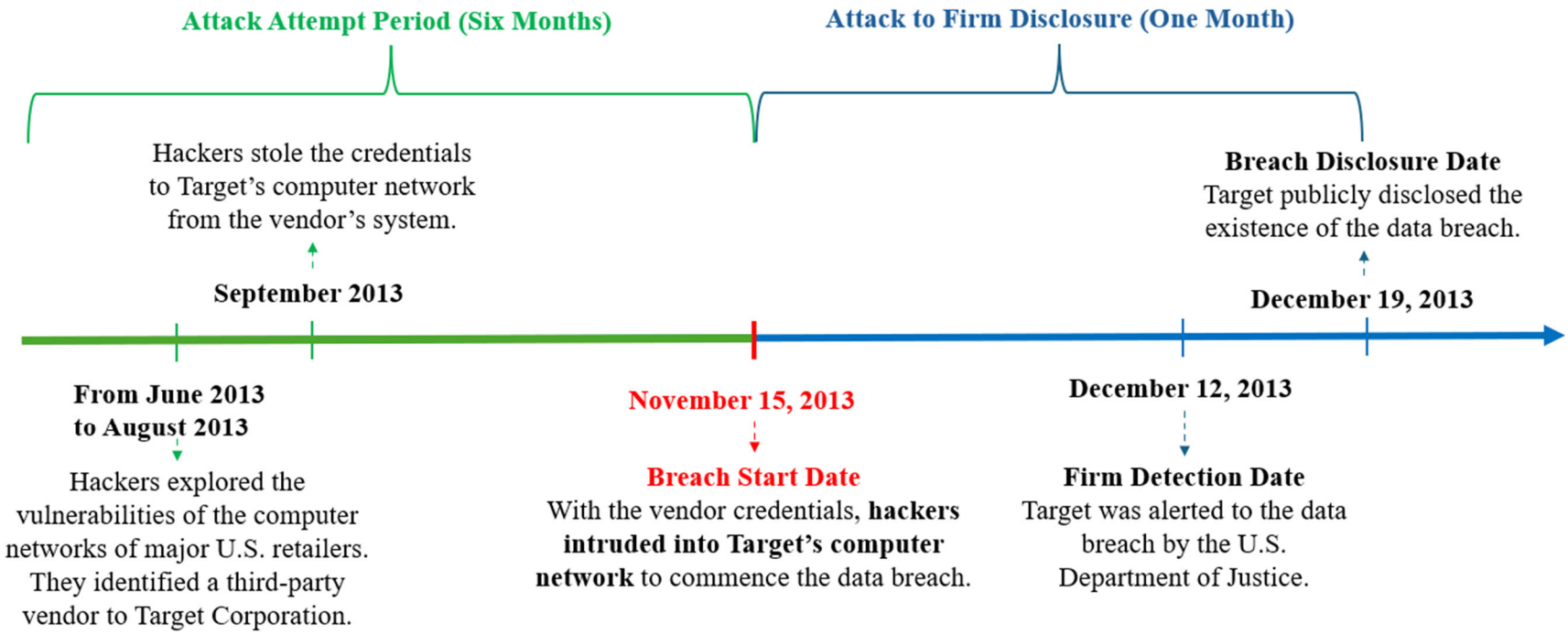
trading. These trades peak just before attacks and dissipate soon after, generating profits much higher than the costs of ransomware or disclosure delays. Moreover, short selling costs and illiquidity also increase, consistent with rising information asymmetry not from insiders but from unknown external actors. Retail investors, typically considered unsophisticated, also preemptively reduce holdings and increase short volume, suggesting access to alternative signals via digital or social channels like Google Search. Meanwhile, insiders and institutions—who might be presumed better positioned—remain inert, either due to the legal risk or the lack of access to this new informational frontier. We also document short selling around the start of different outsider attacks, indicating that the paradigm shift in the information dynamics has more general implications.

Taken together, our results point to a novel and underexplored form of informational asymmetry rooted in outsider-generated shocks. This shift challenges classic theories of market microstructure, disclosure policy, and asset pricing, all of which were designed around the assumption that insiders dominate the information hierarchy. As cyber threats and other external shocks become more prevalent, existing regulatory frameworks may inadvertently aid malicious actors while leaving uninformed investors increasingly exposed. Our findings underscore the need for dynamic regulatory adaptation and a broader theoretical rethinking of who holds, creates, and monetizes material information in the modern financial market.

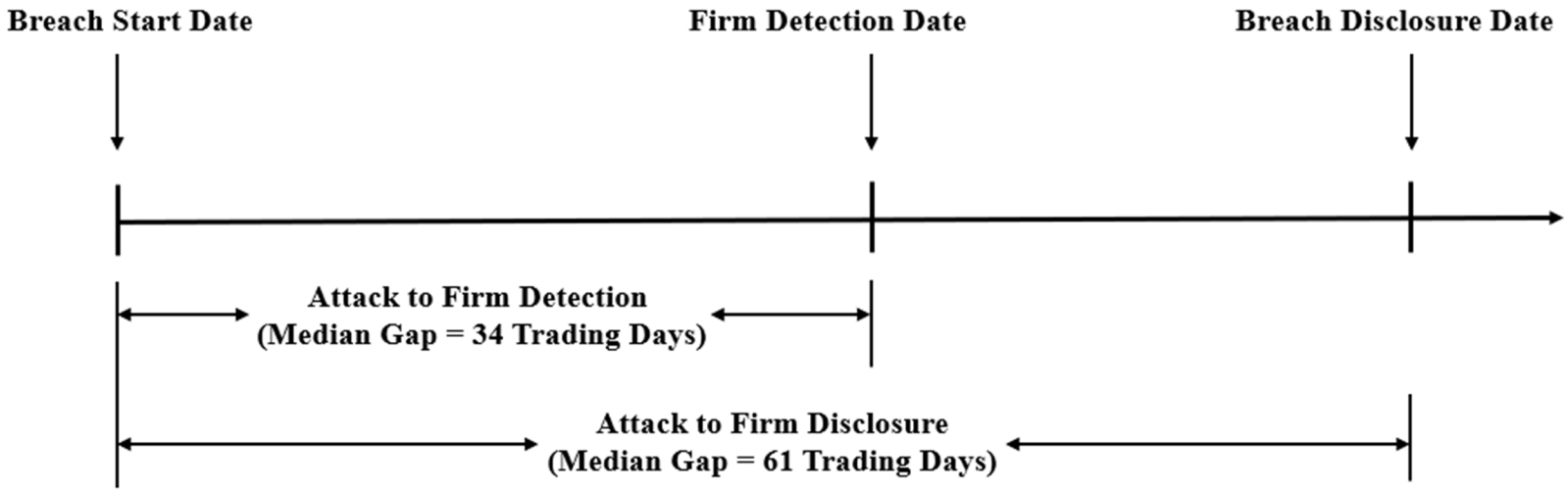
Appendix A. Data Breach Chronology

Appendix A presents the chronological sequence of cyber data breaches. Panel A illustrates the timeline of Target's 2013 data breach, which was incurred by a sophisticated cyberattack that led to the theft of over 40 million credit and debit card records and 70 million customer records. Key facts about the event are derived from court filings in the subsequent customer lawsuits, as well as other publicly available online sources. Panel B shows the timeline of massive data breaches, reporting the median gaps between attack initiation, firm detection, and firm disclosure. The start date is the date when the firm's central system was first compromised by an external intruder; the detection date is the date when the firm first became aware of the breach, either by internally observing suspicious activities or being notified of the breach by an external third party; the disclosure date is the date when the firm publicly announced the incident.

Panel A Timeline of 2013 Target Data Breach



Panel B Timeline of Massive Data Breaches



References

[1] Akey, P.; V. Grégoire; and C. Martineau. "Price Revelation from Insider Trading: Evidence from Hacked Earnings News." *Journal of Financial Economics*, 143 (2022),1162–1184.

[2] Allen, F.; and D. Gale. "Stock-Price Manipulation." *The Review of Financial Studies*, 5 (1992), 503–529.

[3] Barber, B. M.; and T. Odean. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance*, 55 (2000), 773–806.

[4] Barber, B. M.; X. Huang; P. Jorion; T. Odean; and C. Schwarz. "A (Sub)penny for Your Thoughts: Tracking Retail Investor Activity in TAQ." *The Journal of Finance*, 79 (2024), 2403–2427.

[5] Blocher, J.; X. Dong; M. C. Ringgenberg; and P. G. Savor. "Short Covering." Working Paper, Available at SSRN: 2634579 (2023).

[6] Boehmer, E.; C. M. Jones; and X. Zhang. "Which Shorts Are Informed?" *The Journal of Finance*, 63 (2008), 491–527.

[7] Boehmer, E.; and W. Song. "Smart Retail Traders, Short Sellers, and Stock Returns." Working Paper, Available at SSRN: 3723096 (2020).

[8] Boehmer, E.; C. M. Jones; X. Zhang; and X. Zhang. "Tracking Retail Investor Activity." *The Journal of Finance*, 76 (2021), 2249–2305.

[9] Boehmer, E.; and J. (Julie) Wu. "Short Selling and the Price Discovery Process." *The Review of Financial Studies*, 26 (2013), 287–322.

[10] Brugler, J.; and C. Comerton-Forde. "Differential Access to Dark Markets and Execution Outcomes." *Journal of Financial Economics*, 171 (2025), 104086.

[11] Brunnermeier, M. K.; and M. Oehmke. "Predatory Short Selling." *Review of Finance*, 18 (2014), 2153–2195.

[12] Bushee, B. J.; D. A. Matsumoto; and G. S. Miller. "Open versus Closed Conference Calls: the Determinants and Effects of Broadening Access to Disclosure." *Journal of Accounting and Economics*, 34 (2003), 149–180.

[13] Campbell, J. Y.; T. Ramadorai; and A. Schwartz. "Caught on Tape: Institutional Trading, Stock Returns, and Earnings Announcements." *Journal of Financial Economics*, 92 (2009), 66-91.

[14] Cao, S.; W. Jiang; J. L. Wang; and B. Yang. "From Man vs. Machine to Man+ Machine: The Art and AI of Stock Analyses." *Journal of Financial Economics*, 160 (2024), 103910.

[15] Chakrabarty, B.; P. C. Moulton; and C. Trzcinka. "The Performance of Short-Term Institutional Trades." *Journal of Financial and Quantitative Analysis*, 52 (2017), 1403–1428.

[16] Chen, Q.; I. Goldstein; and W. Jiang. "Price Informativeness and Investment Sensitivity to Stock Price." *The Review of Financial Studies*, 20 (2007), 619–650.

[17] Chen, Y.; Z. Da; and D. Huang. "Arbitrage Trading: The Long and the Short of It." *The Review of Financial Studies*, 32 (2018), 1608–1646.

[18] Chen, Y.; Z. Da; and D. Huang. "Short Selling Efficiency." *Journal of Financial Economics*, 145 (2022), 387–408.

[19] Christophe, S. E.; M. G. Ferri; and J. J. Angel. "Prior to Earnings Announcements." *The Journal of Finance*, 59 (2004), 1845–1875.

[20] Christophe, S. E.; M. G. Ferri; and J. Hsieh. "Informed Trading before Analyst Downgrades: Evidence from Short Sellers." *Journal of Financial Economics*, 95 (2010), 85–106.

[21] Cohen, L.; C. Malloy; and L. Pomorski. "Decoding Inside Information." *The Journal of Finance*, 67 (2012), 1009–1043.

[22] Comerton-Forde, C.; T. J. Putniņš; and K. M. Tang. "Why Do Traders Choose to Trade Anonymously?" *Journal of Financial and Quantitative Analysis*, 46 (2011), 1025–1049.

[23] Da, Z.; J. Engelberg; and P. Gao (2011). In search of attention. *The Journal of Finance*, 66 (2011), 1461-1499.

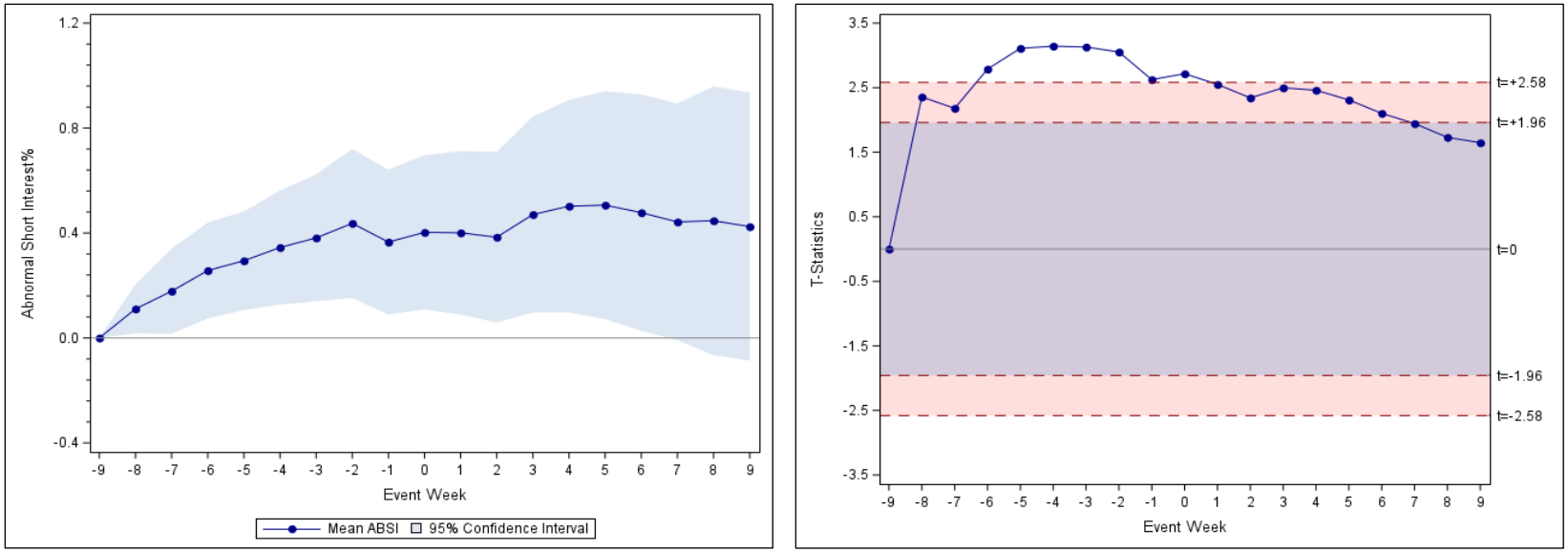
[24] Daniel, K.; M. Grinblatt; S. Titman; and R. Wermers. “Measuring Mutual Fund Performance with Characteristic-Based Benchmarks.” *The Journal of Finance*, 52 (1997), 1035–1058.
[25] D’Avolio, G. “The Market for Borrowing Stock.” *Journal of Financial Economics*, 66 (2002), 271-306.
[26] Dechow, P. M.; A. Lawrence; and J. P. Ryans. “SEC Comment Letters and Insider Sales.” *The Accounting Review*, 91 (2016), 401–439.
[27] Diamond, D. W.; and R. E. Verrecchia. “Constraints on Selling and Asset Price Adjustment to Private Information.” *Journal of Financial Economics*, 18 (1987), 277–311.
[28] Diether, K. B.; K-H. Lee; and I. M. Werner. “Short-Sale Strategies and Return Predictability.” *The Review of Financial Studies*, 22 (2009), 575–607.
[29] Dong, X.; and C. Yang. “Anomalies Never Disappeared: The Case of Stubborn Retail Investors.” Working Paper, Available at SSRN 4417278 (2024).
[30] Easley, D.; and M. O'hara. "Information and the cost of capital." *The Journal of Finance* 59 (2004), 1553-1583.
[31] Edmans, A.; I. Goldstein; and W. Jiang. “Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage.” *The American Economic Review*, 105 (2015), 3766–3797.
[32] Engelberg, J. E.; A. V. Reed; and M. C. Ringgenberg. “How Are Shorts Informed?: Short Sellers, News, and Information Processing.” *Journal of Financial Economics*, 105 (2012), 260–278.
[33] Engelberg, J. E.; A. V. Reed; and M. C. Ringgenberg. “Short-Selling Risk.” *The Journal of Finance*, 73 (2018), 755–786.
[34] Fama, E. F. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance*, 25 (1970), 383–417.
[35] Florackis, C.; C. Louca; R. Michaely; and M. Weber. “Cybersecurity Risk.” *The Review of Financial Studies*, 36 (2023), 351–407.
[36] Geczy, C. C.; D. K. Musto; and A. V. Reed. “Stocks Are Special Too: An Analysis of the Equity Lending Market.” *Journal of Financial Economics*, 66 (2002), 241–269.
[37] Glosten, L. R.; and P. R. Milgrom. “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders.” *Journal of Financial Economics*, 14 (1985), 71-100.
[38] Goldstein, I.; and A. Guembel. “Manipulation and the Allocational Role of Prices.” *The Review of Economic Studies*, 75 (2008), 133–164.
[39] Gordon, L. A.; M. P. Loeb; and T. Sohail. “Market Value of Voluntary Disclosures Concerning Information Security.” *MIS Quarterly*, 34 (2010), 567–594.
[40] Grossman, S. J.; and J. E. Stiglitz. “On the Impossibility of Informationally Efficient Markets.” *The American Economic Review*, 70 (1980), 393-408.
[41] Hendershott, T.; D. Livdan; and N. Schürhoff. “Are Institutions Informed about News?” *Journal of Financial Economics*, 117 (2015), 249–287.
[42] Honkanen, P. “Securities Lending and Trading by Active and Passive Funds.” *Journal of Financial and Quantitative Analysis*, 60 (2025), 1272-1309.
[43] Hu, G. X.; J. Pan; and J. Wang. “Early Peek Advantage? Efficient Price Discovery with Tiered Information Disclosure.” *Journal of Financial Economics*, 126 (2017), 399-421.
[44] Huang, H. H.; and C. Wang. “Do Banks Price Firms’ Data Breaches.” *The Accounting Review*, 96 (2021), 261–286.
[45] Huddart, S.; B. Ke; and C. Shi. “Jeopardy, Non-Public Information, and Insider Trading around SEC 10-K and 10-Q Filings.” *Journal of Accounting and Economics*, 43 (2007), 3–36.
[46] Hvidkjaer, S. “Small Trades and the Cross-Section of Stock Returns.” *The Review of Financial Studies*, 21 (2008), 1123–1151.
[47] Kamiya, S.; J-K. Kang; J. Kim; A. Milidonis; and R. M. Stulz. “Risk Management, Firm Reputation, and the Impact of Successful Cyberattacks on Target Firms.” *Journal of Financial Economics*, 139 (2021), 719–749.
[48] Kaniel, R.; S. Liu; G. Saar; and S. Titman. “Individual Investor Trading and Return Patterns around Earnings Announcements.” *The Journal of Finance*, 67 (2012), 639–680.

[49] Karpoff, J. M.; and X. Lou. “Short Sellers and Financial Misconduct.” *The Journal of Finance*, 65 (2010), 1879–1913.
[50] Kelley, E. K.; and P. C. Tetlock. “How Wise Are Crowds? Insights from Retail Orders and Stock Returns.” *The Journal of Finance*, 68 (2013), 1229–1265.
[51] Kolasinski, A. C.; A. V. Reed; and M. C. Ringgenberg. “A Multiple Lender Approach to Understanding Supply and Search in the Equity Lending Market.” *The Journal of Finance*, 68 (2013), 559–595.
[52] Kyle, A. S. “Continuous Auctions and Insider Trading.” *Econometrica: Journal of the Econometric Society*, (1985), 1315-1335.
[53] Lee, C. M. C.; and B. Radhakrishna. “Inferring Investor Behavior: Evidence from TORQ Data.” *Journal of Financial Markets*, 3 (2000), 83–111.
[54] Lee, D. S.; W. H. Mikkelson; and M. M. Partch. “Managers’ Trading Around Stock Repurchases.” *The Journal of Finance*, 47 (1992), 1947–1961.
[55] Lin, Z.; T. R. Sapp; J. R. Ulmer; and R. Parsa. “Insider Trading ahead of Cyber Breach Announcements.” *Journal of Financial Markets*, 50 (2020), 100527.
[56] Massoud, N.; D. Nandy; A. Saunders; and K. Song. “Do Hedge Funds Trade on Private Information? Evidence from Syndicated Lending and Short-Selling.” *Journal of Financial Economics*, 99 (2011), 477–499.
[57] Michel, A.; J. Oded; and I. Shaked. “Do Security Breaches Matter? The Shareholder Puzzle.” *European Financial Management*, 26 (2020), 288-315.
[58] Miller, E. M. “Risk, Uncertainty, and Divergence of Opinion.” *The Journal of Finance*, 32 (1977), 1151–1168.
[59] Piotroski, J. D.; and D. T. Roulstone. “Do Insider Trades Reflect Both Contrarian Beliefs and Superior Knowledge about Future Cash Flow Realizations?” *Journal of Accounting and Economics*, 39 (2005), 55–81.
[60] Rapach, D. E.; M. C. Ringgenberg; and G. Zhou. “Short Interest and Aggregate Stock Returns.” *Journal of Financial Economics*, 121 (2016), 46–65.
[61] Solomon, D.; and E. Soltes. “What Are We Meeting For? The Consequences of Private Meetings with Investors.” *The Journal of Law and Economics*, 58 (2015), 325–355.
[62] Spanos, G.; and L. Angelis. “The Impact of Information Security Events to the Stock Market: A Systematic Literature Review.” *Computers & Security*, 58 (2016), 216–229.
[63] Wang, X.; X. (Sterling) Yan; and L. Zheng. “Shorting Flows, Public Disclosure, and Market Efficiency.” *Journal of Financial Economics*, 135 (2020), 191–212.
[64] Yan, X. (Sterling); and Z. Zhang. “Institutional Investors and Equity Returns: Are Short-Term Institutions Better Informed?” *The Review of Financial Studies*, 22 (2009), 893–924.

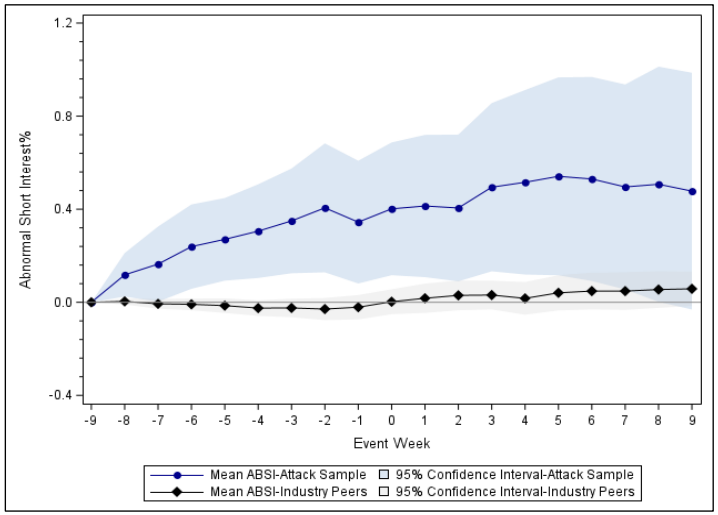
Figure 1. Abnormal Short Interest around Breach Start Date

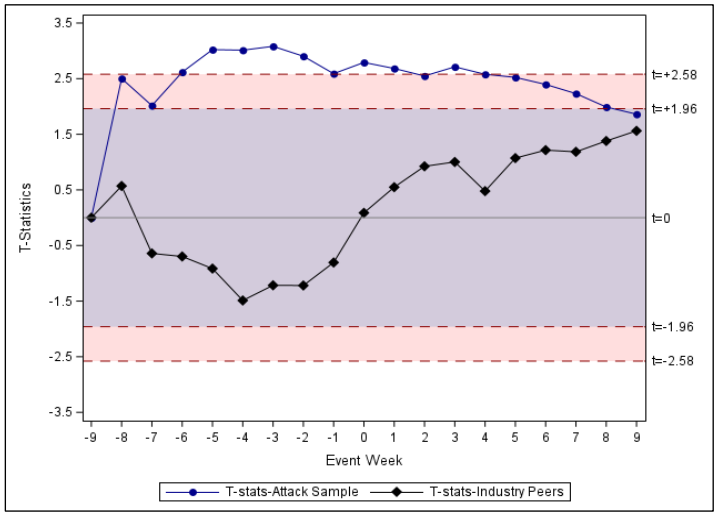
Figure 1 plots the change in the weekly abnormal short interest (*ABSI*) and the corresponding t-statistics of 109 massive data breaches, non-attacked industry peers, and non-massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Massive data breaches are defined as incidents initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data, while non-massive breaches are mainly attacks targeting a firm's local retail facilities. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal short interest *ABSI* is the average of daily abnormal short interest over the event week. To estimate abnormal short interest, we remove firm-specific systematic time trends and purge the variation in short interest associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABSI* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

Panel A Massive Data Breaches

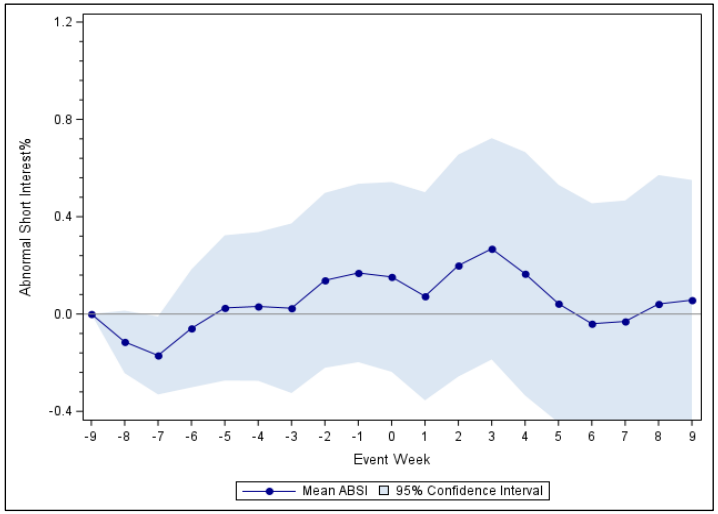


Panel B Massive Data Breaches VS. Industry Peers





Panel C Non-Massive Data Breaches



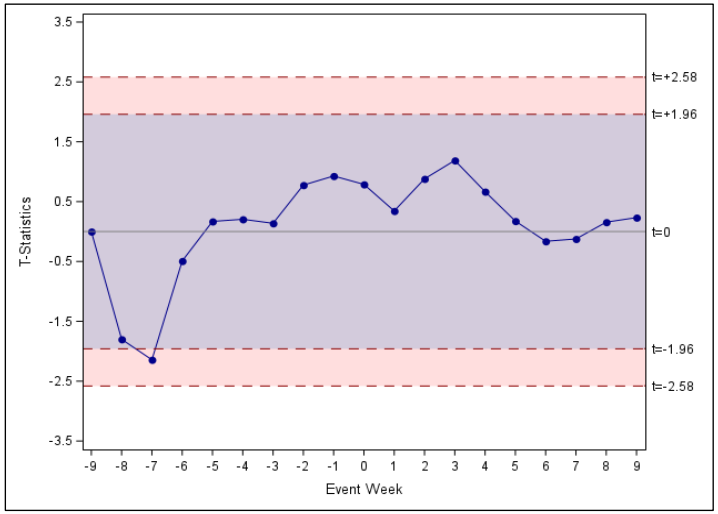
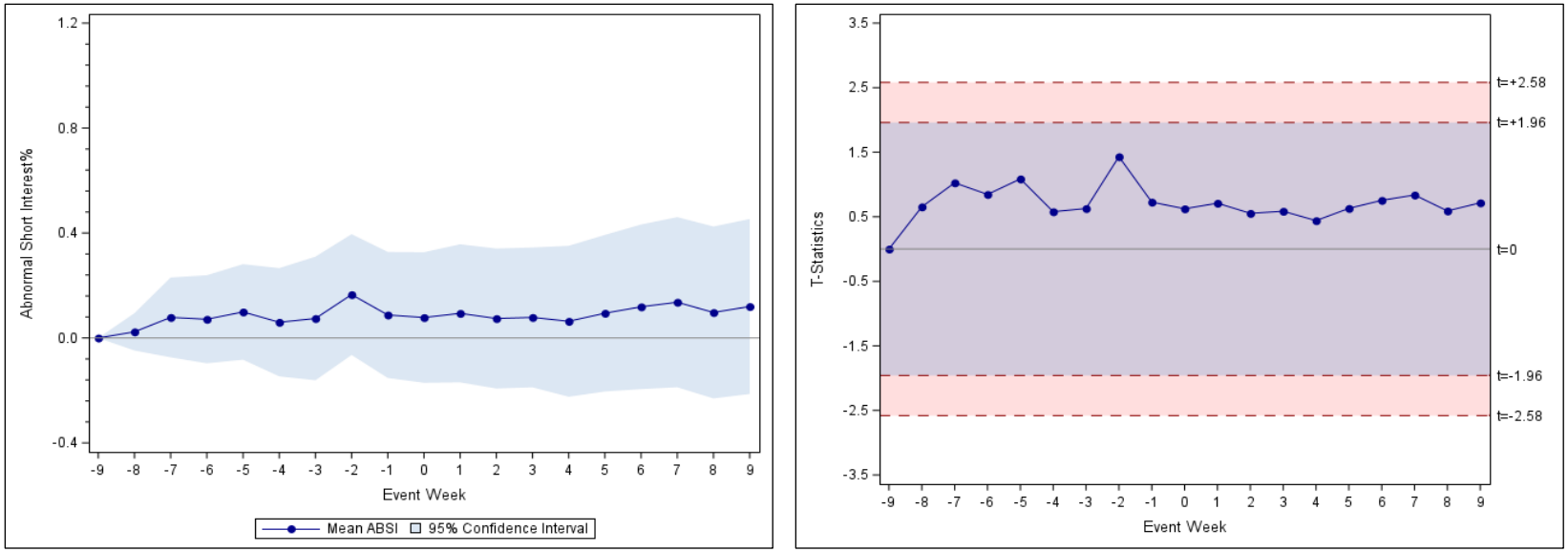


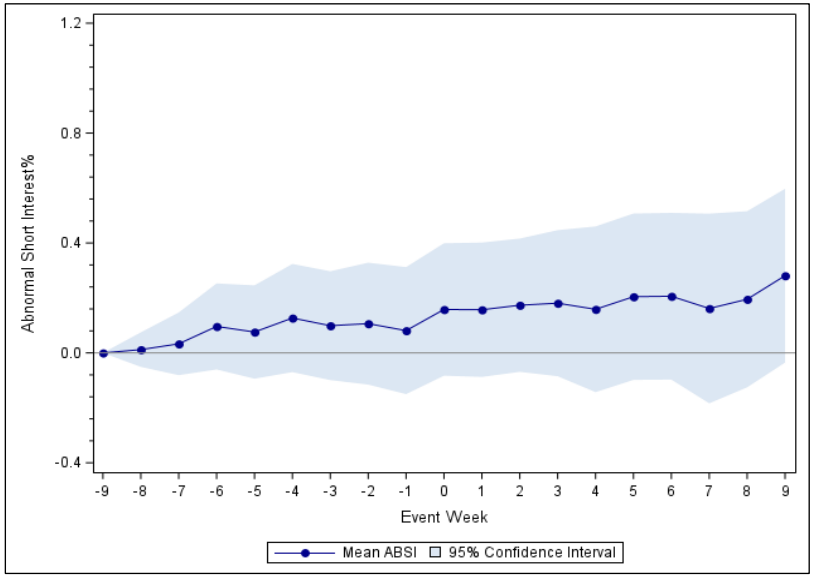
Figure 2. Abnormal Short Interest around Firm Detection Date or Breach Disclosure Date

Figure 2 plots the change in the weekly abnormal short interest (*ABSI*) and the corresponding t-statistics of 109 massive data breaches around other key event dates (*Firm Detection Date* and *Breach Disclosure Date*) over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal short interest *ABSI* is the average of daily abnormal short interest over the event week. To estimate abnormal short interest, we remove firm-specific systematic time trends and purge the variation in short interest associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Firm Detection Date* (*Breach Disclosure Date*) or the first trading day after *Firm Detection Date* (*Breach Disclosure Date*). Internet Appendix IA1 provides the detailed definitions of *ABSI*, *Firm Detection Date*, and *Breach Disclosure Date*. The dashed lines in the t-statistic figures indicate the critical values at the 5% and 1% significance levels.

Panel A Abnormal Short Interest around Firm Detection Date



Panel B Abnormal Short Interest around Breach Disclosure Date



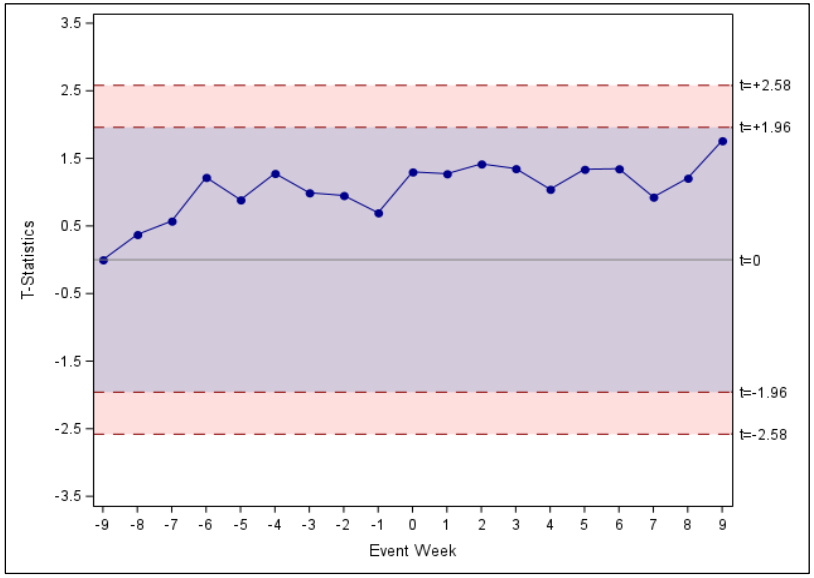


Figure 3. Abnormal Short Volume around Breach Start Date

Figure 3 plots the change in the weekly abnormal short volume (*ABSV*) and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal short volume *ABSV* is the average of daily abnormal short volume over the event week. To estimate abnormal short volume, we remove firm-specific systematic time trends and purge the variation in short volume associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABSV* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

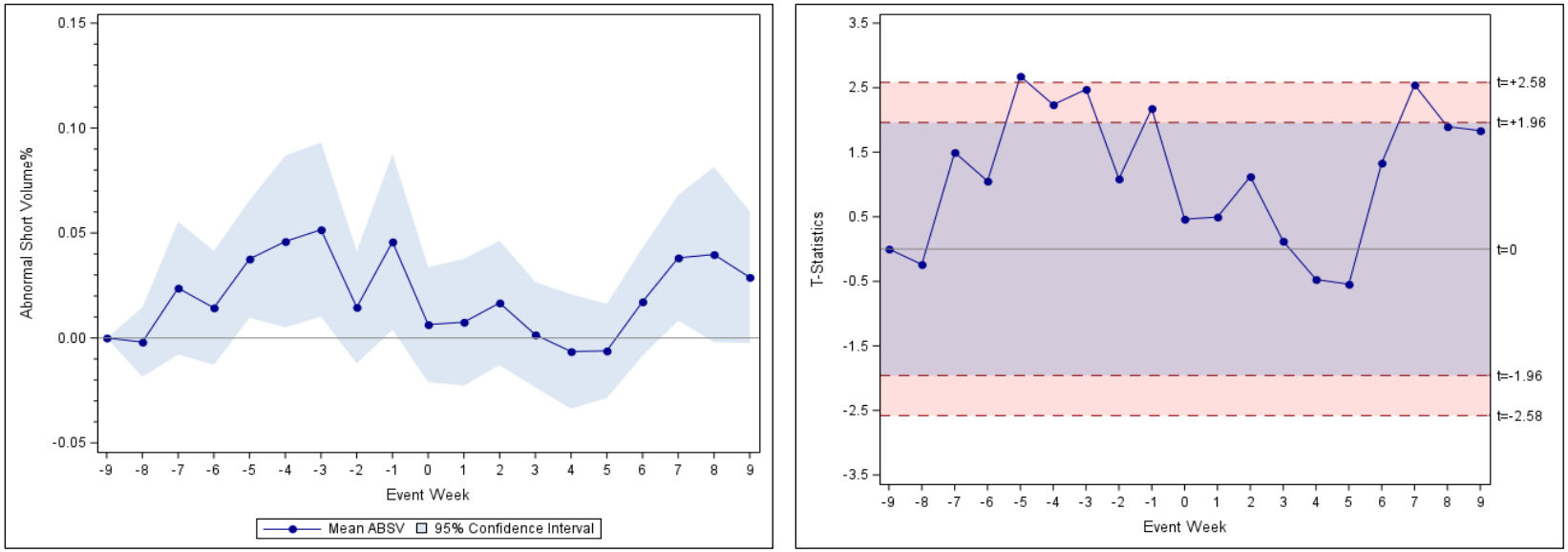


Figure 4. Abnormal Retail Holdings around Breach Start Date

Figure 4 plots the change in the weekly abnormal retail holdings (*ABRH*) and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal retail holdings *ABRH* is the average of daily abnormal retail holdings over the event week. To estimate abnormal retail holdings, we remove firm-specific systematic time trends and control for *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABRH* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

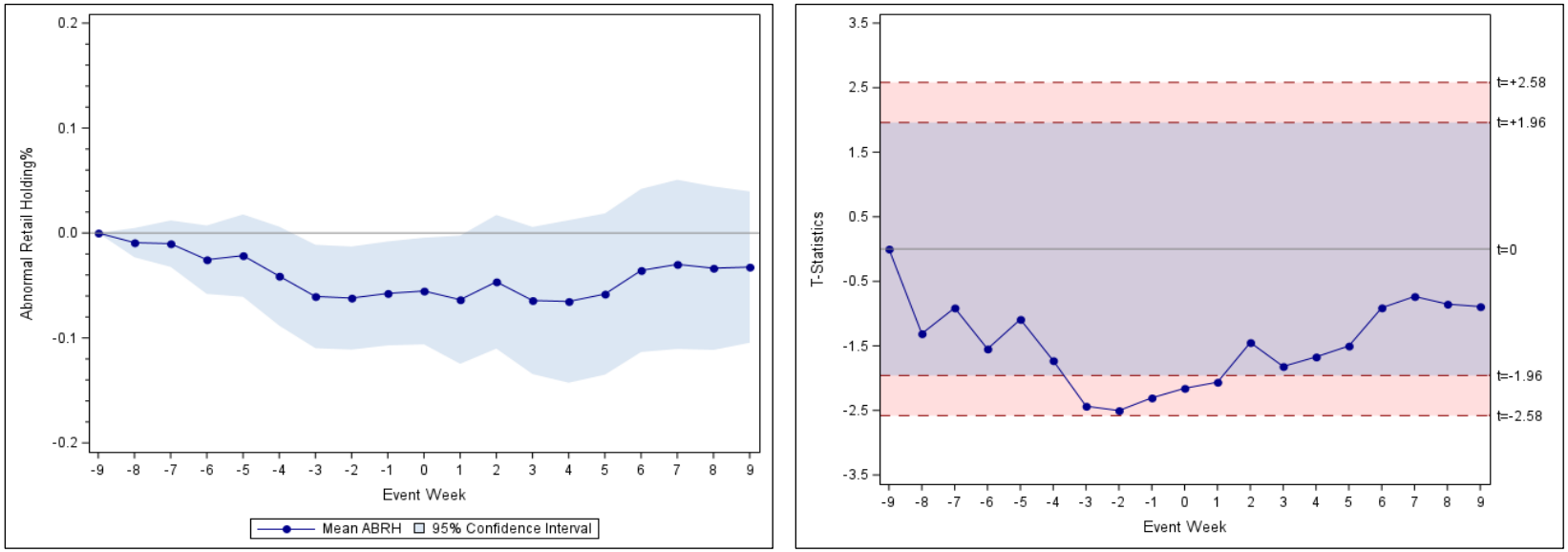


Figure 5. Abnormal Retail Short Volume around Breach Start Date

Figure 5 plots the change in the weekly abnormal retail short volume (*ABRSV*) and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal retail short volume *ABRSV* is the average of daily abnormal retail short volume over the event week. To estimate abnormal retail short volume, we remove firm-specific systematic time trends and purge the variation in retail short volume associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABRSV* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

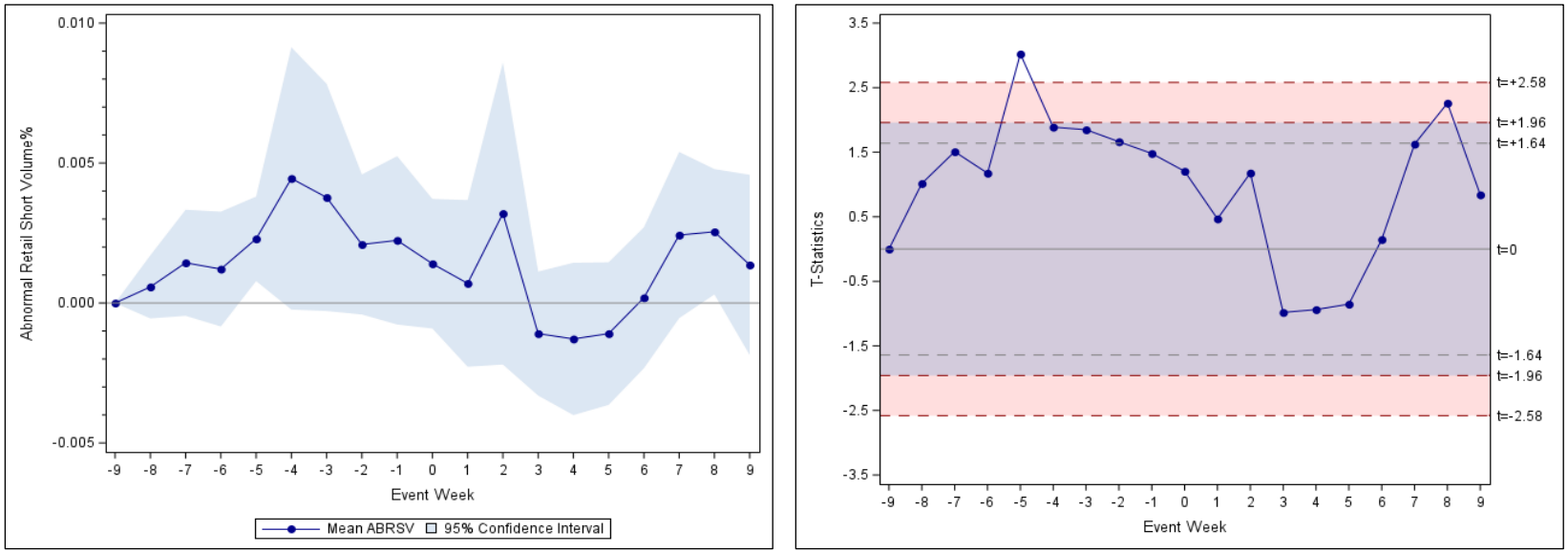


Figure 6. Abnormal Insider Holdings around Breach Start Date

Figure 6 plots the change in the weekly abnormal insider holding (*ABIH*), computed with opportunistic insider trades (Cohen et al., 2012), and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal insider holding is the average of daily abnormal measures over the event week. To estimate abnormal insider holding, we remove firm-specific systematic time trends and purge the variation in insider holding associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABIH* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

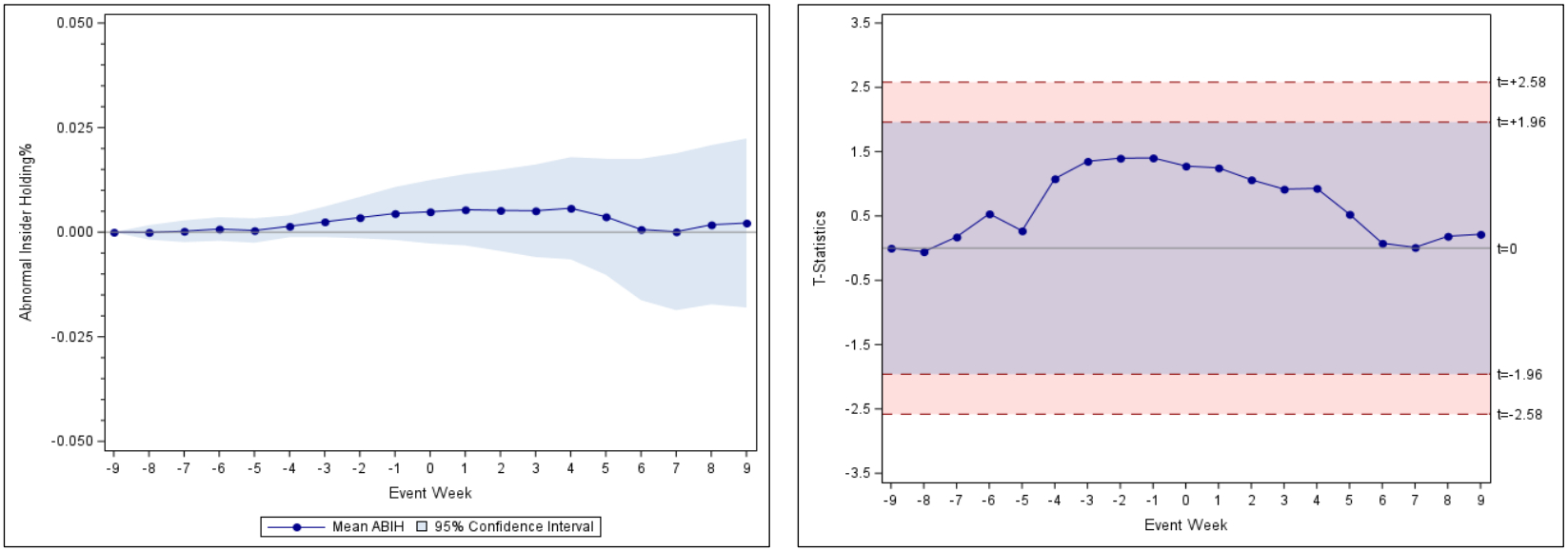


Figure 7. Abnormal Institutional Holdings around Breach Start Date

Figure 7 plots the change in the weekly abnormal institutional holding (*ABIT*) and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Weekly abnormal institutional holding *ABIT* is the average of daily abnormal institutional holding over the event week. To estimate abnormal institutional holding, we remove firm-specific systematic time trends and purge the variation in institutional holding associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of *ABIT* and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

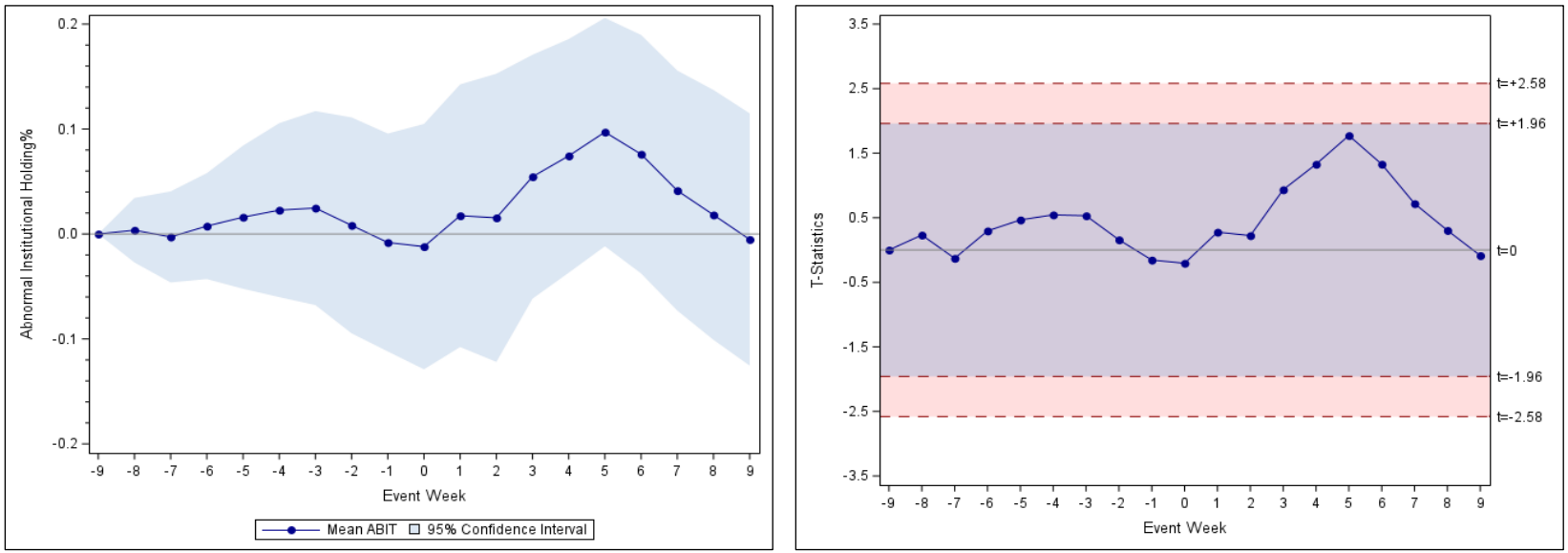
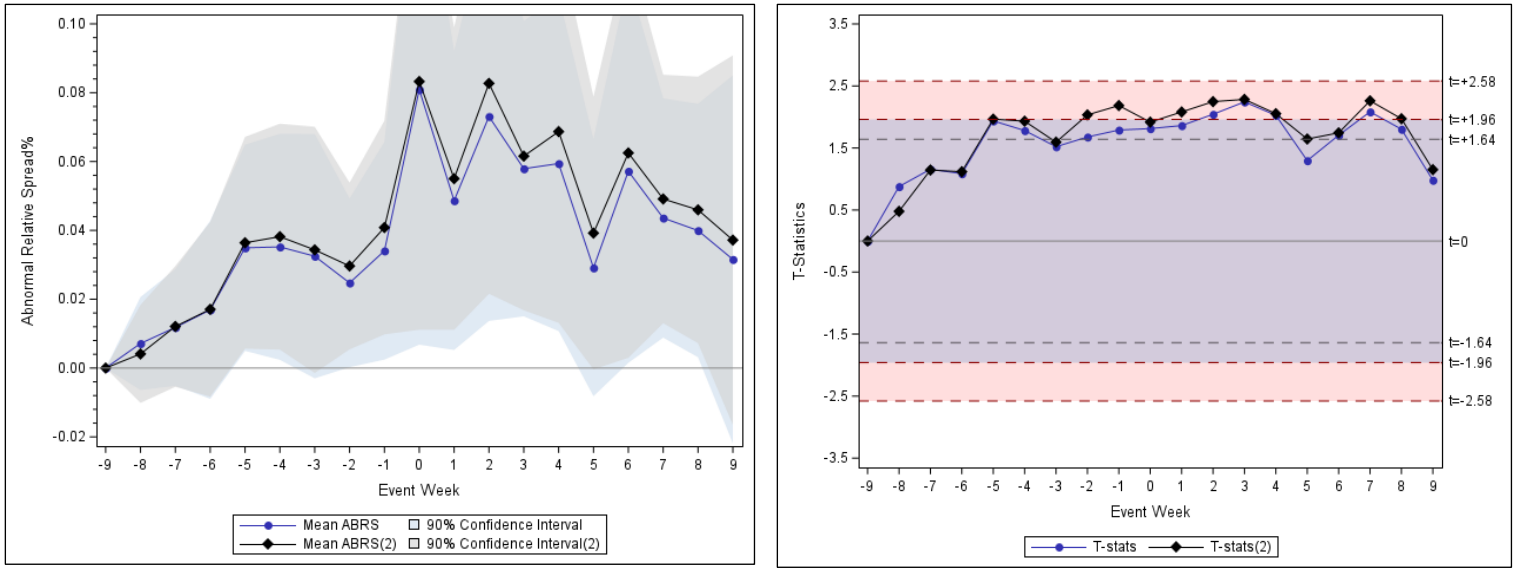


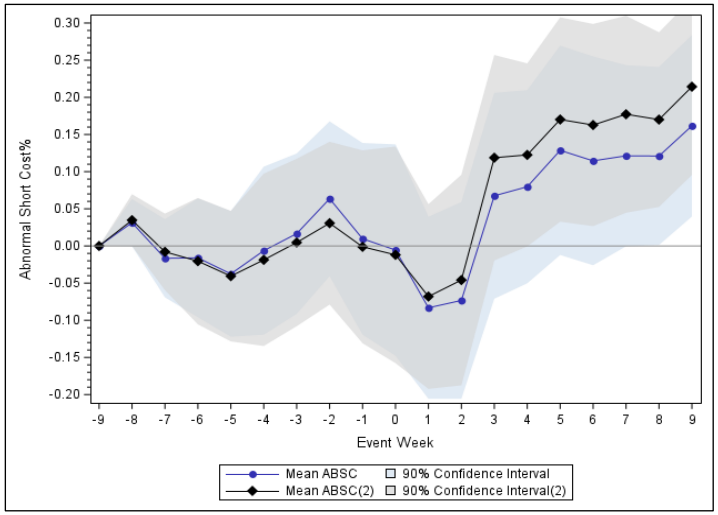
Figure 8. Abnormal Trading and Short Selling Costs around Breach Start Date

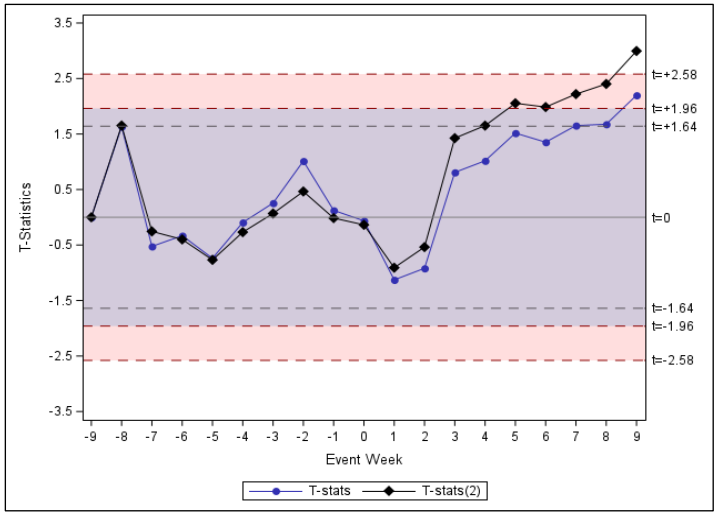
Figure 8 plots the change in the weekly abnormal trading and short selling costs of different market participants of 109 massive data breaches around *Breach Start Date* over the event week window [-9, 9]. Such incidents are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems, leading to the unauthorized access and exfiltration of employee or customer data. We also require that in these data breaches, earnings announcements do not occur within 30 trading days after *Breach Start Date*. Panel A, B and C presents the weekly average of abnormal relative spread, borrowing costs of short sellers, and price impact coefficient of the stock along with their corresponding t-statistics respectively. To estimate the abnormal trading and short selling costs, we remove firm-specific systematic time trends and purge the variation in trading costs associated with *Size*, *Book-to-Market*, *Momentum* and industry fixed effects (*ABRS*, *ABSC*, *ABPI*), and further for *Share Turnover* and *Institutional Ownership* (*ABRS(2)*, *ABSC(2)*, *ABPI(2)*) using a regression-based approach, following Karpoff and Lou (2010). The weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*. Internet Appendix IA1 provides the detailed definitions of abnormal trading cost measures and *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 5% and 1% significance levels.

Panel A Abnormal Relative Spread

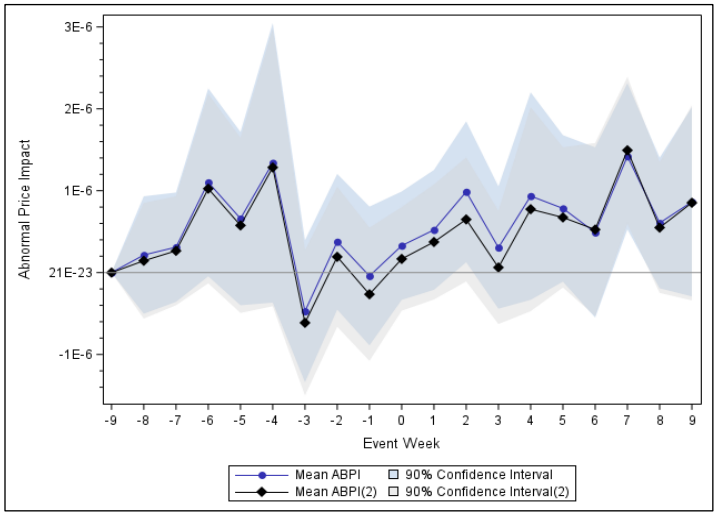


Panel B Abnormal Short Cost





Panel C Abnormal Price Impact



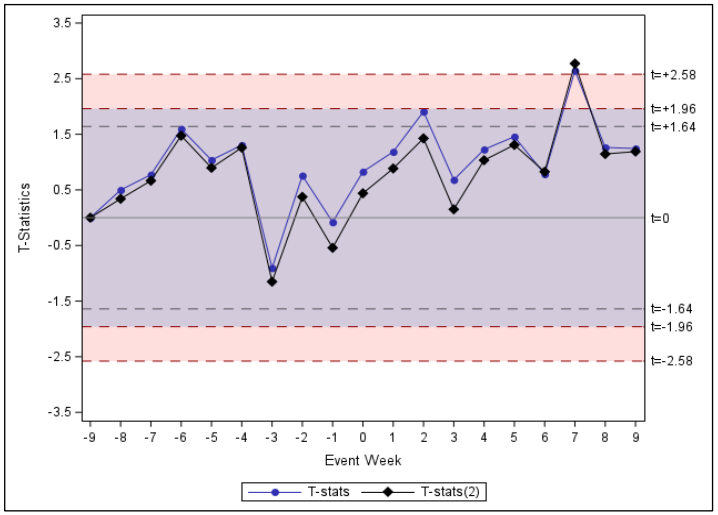
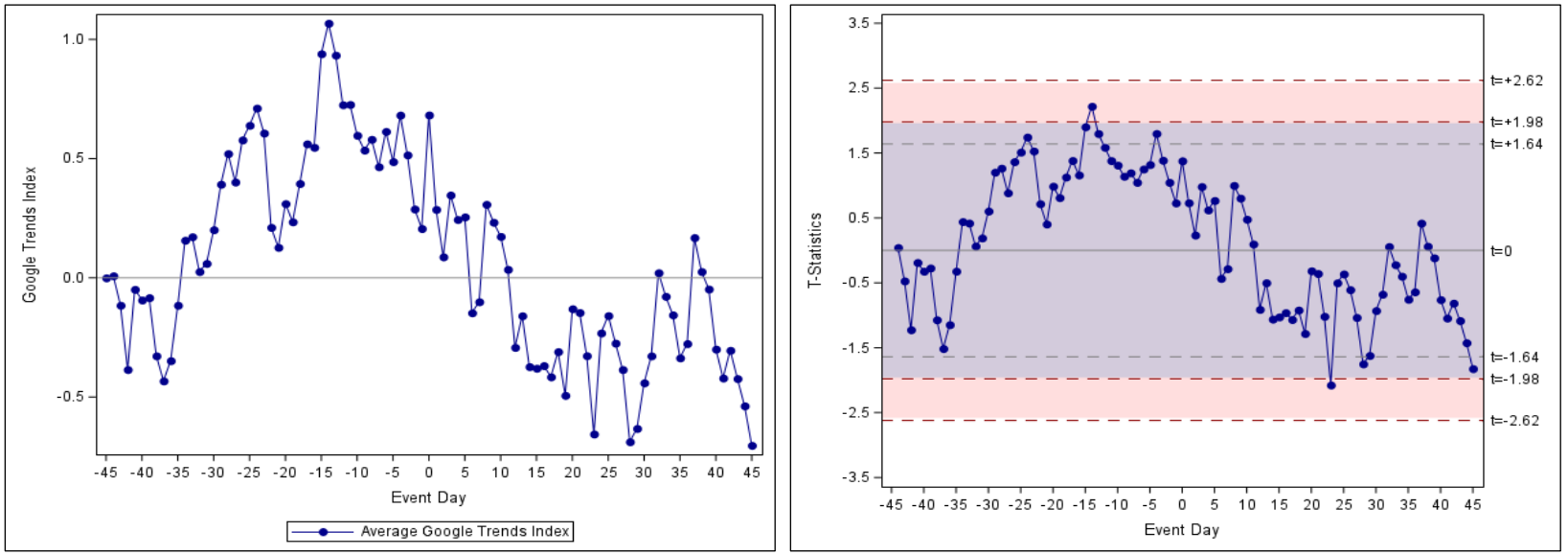


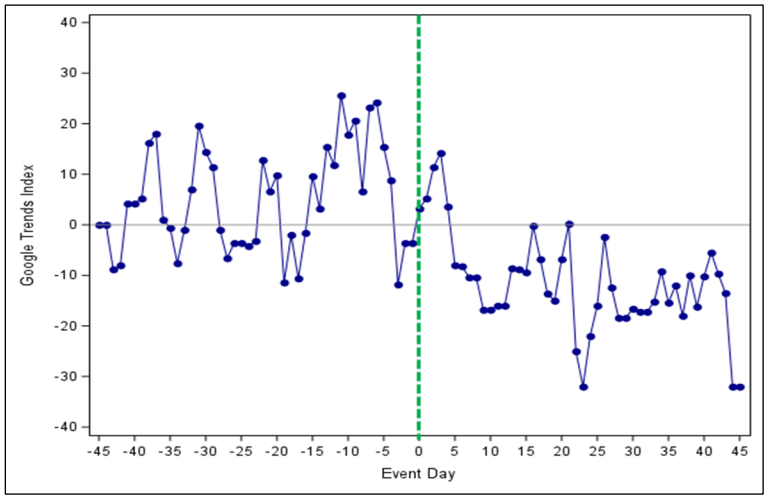
Figure 9. Abnormal Google Search around Breach Start Date

Figure 9 plots the change in daily abnormal google search, measured by Google Trends index, around *Breach Start Date* over the event day window [-45, 45] (event week [-9, 9]). We obtained a keywork-specific Google Trends index for each incident on each event day by defining the search as web users simultaneously inputting the victim firm's name (<company-name>) along with attack-related keywords ("hack"/ "cyberattack"/ "data breach"), with the "Worldwide" searching scope. Panel A presents the average of incident-level daily Google Trends Index, which is constructed by averaging the keyword-specific indices for each incident-day combination across three pre-specified keywords, and the corresponding t-statistics of 109 massive data breaches around *Breach Start Date*. Panel B shows the change in Google search for the four notable data breach incidents over the same event window. We normalize the daily search index to zero at the start of the event window, and use a five-day moving average of daily index to attenuate volatility in the daily measure for clearer graphical representation. Internet Appendix IA1 provides the detailed definitions of *Breach Start Date*. The dashed lines in the t-statistic figure indicate the critical values at the 10%, 5% and 1% significance levels.

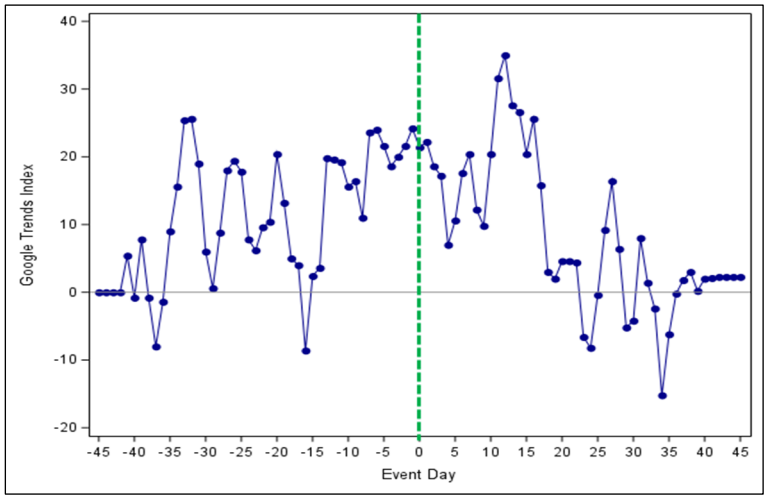
Panel A Abnormal Google Search for Massive Data Breach with Search Keywords <Company Name> + "hack"/ "cyberattack"/ "data breach"



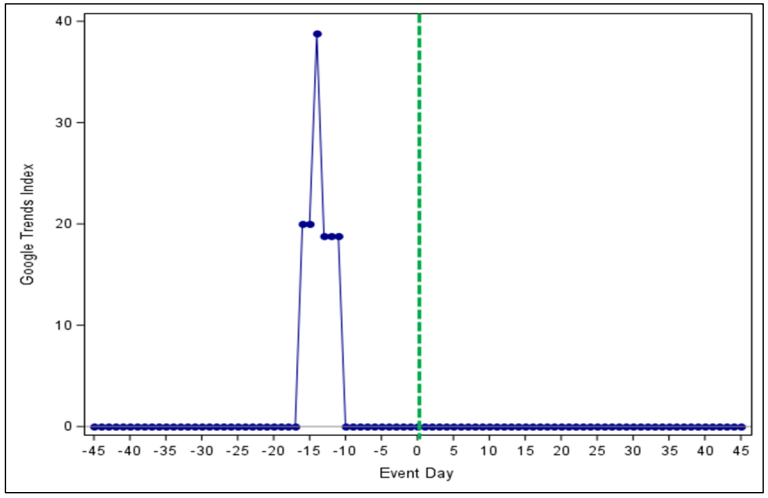
Panel B Notable Examples of Abnormal Google Search around Breach Start Date



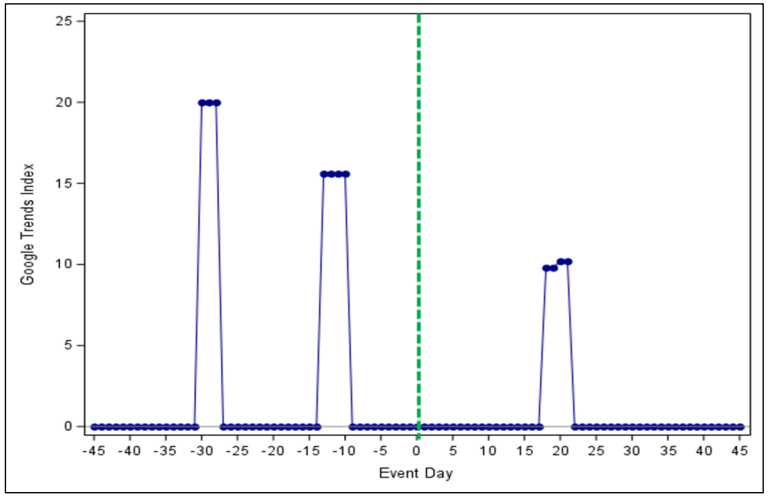
9B1. Search Keyword "Yahoo"+ "Hacking"



9B2. Search Keyword "eBay"+ "Hack"



9B3. Search Keyword "Equifax" + "Cyberattack"

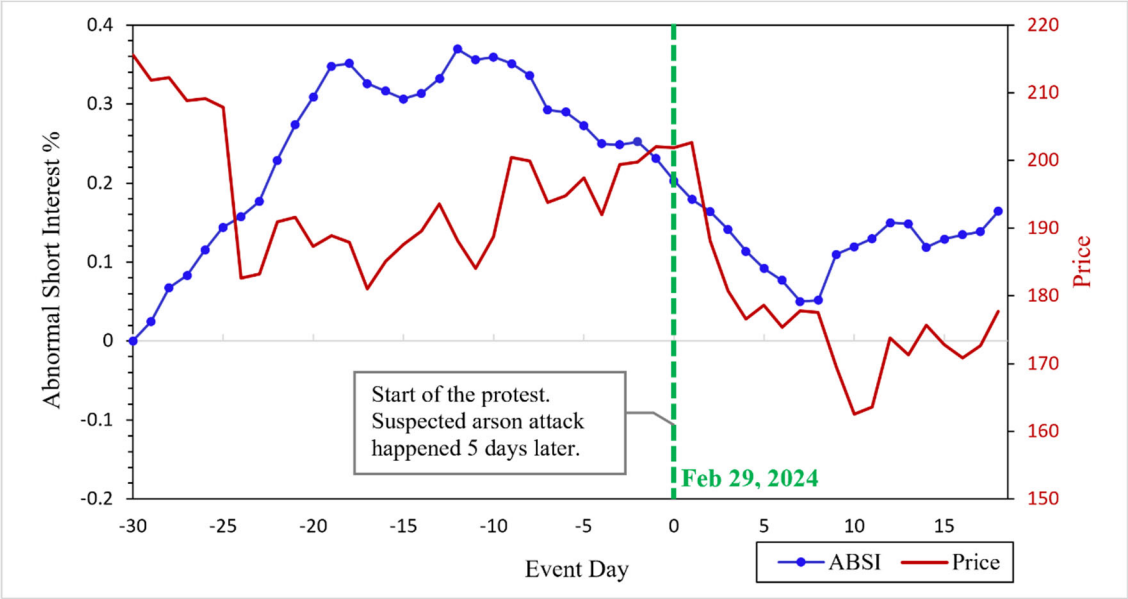


9B4. Search Keyword "Global Payments" + "Hack"

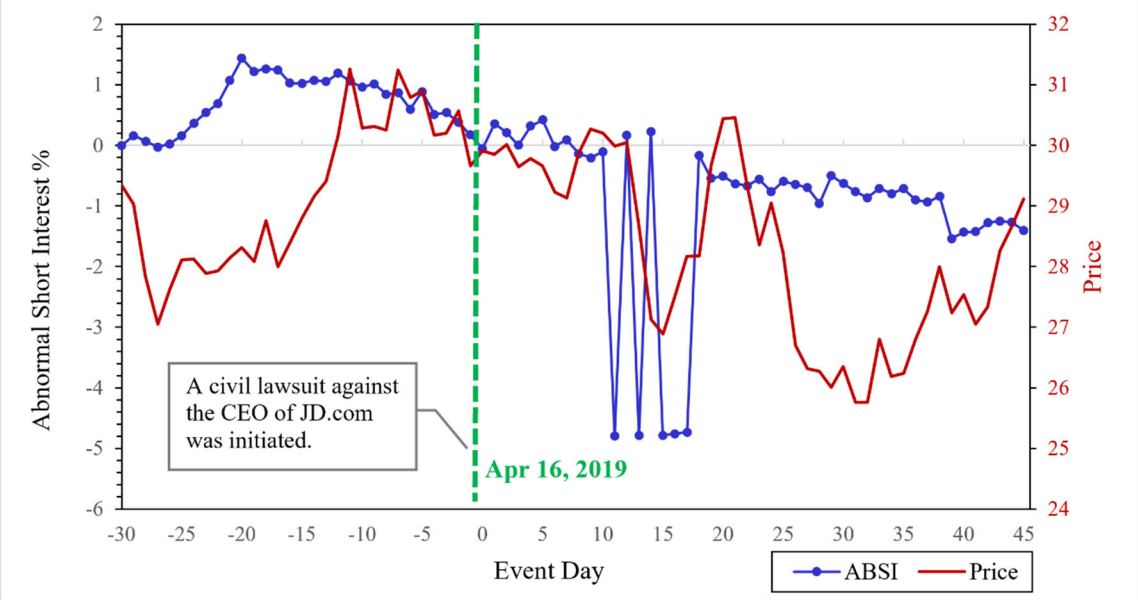
Figure 10. Abnormal Short Interest around the Start of Other Outsider Attacks

Figure 10 explores the change in the daily abnormal short interest (*ABSI*) and stock price around the starts of three recent noted attacks on corporate investor confidence, operations, and ethical reputation. We remove firm-specific systematic time trends to estimate abnormal short interest. We normalize the daily abnormal short interest to zero at the start of the event window. For better presentation, we use the five-day moving average of the abnormal short interest, except for the case of JD, for which we use the raw abnormal short interest measure without smoothing. Day 0 is the earliest identified initiation date of the event (or the first trading day succeeding the date). Other relevant event dates are also indicated in the figures for reference.

Panel A *Operational Disruption*: Tesla Stock Price and Short Selling around Environmentalist Protest against Tesla German Factory



Panel B *Reputational Crisis*: JD.Com Stock Price and Short Selling around the Lawsuit against its CEO for his Alleged Sexual Assault



Panel C *Externally Generated Fundamental Shock*: Nvidia Stock Price and Short Selling around DeepSeek Release

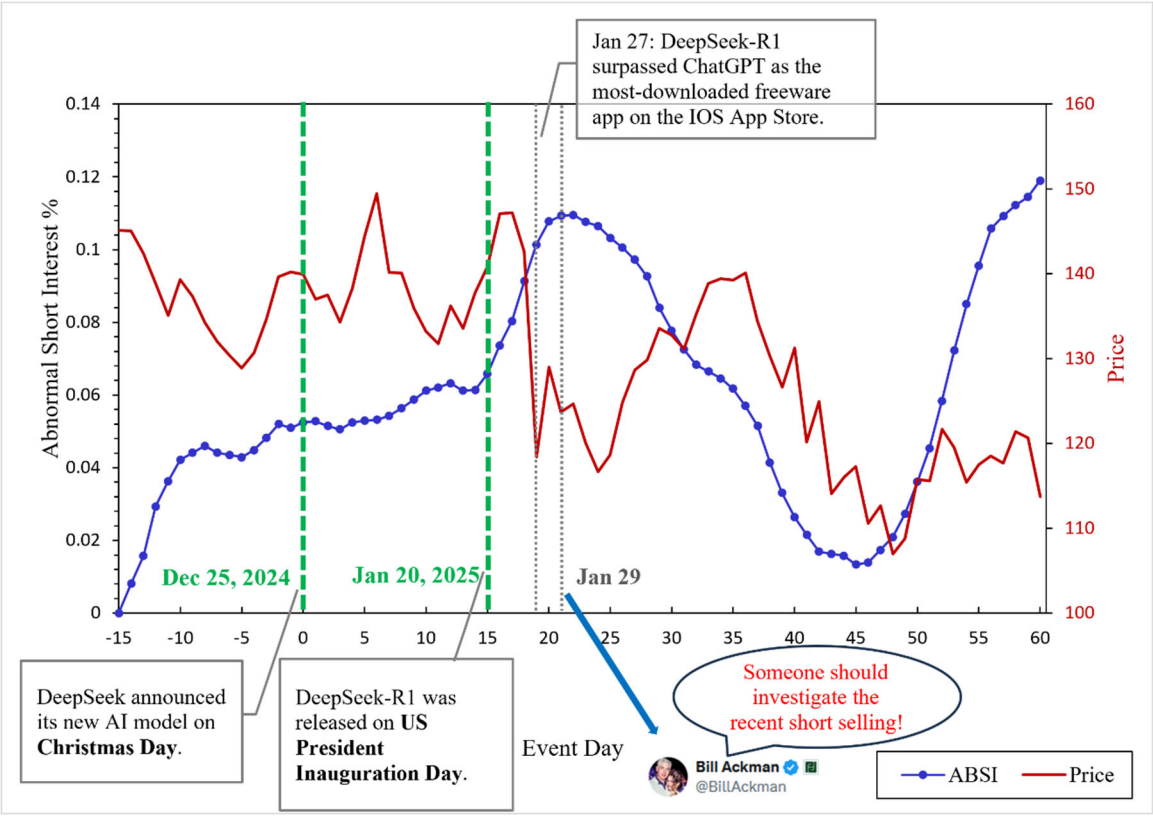


Table 1. Sample Selection

The cybersecurity incidents between 2007 and 2018 are collected from the California Attorney General Website, Audit Analytics Cybersecurity Dataset, and Privacy Rights Clearinghouse's Chronology of Data Breach. We focus on the unique incidents involving public firms that results in the compromise of customer or employee information through cyberattacks, such as hacking intrusions, phishing scams, or local attacks. The massive data breach sample comprises the most consequential incidents initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems. We also require the massive data breaches do not have confounding earnings announcements in the 30 trading-day window following *Breach Start Date*. See Internet Appendix IA2 for the details of our data breach taxonomy.

	#Incidents	#Unique Firms	%S&P500 Market Cap
Unique Cybersecurity Incidents Involving Public Firms from: • The California Attorney General Website "Data Security Breaches" • Audit Analytics Cybersecurity Dataset • Privacy Rights Clearinghouse's Chronology of Data Breaches	527		
Less: Incidents not compromising customer or employee information	(130)		
Total Data Breaches	**397**		
Less: Firms not covered by Compustat, CRSP, Markit, or LSEG databases	(33)		
Less: Data breach(es) lacking verifiable *Breach Disclosure Date*	(1)		
Less: Firms lacking short-selling or stock-trading data for the 45 trading days after *Breach Disclosure Date*	(2)		
Less: Data breach(es) lacking verifiable *Breach Start Date*	(93)		
Less: Firms lacking short-selling or stock-trading data for the 45 trading days prior to *Breach Start Date*	(6)		
Final Data Breach Sample	**262**	210	28%
Less: Incidents resulting from local attacks	(45)		
Less: Firms making earnings announcements within 30 trading days following *Breach Start Date*	(108)		
Massive Data Breach Sample (Attack Sample)	**109**	89	14%

Table 2. Descriptive Statistics

Table 2 provides the descriptive statistics of the variables used in our analyses. Δ*ABSI_PRE*, Δ*ABSI(2)_PRE*, and Δ*ABSI(3)_PRE* are the changes in abnormal short interest *ABSI*, *ABSI(2)*, and *ABSI(3)* during the pre-start period that covers nine weeks before *Breach Start Date*. We remove firm-specific systematic time trends and control for *Size*, *Book-to-Market*, *Momentum* and industry fixed effects to estimate *ABSI*; additionally for *Share Turnover* and *Institutional Ownership* to estimate *ABSI(2)*; and further for *Total Accruals* and *Insider Trading* to estimate *ABSI(3)* using a regression-based approach, following Karpoff and Lou (2010). *Massive Data Breaches* is a dummy variable equal to 1 if the data breach is initiated by criminals through a phishing scam or hacking intrusion against victim firms' central systems and meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*. *Breach Start BHAR* is the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Start Date*. *Breach Disclosure BHAR* is the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Disclosure Date*. *Lender Concentration Ratio* is the concentration ratio of lender value on loan, measured at the beginning of the pre-attack-start period. *Breach Duration* is the logarithm value of the number of days between *Breach Start Date* and *Firm Detection Date*. *Log Size* is the logarithm of market capitalization. *Book-to-Market* is the ratio of equity book value to equity market value. Ret6m and Ret30d are the six-month and 30-day buy-and-hold returns, respectively, prior to *Breach Start Date*. *Share Turnover* is the monthly share turnover of the month prior to *Breach Start Date*. *Institutional Ownership* is the percentage of shares held by institutional investors among outstanding shares, measured at the end of the calendar quarter prior to *Breach Start Date*. *Idiosyncratic Volatility* is the standard deviation of daily idiosyncratic return of the month prior to *Breach Start Date*. *Illiquidity* is the average of the victim firm's daily stock trading illiquidity, measured in the month prior to *Breach Start Date*. Internet Appendix IA1 provides the detailed variable definitions.

Variable	#Obs	Mean	Median	STD	P25	P75
ΔABSI_PRE	262	0.212	0.114	1.482	-0.375	0.562
ΔABSI(2)_PRE	262	0.207	0.109	1.485	-0.387	0.587
ΔABSI(3)_PRE	262	0.212	0.098	1.489	-0.391	0.613
Massive Data Breaches	262	0.416	0.000	0.494	0.000	1.000
Breach Start BHAR	262	-0.001	-0.006	0.117	-0.053	0.048
Breach Disclosure BHAR	262	-0.010	-0.012	0.092	-0.061	0.039
Lender Concentration Ratio	262	0.303	0.260	0.165	0.196	0.347
Breach Duration	262	3.322	3.663	1.814	2.079	4.745
Log Size	262	8.738	8.771	1.933	7.480	10.263
Book-to-Market	262	0.517	0.410	0.459	0.210	0.743
Ret6m	262	0.076	0.078	0.220	-0.040	0.186
Ret30d	262	0.005	0.009	0.095	-0.042	0.057
Share Turnover	262	0.190	0.148	0.143	0.096	0.255
Institutional Ownership	262	0.717	0.790	0.273	0.615	0.895
Idiosyncratic Volatility	262	0.016	0.012	0.010	0.009	0.020
Illiquidity	262	0.007	0.000	0.024	0.000	0.001

Table 3. Short Selling before Breach Start Date

Table 3 reports the OLS regression results about the determinants of short-selling activities before *Breach Start Date*. The sample comprises 262 data breaches which compromise customer or employee information through hacking intrusions, phishing scams, or local attacks. Δ*ABSI_PRE, ΔABSI(2)_PRE, and ΔABSI(3)_PRE* are the changes in abnormal short interest during the pre-start period which covers nine weeks before *Breach Start Date*. *Massive Data Breaches* is a dummy variable equal to 1 if the data breach is initiated by criminals through a phishing scam or hacking intrusion against victim firms' central systems and meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*. *Breach Start BHAR* is the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Start Date*. All the variables are defined in Internet Appendix IA1. Robust t-statistics are reported in parentheses. *, **, and *** indicate two-tailed statistical significance of 10%, 5%, and 1% levels respectively.

	(1) ΔABSI_PRE	(2) ΔABSI(2)_PRE	(3) ΔABSI(3)_PRE
Massive Data Breaches	**0.345****	**0.348****	**0.329***
	(1.97)	**(1.99)**	**(1.88)**
Breach Start BHAR	**-2.560*****	**-2.678*****	**-2.693*****
	(-3.04)	**(-3.26)**	**(-3.28)**
Breach Disclosure BHAR	1.206	1.152	1.192
	(1.05)	(1.00)	(1.03)
Breach Duration	-0.005	-0.013	-0.011
	(-0.09)	(-0.26)	(-0.22)
Log Size	-0.007	-0.023	-0.022
	(-0.14)	(-0.46)	(-0.43)
Book-to-Market	0.171	0.120	0.132
	(0.72)	(0.50)	(0.54)
Ret6m	-0.485	-0.486	-0.519
	(-0.86)	(-0.87)	(-0.93)
Ret30d	0.634	0.541	0.554
	(0.69)	(0.57)	(0.58)
Share Turnover	2.972***	2.939***	2.967***
	(2.81)	(2.72)	(2.72)
Institutional Ownership	-0.132	-0.105	-0.132
	(-0.34)	(-0.27)	(-0.34)
Idiosyncratic Volatility	-9.564	-10.614	-10.629
	(-0.78)	(-0.86)	(-0.85)
Illiquidity	-0.131	-0.565	-0.586
	(-0.04)	(-0.17)	(-0.18)
Constant	-0.218	-0.022	-0.016
	(-0.35)	(-0.04)	(-0.03)
#Observations	262	262	262
R-Squared	0.116	0.117	0.118

Table 4. Lender Concentration Effect

Table 4 reports the OLS regression results about the lender concentration effect on short-selling activities before *Breach Start Date*. The sample comprises 262 data breaches which compromise customer or employee information through hacking intrusions, phishing scams, or local attacks. Δ*ABSI_PRE, Δ ABSI(2)_PRE, and* Δ*ABSI(3)_PRE* are the changes in abnormal short interest during the pre-start period which covers nine weeks before *Breach Start Date*. *Massive Data Breaches* is a dummy variable equal to 1 if the data breach is initiated by criminals through a phishing scam or hacking intrusion against victim firms' central systems and meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*. *Lender Concentration Ratio* is the concentration ratio of lender value on loan measured at the beginning of the pre-attack-start period (the beginning of week -9 relative to *Breach Start Date*). A higher concentration ratio indicates that the sources of shares sold short concentrate in fewer lenders. The control variables include *Log Size*, *Book-to-Market*, and *Ret6m*, *Share Turnover*, *Institutional Ownership*, *Idiosyncratic Volatility*, and *Illiquidity*. All the variables are defined in Internet Appendix IA1. Robust t-statistics are reported in parentheses. *, **, and *** indicate two-tailed statistical significance of 10%, 5%, and 1% levels respectively.

	(1) ΔABSI_PRE	(2) ΔABSI(2)_PRE	(3) ΔABSI(3)_PRE
Massive Data Breaches	0.918**	0.905**	0.864**
	(2.50)	(2.46)	(2.34)
Lender Concentration Ratio	0.887	0.869	0.790
	(1.62)	(1.59)	(1.43)
Massive Data Breaches* Lender Concentration Ratio	**-1.912****	**-1.853****	**-1.784****
	(-2.16)	**(-2.11)**	**(-2.02)**
Breach Initiation BHAR	-2.389***	-2.511***	-2.527***
	(-2.84)	(-3.02)	(-3.05)
Breach Disclosure BHAR	1.137	1.082	1.128
	(1.00)	(0.95)	(0.99)
Breach Duration	-0.012	-0.020	-0.018
	(-0.24)	(-0.40)	(-0.35)
Constant	-0.508	-0.304	-0.274
	(-0.78)	(-0.47)	(-0.41)
Control Variables	Yes	Yes	Yes
#Observations	262	262	262
R-Squared	0.125	0.126	0.125

Table 5. Market Reaction to Data Breach Start

Table 5 reports the univariate analyses of the market reaction to the start of data breach. The sample comprises 109 massive data breaches and the other data breaches. Massive data breaches are initiated by criminals through phishing scams or hacking intrusions against victim firms' central systems and meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*. Other data breaches consist of local attacks and the incidents whose breach start is confounded with earnings announcements. Δ*ABSI_PRE, ΔABSI(2)_PRE, and ΔABSI(3)_PRE* are the changes in abnormal short interest during the pre-start period which covers nine weeks before *Breach Start Date*. For each type of abnormal short interest, the sample is divided into two groups in accordance with the direction of pre-start short interest change. *Breach Start BHAR* is the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Start Date*. The table presents the group average stock returns which are equal-weighted *Breach Start BHAR*. Robust t-statistics are reported after average stock returns. *, **, and *** indicate two-tailed statistical significance of 10%, 5%, and 1% levels respectively.

Breach Start BHAR (Equal-Weighted Average)

Short Interest Change	#Obs	Massive Data Breaches	t-stat	#Obs	Other Data Breaches	t-stat
ΔABSI_PRE > 0	**72**	**-0.032****	**(-2.56)**	76	-0.006	(-0.43)
ΔABSI_PRE <= 0	37	0.020	(0.98)	77	0.023*	(1.80)
Difference		**-0.052****	**(-2.17)**		-0.029	(-1.54)
ΔABSI(2)_PRE > 0	**69**	**-0.034****	**(-2.60)**	75	-0.003	(-0.20)
ΔABSI(2)_PRE <= 0	40	0.019	(0.98)	78	0.020	(1.50)
Difference		**-0.053****	**(-2.26)**		-0.030	(-1.60)
ΔABSI(3)_PRE > 0	**68**	**-0.034****	**(-2.59)**	75	-0.007	(-0.48)
ΔABSI(3)_PRE <= 0	41	0.018	(0.96)	78	0.024*	(1.87)
Difference		**-0.052****	**(-2.26)**		-0.030	(-1.60)

Table 6. Pseudo Breach Start Date

Table 6 reports the analyses of pseudo breach start date. The sample is based on 109 massive data breaches which are initiated by criminals through phishing scams or hacking intrusions. Meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*. Pseudo breach start date is randomly drawn from the non-event day windows [-250, -46] and [46, 250] around *Breach Start Date* with replacement. Panel A, B, and C report the regression results of 100 times simulation, 200 times simulation, and 300 times simulation respectively. Δ*ABSI_PRE, ΔABSI(2)_PRE, and ΔABSI(3)_PRE* are the changes in abnormal short interest during the pre-start period which covers nine weeks before pseudo or real breach start dates. *Actual* is a dummy variable equal to 1 if the observation reflects the real breach start date while equal to 0 for pseudo events. All the variables are defined in Internet Appendix IA1. Robust t-statistics are reported in parentheses. *, **, and *** indicate two-tailed statistical significance of 10%, 5%, and 1% levels respectively.

Panel A 100 Times Simulation			
	(1) ΔABSI_PRE	(2) ΔABSI(2)_PRE	(3) ΔABSI(3)_PRE
Actual	**0.369***** **(2.72)**	**0.366***** **(2.72)**	**0.362***** **(2.69)**
Constant	-0.012 (-0.89)	-0.013 (-0.91)	-0.015 (-1.06)
#Observations	11009	11009	11009
R-Squared	0.001	0.001	0.001

Panel B 200 Times Simulation			
	(1) ΔABSI_PRE	(2) ΔABSI(2)_PRE	(3) ΔABSI(3)_PRE
Actual	**0.360***** **(2.65)**	**0.357***** **(2.64)**	**0.354***** **(2.61)**
Constant	-0.000 (-0.02)	0.001 (0.10)	-0.003 (-0.34)
#Observations	21909	21909	21909
R-Squared	0.000	0.000	0.000

Panel C 300 Times Simulation			
	(1) ΔABSI_PRE	(2) ΔABSI(2)_PRE	(3) ΔABSI(3)_PRE
Actual	**0.363***** **(2.67)**	**0.358***** **(2.65)**	**0.356***** **(2.63)**
Constant	-0.002 (-0.24)	-0.000 (-0.05)	-0.005 (-0.64)
#Observations	32809	32809	32809
R-Squared	0.000	0.000	0.000

Internet Appendix IA. Variable Definition, Data Collection, and Robustness Tests

IA1. Variable Definition

1. Data Breach Variables

- *Breach Start Date*: the date when the victim firm's cyber network was intruded or when hackers started to collect customer or employee data stored by the victim firm (see Internet Appendix IA3 for details).
- *Firm Detection Date*: the date when the data breach was exposed to the victim firm's management (see Internet Appendix IA3 for details).
- *Breach Disclosure Date*: the date when the data breach was publicly announced or confirmed by the victim firm (see Internet Appendix IA3 for details).
- *Breach Duration*: the logarithm value of the number of days between *Breach Start Date* and *Firm Detection Date. Breach Disclosure Date* is used to substitute for missing *Firm Detection Date*.
- *Massive Data Breaches*: a dummy variable equal to 1 if the data breach is initiated by criminals through a phishing scam or hacking intrusion and meanwhile, no earnings announcement occurs within 30 trading days after the data breach's *Breach Start Date*.
- *Breach Start BHAR*: the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Start Date*, calculated as

$$Breach\ Start\ BHAR_i = \prod_{d=0}^{30}(1 + Ret_{id}) - \prod_{d=0}^{30}(1 + dgtwRet_d),$$

  where Ret_{id} is the daily stock return of firm *i* on event day *d* and $dgtwRet_d$ is the daily value-weighted average return of the benchmark portfolio. Following Daniel, Grinblatt, Titman, and Wermers (1997), we construct the benchmark portfolio for each victim firm based on cross-sectional quintiles of market capitalization, book-to-market, and prior 12-month compound returns.
- *Breach Disclosure BHAR*: the DGTW-adjusted buy-and-hold abnormal return over the event day window [0, 30] surrounding *Breach Disclosure Date*.

2. Short Selling

- $ABSI, ABSI(2), ABSI(3)$: the weekly abnormal short interest (in percentage points) estimated by three regression models in Karpoff and Lou (2010). Weekly abnormal short interest is the weekly average of daily abnormal short interest over the event week.[35] To estimate daily abnormal short interest, we first remove firm-specific systematic time trends by subtracting the average short interest over the window [*d*-90, *d*-1] from trading day *d*'s raw short interest based on the following formula:[36]

$$\text{D}etrended\ SI_t = SI_t - \frac{SI_{t-1} + SI_{t-2} + \cdots + SI_{t-90}}{90},$$

  Then, we estimate the abnormal short interest with the detrended short interest using a regression-based approach. Following Karpoff and Lou (2010) Model 1, we first sort the universe of companies into terciles by *Size*, *Book-to-Market*, and *Momentum* respectively, and create

[35] If we focus on the breach start event, event day 0 denotes Breach Start Date or the first trading day after Breach Start Date. The weekly abnormal short interest of week 0 is the average of daily abnormal short interest over the event day window [0, 4]

[36] Raw short interest is defined as the percentage of shares sold short (total quantity of stock on loan) over shares outstanding.

indicator variables for each group.[37] Next, we regress the detrended daily short interests of all-non victim firms on these indicators and further control for firms' industry classifications to estimate coefficients, which we use to predict the short interest for the victim firms for each trading day *d*.[38] Victim firm *i*'s daily abnormal short interest $ABSI_{id}$ is the difference between its actual detrended short interest and its predicted short interest from the model on trading day *d*. The regression model used for estimation and prediction is shown below.

$$SI_{id} = \sum_{g=low}^{medium} s_{gd} Size_{igd} + \sum_{g=low}^{medium} b_{gd} Book - to - Market_{igd} + \sum_{g=low}^{medium} m_{gd} Momentum_{igd} + \sum_{k=1}^{K} \phi_{kd} Ind_{ikd} + u_{id},$$

$ABSI(2)_{id}$ and $ABSI(3)_{id}$ are constructed analogously, with regression KL Model 2 adding *Share Turnover* and *Institutional Ownership,* and model 3 further adding *Total Accruals* and *Insider Trading* to estimate the abnormal short interest for each victim firm. *Insider Trading* is defined as the difference between net insider selling on the current day and its average over the previous 30 trading days.

Next, we construct weekly abnormal short interest measures by taking the average of daily abnormal short interests over the event week. Event week 0 covers the event day window [0, 4] in which day 0 is the event date or the first trading day after the event date.

- $\Delta ABSI\_PRE$, $\Delta ABSI\_PRE(2)$, $\Delta ABSI\_PRE(3)$: change in abnormal short interest $ABSI$, $ABSI(2)$, and $ABSI(3)$ over the pre-start period which covers nine weeks before *Breach Start Date*.
- $ABSV$: the weekly abnormal short volume (in percentage points) constructed by removing firm-specific systematic time trends and purging the variation associated with *Size*, *Book-to-Market*, and *Momentum*. We first obtain the transaction-level short trading information from FINRA Short Sale Volume Data and aggregate the trades to the daily level to estimate daily short volume. Then, we construct the weekly abnormal measure using a detrending procedure and a Karpoff-Lou regression-based approach, analogously to the construction of abnormal short interest.
- $ABRSV$: the weekly abnormal retail short volume (in percentage points) constructed by removing firm-specific systematic time trends and purging the variation associated with *Size*, *Book-to-Market*, and *Momentum*. We first obtain the transaction-level short trading information from FINRA, identify trades likely initiated by retail short sellers with the methodology proposed by Boehmer and Song (2020), and aggregate these trades to the daily level to estimate daily retail short volume. Then, we construct the weekly abnormal measure using a detrending procedure and a Karpoff-Lou regression-based approach, analogously to the construction of abnormal short interest.
- *Lender Concentration Ratio*: the concentration ratio measured at the beginning of the pre-attack-start period, which is the beginning of week -9 relative to *Breach Start Date*. This measure represents the distribution of lender value on loan. A higher concentration ratio indicates that shares sold short come from a smaller number of lenders.

3. Retail Holdings, Institutional Holdings, and Insider Holdings

- $ABRH$: the weekly abnormal retail holdings (in percentage points) constructed by removing firm-

[37] We sort all companies with non-missing data from Compustat, CRSP, Markit, and Thomson Reuters.
[38] Non-victim firms did not experience data breaches during the sample period.

specific systematic time trends and purging the variation associated with *Size*, *Book-to-Market*, and *Momentum*. We first obtain the transaction-level trading information from TAQ data, identified retail buys and sells with the methodology proposed by Barber et al.(2024), and aggregate these trades to calculate retail net buys at the daily level (RNB_d).[39]Then, we calculate the daily detrended retail holdings ($Detrended\ RH_d$) with the detrend window [d-90, d-1] based on the following formula:

$$Detrended\ RH_t = RH_t - \frac{RH_{t-1} + RH_{t-2} + \cdots + RH_{t-90}}{90}$$

$$= (RH_t - RH_{t-1})$$

$$+ \frac{89 \times (RH_{t-1} - RH_{t-2}) + 88 \times (RH_{t-2} - RH_{t-3}) + \cdots + 1 \times (RH_{t-89} - RH_{t-90})}{90}$$

$$= RNB_t + \frac{89}{90} \times RNB_{t-1} + \frac{88}{90} \times RNB_{t-2} + \cdots + \frac{1}{90} \times RNB_{t-89}.$$

Next, we construct the weekly abnormal retail holding with this detrended daily measure using Karpoff and Lou (2010)'s regression approach, which is analogously to the way to the construction of abnormal short interest.

- $ABIH$: the weekly abnormal insider holdings (in percentage points). First, we obtain insider trading data from LSEG database (formerly Thomson/Refinitiv) and identify opportunistic insider trades using Cohen et al. (2012)'s trader-based classification. Then, we aggregate all opportunistic insider buys and sells to calculate daily insider net buys. Next, we follow the same procedure used to measure weekly abnormal retail holdings to compute weekly abnormal insider holdings from daily opportunistic insider net buys.
- $ABIT$: the weekly abnormal institutional holdings (in percentage points). The calculation of abnormal institutional holdings is identical to that of abnormal retail holdings and insider holdings. We use institutional net buys to construct detrended institutional holdings. All the institutional trades exceed 20,000 dollars in value.

4. Trading Costs

- $ABSC, ABSC(2),$: the weekly abnormal short seller borrowing cost constructed by removing firm-specific systematic time trends and purging the variation associated with *Size*, *Book-to-Market*, and *Momentum* ($ABSC$), and further the variation associated with *Share Turnover* and *Institutional Ownership* ($ABSC(2)$) using a regression-based approach suggested by Karpoff and Lou (2010). The construction of this variable follows the same procedure as that of abnormal short interest. Short cost data is obtained from Markit.
- $ABRS, ABRS(2)$: the weekly abnormal relative spread (in percentage point), constructed identically to $ABSC$. Relative spread is calculated as the bid-ask spread divided by the midpoint of the bid and ask prices.
- $ABPI, ABPI(2)$: the weekly abnormal price impact coefficient, constructed identically to $ABSC$.

5. Control Variables

[39] Retail net buys equal the difference between retail buys and retail sells. Retail holding is the percentage of retail investors' shares among outstanding shares. The same calculation applies to institutional net buys and institutional holding used to compute ABIT.

- *Log Size*: the logarithm of market capitalization. Market capitalization is measured on the last trading day of the month prior to *Breach Start Date*.
- *Book-to-Market*: the ratio of equity book value to equity market value. The book value is measured at the end of the fiscal quarter prior to *Breach Start Date*, and the market value is measured on the last trading day of the month prior to *Breach Start Date*.
- *Ret6m*: the buy-and-hold return for the six-month period ending in the month prior to *Breach Start Date*.
- *Ret30d*: the DGTW-adjusted buy-and-hold abnormal return for the 30-day period before *Breach Start Date* (event day window [-30, -1]).
- *Share Turnover*: the monthly share turnover of the month prior to *Breach Start Date*.
- *Institutional Ownership*: the percentage of shares held by institutional investors over shares outstanding, measured at the end of the calendar quarter prior to *Breach Start Date*.
- *Idiosyncratic Volatility*: the standard deviation of daily idiosyncratic return in the month prior to *Breach Start Date*. Daily idiosyncratic return is the difference between the victim firm's daily stock return and the CRSP daily value-weighted market return.
- *Illiquidity*: the average of the victim firm's daily stock trading illiquidity, measured in the month prior to *Breach Start Date*. The calculation of daily stock trading illiquidity is shown below.

$$Stock\ Trading\ Illiquidity_{id} = \frac{abs(Ret_{id})}{Dollar\ Volume_{id}}$$

where Ret_{id} is the daily stock return of firm i on day d and $Dollar\ Volume_{id}$ is the daily dollar trading volume of firm i on day d.

IA2. Data Breach Taxonomy

The California Attorney General website, Audit Analytics Cybersecurity Dataset, and Privacy Rights Clearinghouse Chronology provide a variety of data breaches that were disclosed between 2007 and 2018.[40] To identify the hacker-initiated data breaches which compromise the customer or employee information of U.S. public firms, we classify all data breaches into following eleven types.

- **Type 1: Hacking intrusions compromising the customer or employee information**
  (1) *Description*: In these data breaches, hackers gained unauthorized access to the customer or employee information by intruding into public firms' computer networks, websites, or systems etc. Hackers usually install malicious codes to initiate attacks. Since these sophisticated data breaches have high likelihood of undermining victim firms' operation, we include them in our sample.
  (2) *Example*: Target Corporation was attacked by hackers on November 15, 2013. The incident led to the exposure of 40 million customers' payment card data and 70 million customers' personal information. Target disclosed this incident on December 19, 2013. After learning Target breach, the impacted customers launched a class action lawsuit in 2014. Target recorded $184 million as its pretax data breach-related costs net of insurance.
- **Type 2: Phishing scams compromising the customer or employee information**
  (1) *Description*: In these data breaches, criminals impersonated as a public firm's senior executives to request the customer or employee information by emails. A common target of phishing attacks is employees' W-2 tax forms. Our sample includes these sophisticated data breaches because they are inclined to significantly increase victim firms' operating expense.
  (2) *Example*: On March 14, 2016, a criminal sent to the payroll department of Sprouts Farmers

[40] See the details of California Attorney General website. Among the data breaches provided by Privacy Rights Clearinghouse Chronology, we focus on the incidents of breach type "HACK" and business types "BSF," "BSO," "BSR," or "MED."

Market an email that pretended to be sent by the company's senior executive. Sprouts Farmers Market then exposed the 2015 W-2 tax forms of its employees, some of whom later sued the company for the compromise of their personal and financial information.

- **Type 3: Local breaches compromising the customer or employee information**
  (1) *Description*: In these data breaches, criminals attacked some retail locations of a public firm, which can be a store, a restaurant, or a gas station etc. Some criminals installed malware on certain locations' point-of-sale system to extract customers' payment card data, whereas some hacked into the computers used at specific locations. However, criminals did not breach the central system of victim firms. Our sample includes these data breaches because the customer or employee information was compromised.
  (2) *Example*: On February 3, 2017, InterContinental Hotels Group ("IHG") disclosed that some criminals had installed malware on the point-of-sale system of 12 properties managed by the company. As a result, certain customers' payment card information was stolen.
- **Type 4: Hacking intrusions NOT compromising the customer or employee information**
  (1) *Description*: In these incidents, although hackers attacked public firms' networks, websites, or systems etc., they did not intend to steal the customer or employee information. Instead, they aimed to disrupt victim firms' operation, intimidate target companies, or achieve other goals which did not involve the customer or employee information. Our sample excludes these incidents.
  (2) *Example*: A group of hackers commenced a distributed denial of service attack against the system of Microsoft Xbox Live on December 24, 2014. The attack caused millions of people to be unable to play game consoles for a few days. The incentives of the hackers were allegedly to manifest the weaknesses of Xbox Live system.
- **Type 5: Phishing scams NOT compromising the customer or employee information**
  (1) *Description*: In these incidents, criminals impersonated as a public firm's senior executive to request a transfer of money to their own bank accounts. No customer or employee information was compromised. Our sample excludes these incidents.
  (2) *Example*: On April 30, 2015, a hacker pretended to be Mattel CEO and sent to the company's finance executive a note which demanded a vendor payment. $3 million were transferred to the hacker's bank account. This phishing scam did not expose customer or employee data.
- **Type 6: Data breaches denied by companies**
  (1) *Description*: In these data breaches, hackers claimed that they had accessed certain data by attacking a public firm's networks, websites, or systems etc., but the company denied their claim. Given the dispute, our sample excludes these data breaches.
  (2) *Example*: According to a report on November 2, 2018, some hackers claimed that they had obtained 120 million accounts of Facebook users. Facebook insisted that its security had not been breached and the hackers had collected only 81,000 users' private messages by deploying malicious browser extensions on users' own personal computers.
- **Type 7: Security vulnerability exposure**
  (1) *Description*: In these events, public firms had security vulnerabilities which might result in a compromise of customer or employee information. These companies acknowledged the risk of entailing data exposure but did not confirm that data had been exposed due to the security vulnerability. Our sample does not include these events.
  (2) *Example*: On November 2, 2016, Cisco Systems Inc. announced that it had been notified about a security vulnerability of its Professional Careers mobile website. The vulnerability was the consequence of an inappropriate security setting caused by previous system maintenance. In spite of its website's security vulnerability, Cisco did not believe that the user information had been illegally accessed.
- **Type 8: Automated attacks**
  (1) *Description*: In these data breaches, criminals logged into victim firms' websites or systems by

using the login credentials obtained from other firms' data breaches which had happened beforehand. Using algorithms, the criminals conducted login trials on a large scale, without breaching victim firms' cybersecurity. We exclude these incidents from our sample because they are essentially attributed to other companies' data breaches.

(2) *Example*: From April 26, 2018 to June 12, 2018, some criminals used valid customer credentials to log into Macy's customer accounts. Macy asserted that the credentials had been obtained from a source other than Macy's systems. The suspicious login activities were detected on June 11, 2018 and were reported by Macy's on June 27, 2018.

- **Type 9: Insiders' theft**

(1) *Description*: In these data breaches, the customer or employee information was compromised due to the misconduct of a public firm's current workers, former workers, or contractors. These data breaches are excluded from our sample because public firms' cybersecurity was not breached.

(2) *Example*: On April 21, 2014, L Brands Inc. reported that an employee who worked at a Victoria's Secret store had stolen customer payment information. Instead of intruding into Brands' network, the employee installed a hidden device which scanned customers' credit cards.

- **Type 10: Data breaches happening to firms that later changed their ownership**

(1) *Description*: In these data breaches, when criminals launched attacks, victim firms had not initiated IPOs and were still privately held. Under other circumstances, victim firms were owned by a public firm when hackers initiated attacks but later became the subsidiaries of another public firm before their incidents were disclosed. Our sample excludes these data breaches because they either do not happen to public firms or involve multiple public firms.

(2) *Example*: In July 2014, some criminals started hacking the Starwood guest reservation system. After Starwood was acquired by Marriot on September 23, 2016, Marriot received an alert about an attempt to intrude into the Starwood guest reservation system on September 8, 2018. Marriot disclosed the incident to the public on November 30, 2018.

- **Type 11: Miscellaneous data breaches**

(1) *Description*: In addition to the aforementioned data breaches, the three databases provide other incidents. Some of these incidents did not leak the customer or employee information; some lacked public coverage and as a result, we cannot determine their types; some involved so many public firms, none of which was seriously affected; some did not breach companies' network security; some happened to private firms. These data breaches are not included in our sample.

(2) *Example*: On October 21, 2013, some hackers began collecting user login credentials by installing malware on users' own personal computers. More than 93,000 websites, including the websites operated by Facebook, Google (now Alphabet), Yahoo, Twitter, Automatic Data Processing, and LinkedIn, were affected. Though these companies had thousands of customers' login credentials compromised, their own cybersecurity was not breached.

Among these eleven types of data breaches, our analysis focuses on the first three types of information-related incidents, where criminals illegally obtained the customer or employee information by breaching the cybersecurity of public firms. We choose these incidents because the loss of information tends to incur significant costs (e.g. litigation expense). As Table 1 shows, after deleting duplicate events, our sample comprises 397 unique data breaches.

IA3. Date Collection

For each of these 397 data breaches, we try to collect its *Breach Start Date*, *Firm Detection Date*, and *Breach Disclosure Date*. Our goal is to find the earliest date which can be confirmed by a public information source. We summarize different types of dates below.

1. Breach Start Date

- **Type 1: The first date of the affected shopping period**

(1) *Description*: On such dates, when customers made payment during shopping, hackers were collecting their information from the breached system of victim firms. This type of breach start date is the first date of a shopping period during which hackers continuously extracted customer data.

(2) *Example*: On May 17, 2017, Rite Aid Corporation reported that it had experienced an attack against its e-commerce platform. Hackers obtained payment card data when customers made payment during online shopping between January 30, 2017 and April 11, 2017. We treat January 30, 2017 as the breach start date.

- **Type 2: The date of hacking intrusions**
  (1) *Description*: On such dates, criminals initiated hacking intrusions against a public firm.
  (2) *Example*: Target Corporation experienced a data breach in 2013. According to the lawsuit complaint filed by the impacted financial institutions, between November 15, 2013 and December 15, 2013, hackers gained an illegal access to customer information by installing malware on Target internal network. We use November 15, 2013 as the breach start date.
- **Type 3: The date of phishing scams**
  (1) *Description*: On such dates, a public firm fell victim to a phishing scam that allowed criminals to steal the customer or employee information.
  (2) *Example*: On March 1, 2016, an employee of Seagate Technology Plc sent the 2015 W-2 forms of Seagate employees to hackers in a phishing scam. We regard March 1, 2016 as the breach start date. The phishing scam triggered a class action suit which was filed on April 14, 2016 against Seagate.
- **Type 4: The date provided by the California Attorney General website**
  (1) *Description*: On the California Attorney General website, reporting companies can provide information in the column "Date(s) of Breach." When we cannot find breach start dates from publicly available articles and documents, the date found in the column "Date(s) of Breach" is regarded as the breach start date.
  (2) *Example*: Nvidia Corporation's network was accessed by hackers in 2014. In its breach notification letter provided to California Attorney General, Nvidia did not mention on which date hackers breached its cybersecurity. We did not find the date from media articles either. However, the company entered October 8, 2014 into the column "Date(s) of Breach" on the California Attorney General website, so we use October 8, 2014 as Nvidia's breach start date.

2. Firm Detection Date

- **Type 1: The date when victim firms were notified of the breaches by external third parties**
  (1) *Description*: On such dates, third parties notified victim firms about their data breaches. Prior to the third-party notification, victim firms were not aware of these data breaches.
  (2) *Example*: Target Corporation was attacked by hackers in 2013. Although Target received an alert from its internal system on December 2, 2013, no serious attention was paid to this alert. After the U.S. Department of Justice notified the company about its data breach on December 12, 2013, Target started an investigation and terminated hackers' intrusion on December 15, 2013. We use December 12, 2013 as the date of firm detection.
- **Type 2: The date when victim firms internally discovered malware, unauthorized access or suspicious activities**
  (1) *Description*: On such dates, the internal security teams of the victim firms detected malware or hints that indicated hackers' unauthorized access. Then they realized that they had been attacked by criminals.
  (2) *Example*: Equifax Inc. experienced a large-scale data breach in 2017. In its press release, Equifax claimed that on July 29, 2017, its security team noticed that hackers had exploited the application vulnerability of its website to gain unauthorized access to its network. Then they acted immediately to stop the intrusion, and began working with Mandiant, a cybersecurity company, to investigate the incident and identify the specific information that had been stolen,

as well as the affected customers. Also, they promptly reported the criminal access to law enforcement and continued to work with authorities. July 29, 2017 is the firm detection date of this incident.

3. Breach Disclosure Date

- **Type 1: The date on which victim firms announced their data breaches**
  (1) *Description*: On such dates, victim firms confirmed that they had experienced data breaches. Under certain circumstances, victim firms disclosed that it started investigating a likely data breach.
  (2) *Example*: On December 19, 2013, Target Corporation announced that it was aware of the compromise of customer payment information and was investigating the data breach through cooperating with financial institutions and law enforcement agencies. December 19, 2013 is the breach disclosure date of this incident.
- **Type 2: The date on which news articles mentioned victim firms' confirmation about their data breaches**
  (1) *Description*: On such dates, news or social media articles reported that public firms had confirmed their data breaches. When we cannot find victim firms' own announcement date, we regard the date of these reporting articles as the breach disclosure date.
  (2) *Example*: On October 28, 2016, a tweet published Converse (Australia)'s announcement about its recent data breach in which hackers used malware to collect customer payment information, so the breach disclosure date is October 28, 2016.
- **Type 3: The date on which public documents disclosed data breaches**
  (1) *Description*: On such dates, a public document disclosed that a public firm had experienced a data breach. Although we do not find the confirmation of victim firm, the document is from a reliable information source. Therefore, the date of the public document is regarded as the breach disclosure date.
  (2) *Example*: On June 26, 2012, the Federal Trade Commission announced on its website that it had filed lawsuit against Wyndham Worldwide Corporation because Wyndham's negligence over customer privacy protection led to three consecutive data breaches. We do not find the date of Wyndham's own confirmation. Considering that FTC is a reliable information source, we use June 26, 2012 as the breach disclosure date.

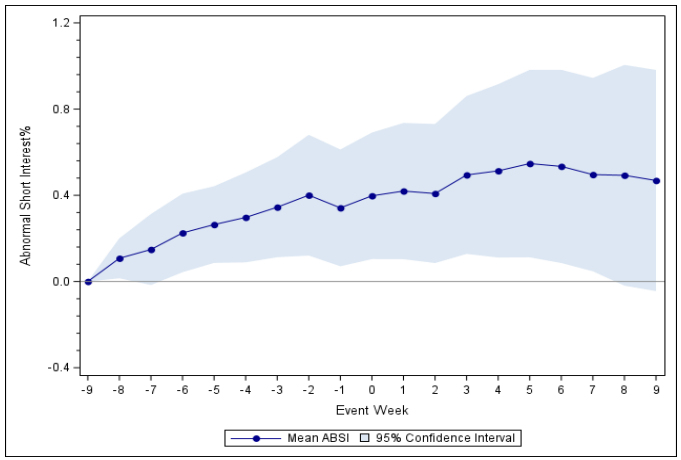
During the process of collecting dates, some incidents provide only a period rather than a concrete date. In that situation, we choose to use approximate dates when the length of the period does not exceed a month. Our approximating rules are described below.

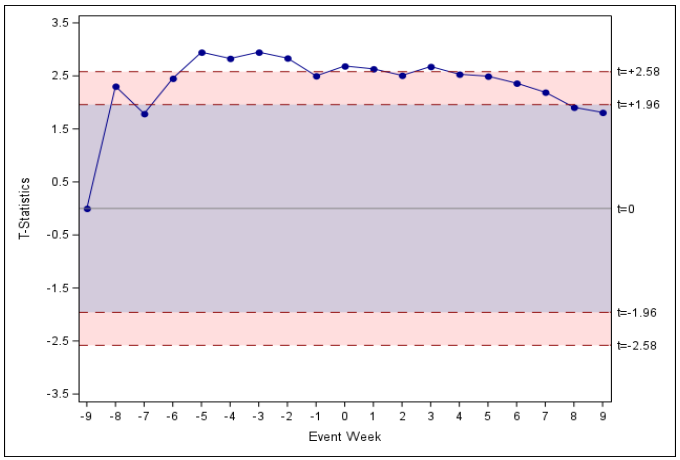
- **Rule 1**: If the period is a particular month or the beginning of a particular month, the approximate date is the first date of that month. For example, (early) April 2003 suggests that the approximate date be April 1, 2003.
- **Rule 2**: If the period is the midpoint of a particular month, the approximate date is the 11th date of that month. For example, mid-April 2003 suggests that the approximate date be April 11, 2003.
- **Rule 3**: If the period approaches the end of a particular month, the approximate date is the 15th date of that month. For example, late April 2003 suggests that the approximate date be April 15, 2003.
- **Rule 4**: If the period is a particular week or weekend, the approximate date should be the first date of the week or weekend.

IA4. Robustness Tests – Alternative Strategies to Identifying Abnormal Holdings and Trading

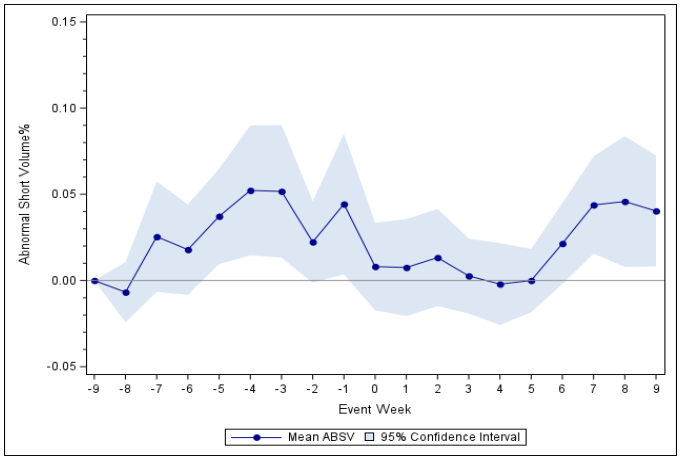
Internet Appendix IA4 replicates our main analyses of 109 massive data breaches around *Breach Start Date* using alternative measures for abnormal holdings and trading. In this table, we replace the Karpoff measures with detrended-only abnormal measures that control solely for firm-specific systematic time trend. Specifically, we detrend holdings or trading activity on Day t by subtracting a moving average calculated over Day t-90 to t-1. Then, we take the weekly average of the abnormal measures over each event week. Weekly average of event week -9 is normalized to zero, allowing the figure to display changes in subsequent weeks relative to event week -9. Week 0 covers the event day window [0, 4] in which day 0 is *Breach Start Date* or the first trading day after *Breach Start Date*.

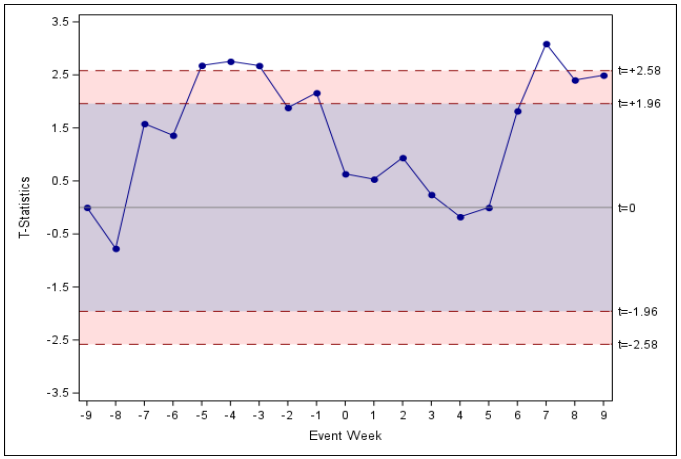
Panel A Abnormal Short Interest



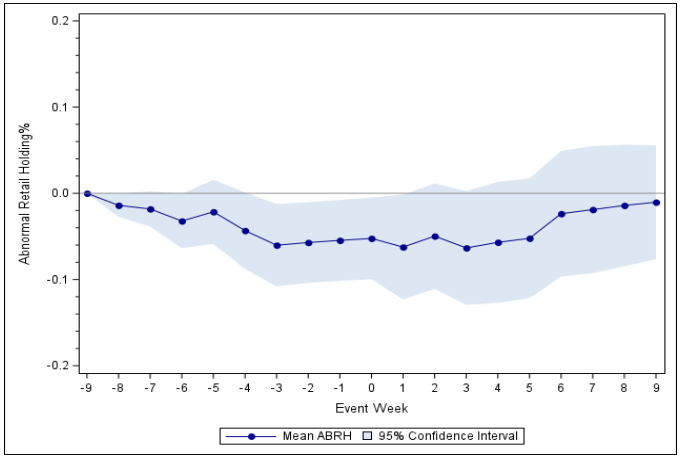


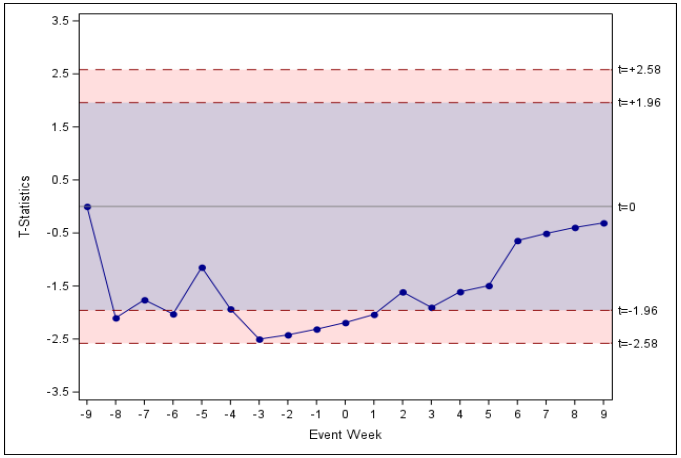
Panel B Abnormal Short Volume



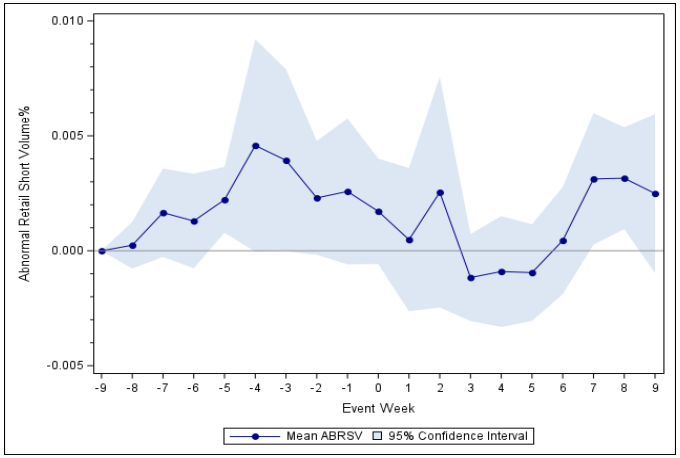


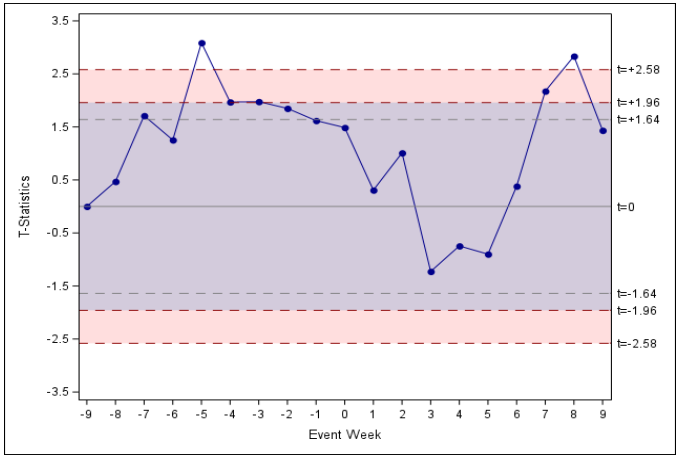
Panel C Abnormal Retail Holdings



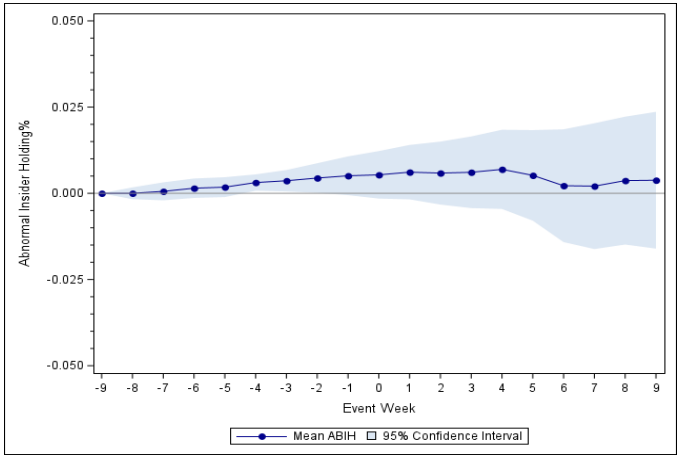


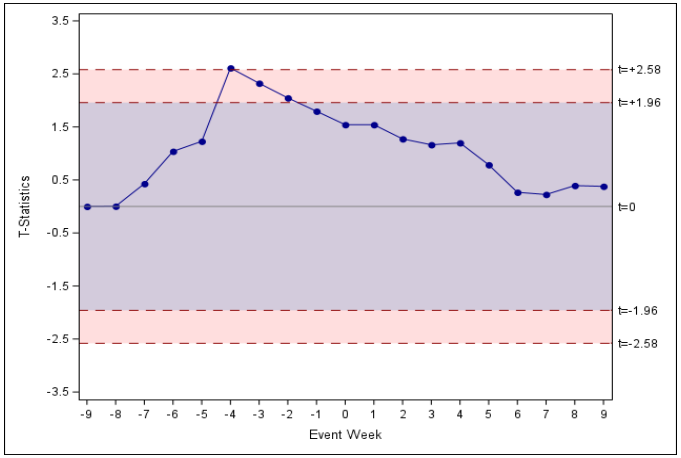
Panel D Abnormal Retail Short Volume



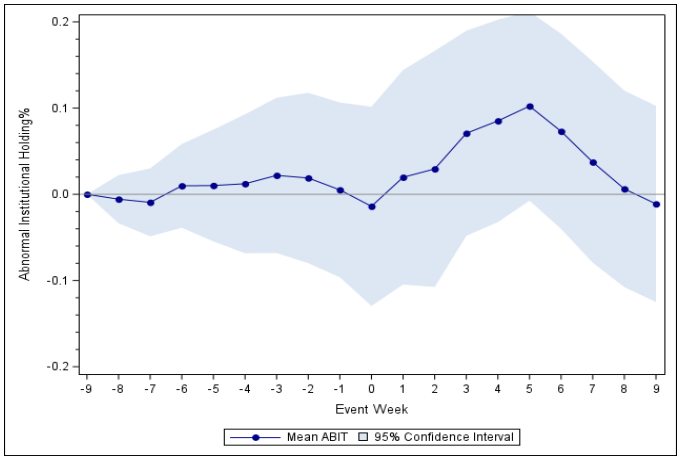


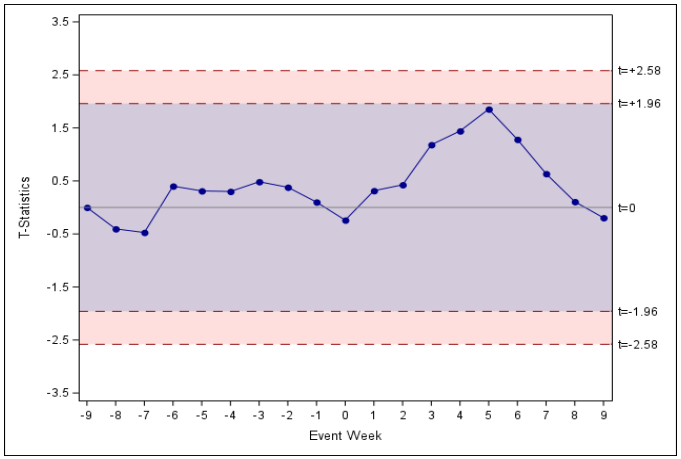
Panel E Abnormal Insider Holdings





Panel F Abnormal Institutional Holdings

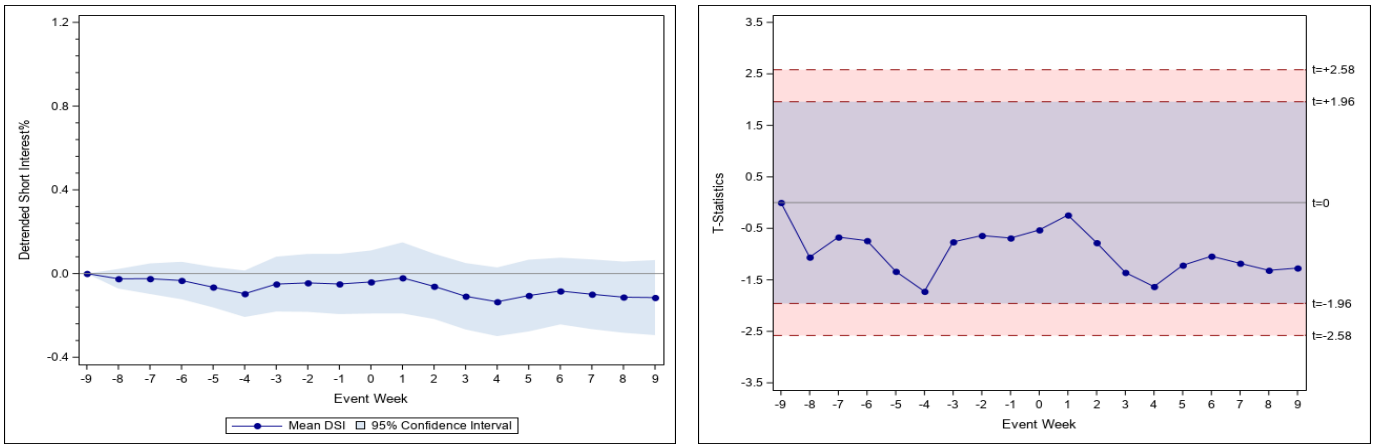




IA5. Robustness Tests – Counterfactual Peer Analyses

Internet Appendix IA5 conducts the counterfactual analyses of changes in abnormal short interest and abnormal Google Search around *Breach Start Date* using 515 peer firms for the 109 massive data breaches. The peers are selected via propensity score matching based on firms' cyber risk score, *Size*, *Book-to-Market*, *Momentum*, industry and year fixed effects. For each attacked firm, five nearest neighbor firms are identified as peers. We then construct the abnormal measure for each peer by removing firm-specific systematic time trends using the detrending procedure that subtracts the average over Day t-90 to Day t-1 from the daily raw value. Finally, we compute weekly averages of the abnormal measures for counterfactual peers in accordance with the event weeks/days of the paired attacked firms. The variable and figure definitions are consistent with our main analyses. We obtain the data of cyber risk scores from Florackis et al. (2023).

Panel A Abnormal Short Interest around Breach Start Date



Panel B Abnormal Google Search around Breach Start Date