Essays on Cybersecurity Risks

ESSAYS ON CYBERSECURITY RISKS

BY

WANGGEFEI WU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS
2025

ACCEPTANCE

This dissertation was prepared under the direction of the *Wanggefei Wu* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Dr. Stephen H. Shore (Chair)
Dr. Ajay Subramanian
Dr. Alejandro Del Valle
Dr. David Maimon (External – Andrew Young School of Policy Studies)

ABSTRACT

ESSAYS ON CYBERSECURITY RISKS

BY

WANGGEFEI WU

July 18th, 2025

Committee Chair: Dr. Stephen H. Shore

Major Academic Unit: Maurice R. Greenberg School of Risk Science

With the rapid increase in major cybersecurity breaches in recent years, cybersecurity risk has become one of the top global concerns, leading regulators and firms to strengthen cybersecurity risk management and disclosure practices. Recognizing the escalating threat landscape, the U.S. Securities and Exchange Commission (SEC) and U.S. state governments have implemented stricter disclosure regulations. The increasing regulatory pressure on firms now requires them to disclose not only cybersecurity breaches but also their risk management practices and board oversight. This dissertation investigates how firms communicate cybersecurity risk information in their disclosures, how these disclosures respond to both the firm's own breach incidents and those of industry peers, whether they carry predictive signals of future breaches, and how markets react to cybersecurity breaches.

The first chapter introduces a firm-level cybersecurity awareness score constructed from annual 10-K filings to capture how firms respond to cybersecurity breaches. By using firm-specific filing timelines, the chapter examines how firms update their cybersecurity awareness following breach incidents. The finding shows that firms significantly increase the use of cybersecurity-related terms in their 10-K fillings after experiencing either their own breaches or those of peers, indicating that the cybersecurity awareness score reflects perceived cybersecurity risk. Notably, higher cybersecurity awareness scores are positively associated with the likelihood of future breaches, suggesting that firms recognize and communicate their heightened exposure to cybersecurity risks. However, while a firm's own past breaches predict future incidents, competitor breaches do not.

The second chapter evaluates the capability of Large Language Models (LLMs), including GPT-4, Llama3, and finBERT, for sentiment analysis of cybersecurity-related disclosures in 10-K filings. These models are compared to traditional dictionary approaches. The findings show that sentiment scores generated by GPT-4 and finBERT significantly predict future breaches, with a more negative tone associated with a lower likelihood of being breached. This suggests that firms using more cautious language around cybersecurity risks may be more proactive in managing those risks. However, it remains difficult to discern whether firms systematically adjust the tone of their 10-K fillings following their own breach incidents. When examining peer effects, evidence of sentiment spillover emerges, but only under certain models. Specifically, Llama3 and traditional dictionary-based models detect modest tone adjustments in response to competitor breaches, while GPT-4 and finBERT do not, indicating that spillover effects in sentiment are subtle and model-dependent.

The third chapter investigates stock market responses to cybersecurity breaches, focusing on changes in abnormal returns, volatility, and intra-industry and inter-industry spillover effects. Empirical analysis reveals significant negative market responses, with increased volatility and decreased abnormal returns. Notably,

peer firms often benefit from malicious breaches within the same industry, exhibiting higher abnormal returns and reduced volatility, suggesting competitive effects rather than contagion. These spillover effects are primarily driven by malicious breaches and are significant at the firm level than at the industry level, likely due to the presence of private firms in the dataset. Overall, the findings highlight the diverse ways in which cybersecurity breaches influence financial markets and underscore the importance of distinguishing between breach types and levels of analysis.

These essays contribute to the growing literature related to cybersecurity risks by offering novel insights into firms' behavior to cybersecurity risks, measurement, and financial market reaction.

ACKNOWLEDGEMENT

My doctoral journey would not have been possible without the guidance, support, encouragement, and love of so many people.

I am deeply grateful to my advisor, Dr. Stephen H. Shore, for his thoughtful guidance, unwavering support, and kind encouragement throughout my doctoral studies. This dissertation would not have been possible without his insights and suggestions.

I would also like to thank the members of my dissertation committee, Dr. Ajay Subramanian, Dr. Alejandro Del Valle, and Dr. David Maimon, for their generous guidance and support.

Special thanks to all the professors I have learned from, the department faculty and staff, and my fellow Ph.D. students, for making my time at Georgia State University a truly unique experience.

Many thanks to my parents for their endless, unconditional support. Thanks also to my friends for standing by me, listening, laughing, and reminding me I was never alone.

I am deeply grateful to everyone who lifted me up, step by step, making it possible for me to come this far.

TABLE OF CONTENTS

ACKNOWLEDGEMENT 5

LIST OF TABLES 8

LIST OF ILLUSTRATIONS 9

CHAPTER 1 : SPILLOVER EFFECTS OF BREACH INCIDENTS ON CYBERSECURITY AWARENESS 1

1.1 Introduction 1

1.2 Data and Key Variables 5

1.3 Regression analysis 13

1.4 Robustness tests 24

1.5 Conclusion 27

CHAPTER 2 : CAN CHATGPT REVEAL MORE? ASSESSING CYBERSECURITY RISK ANALYSIS IN CORPORATE DISCLOSURES 29

2.1 Introduction 29

2.2 Data and Variable Construction 33

2.3 Regression analysis 43

2.4 Conclusion 52

CHAPTER 3 : THE IMPACT OF DATA BREACHES ON STOCK RETURNS 54

3.1 Introduction 54

3.2 Data and Basic Characteristics 57

3.3 Regression analysis 64

3.4 Conclusion 77

APPENDIX A : APPENDIX FOR CHAPTER 1 78

APPENDIX B : APPENDIX FOR CHAPTER 2 86

APPENDIX C : APPENDIX FOR CHAPTER 3 89

LIST OF TABLES

TABLE 1.1 Summary Statistics for Cybersecurity Awareness Scores . . . 9
TABLE 1.2 Summary Statistics of Observation Periods Between Consecutive 10-K Filings (in Days) . . . 12
TABLE 1.3 Can firms predict their own subsequent breaches? . . . 15
TABLE 1.4 Can firms' disclosures predict their own subsequent breaches? . . . 17
TABLE 1.5 Do firms' disclosures respond to breaches? . . . 19
TABLE 1.6 Do firms' disclosures respond to their competitors past breaches? . . . 22
TABLE 1.7 Do competitors past breaches predict firms' future breaches? . . . 24
TABLE 1.8 Summary Statistics for Inflation Risk Awareness Scores . . . 25
TABLE 1.9 Pseudo test for Research Question 2 . . . 25
TABLE 1.10 Pseudo test for Research Question 3 . . . 26

TABLE 2.1 Summary statistics . . . 40
TABLE 2.2 Do firms' sentiment in 10-K fillings predict their own subsequent breaches? [1ex] 45
TABLE 2.3 Do past breaches affect how firms communicate about cybersecurity risks in 10-K fillings . . . 48
TABLE 2.5 Do firms' sentiment in 10-K fillings responds to their competitors' past breaches? 51

TABLE 3.1 Descriptive Statistics . . . 60
TABLE 3.2 Industry-level Summary Statistics for Volatility and Abnormal Return . . . 63
TABLE 3.3 Fixed effects logistic regression . . . 66
TABLE 3.4 Market Reactions to Data Breaches (firm-level) . . . 69
TABLE 3.5 Market Reactions to Malicious and Non-malicious Data Breaches (firm-level) . . . 71
TABLE 3.6 Market Reactions to Data Breaches (Industry-level) . . . 74
TABLE 3.7 Market Reactions to Malicious and Non-malicious Data Breaches (Industry-level) 76

TABLE A.1 Research Question 2 . . . 78
TABLE A.2 Research Question 3 . . . 79
TABLE A.3 Spillover effect:Research Question 4 . . . 79
TABLE A.4 Summary Statistics for Cybersecurity Awareness Score with a focus on cybersecurity risk management . . . 81
TABLE A.5 Research Question 2 . . . 81
TABLE A.6 Research Question 3 . . . 82
TABLE A.7 Spillover effect:Research Question 4 . . . 82
TABLE A.8 Research Question 5 . . . 83
TABLE A.9 Logistic Regression: Breach Likelihood and Interaction with Time Difference . . . 84
TABLE A.10 Effect of Cybersecurity Breaches on Firm Fundamentals . . . 85

TABLE B.1 ChatGPT Sentiment Analysis Outputs . . . 87
TABLE B.2 LLM Sentiment Analysis Results . . . 88

TABLE C.1 Market Reactions . . . 90
TABLE C.2 Industry Classification by Type and SIC Code . . . 91

LIST OF ILLUSTRATIONS

FIGURE 1.1 Time trends: this figure plots the cybersecurity awareness score in 10-Ks for each year, based on a sample of 8,727 publicly traded firms with 70,105 annual fillings. A higer score reflects a higher level of cybersecurity awareness. 9
FIGURE 1.2 This figure displays the distribution of time intervals between consecutive 10-K filings, measured in days. The histogram uses a bin width of 10 days, which provides sufficient granularity to visualize annual reporting patterns. 12
FIGURE 1.3 Set up 13
FIGURE 1.4 Research question 1 14
FIGURE 1.5 Research question 2 17
FIGURE 1.6 Research question 3 19
FIGURE 1.7 Research question 4 22
FIGURE 1.8 Research question 5 23

FIGURE 2.1 Time trends: This figure plots the average GPT sentiment score in 10-Ks for each year, based on a sample of 6,594 observations from 803 firms. A score closer to 3 reflects a more negative tone, whereas a score closer to 1 indicates a more positive tone. 42

FIGURE 3.1 Weekly breaches from January 2005 to December 2018. 59
FIGURE 3.2 Daily volatility of the value-weighted market volatility for the period 2004 through 2019. 62
FIGURE 3.3 Daily value-weighted industry volatility 63

CHAPTER 1

SPILLOVER EFFECTS OF BREACH INCIDENTS ON CYBERSECURITY AWARENESS

1.1. Introduction

Cybersecurity risks remain a top concern for businesses as they increasingly depend on digital platforms and online services. Cybersecurity threats continue to grow, with cyber-attacks affecting businesses of all sizes in many industries, including supply chain, financial, health, and retail. According to a 2024 FBI report, cybercrime losses in the U.S. hit a record $12.5 billion in 2023, driven primarily by investment fraud and email scams. More and more states have established rigorous laws to protect data privacy. The U.S. Securities and Exchange Commission (SEC) has taken several steps over the years to mandate and enhance risk disclosures, particularly regarding cybersecurity risks and incidents, for publicly traded companies. This effort began with the introduction of mandatory risk disclosures in 2005 (Item 1A), followed by detailed guidance on cybersecurity risk disclosure practices in 2011 (SEC, 2011). In 2018, the SEC issued further guidance emphasizing the importance of disclosing information about cybersecurity risk management programs, preventative measures, and specific disclosure areas that companies should address when evaluating their cybersecurity risks (SEC, 2018). Most recently, the SEC released new rules regarding data security rules to enhance and standardize disclosures related to cybersecurity risk management, strategy, governance, and incident reporting by public companies (SEC, 2023). Although these regulations are specifically designed for publicly traded companies, their influence extends far beyond. A data breach can result in financial losses, reputational harm, and the erosion of trust from customers and vendors. Moreover, because businesses are so interconnected, a breach at one firm can affect not only the customers and vendors but also cast doubt on cybersecurity risk management across its entire industry. Therefore, companies of all types should become familiar with these regulatory developments and historical breach incidents as they design and maintain effective cybersecurity risk management plans.

Both breach disclosures and financial disclosures regarding cybersecurity risks and incidents serve not only as legal obligations but also create a dynamic information environment in the market. Thus, while firms are expected to disclose information transparently, they simultaneously gain access to their competitors' cybersecurity practices and vulnerabilities. This dual role raises several important questions: What are the market-wide effects of public cybersecurity disclosures? How do firms interpret and respond to this information? Do they revise or strengthen their own cybersecurity risk management strategies as a result? This paper exploresthese questions by constructing firm-level cybersecurity awareness measures based on annual disclosures and investigating whether breach incidents, either within the firm or among competitors, have spillover effects on firms' cybersecurity awareness.

In this paper, I first construct firm-level cybersecurity awareness scores using SEC 10-K filings, based on the frequency of cybersecurity-related terms, to quantify firms' responses and proxy their level of implementation of cybersecurity practices. Then by performing the empirical analysis, the result shows that firms exhibit heightened concern about cybersecurity risks following both their own breach incidents and those of their competitors. Cybersecurity awareness scores incorporate future risks. Firms with higher scores have a higher likelihood of experiencing future breaches. Moreover, disclosures of both a firm's own breach and its competitors' breaches generate spillover effects. This spillover is of broader public concern, as regulators actively bring such information into the marketplace. By examining these dynamics, this paper finds that firms increase their cybersecurity awareness after observing breach incidents, either internally or within their industry, indicating that firms recognize such events as signals of systemic vulnerabilities that warrant proactive cybersecurity risk management.

To validate the robustness of the main findings, I also conduct a pseudo test using inflation risk disclosures, which are also subject to the SEC disclosure regulations but are theoretically unrelated to cybersecurity breaches. By constructing an inflation risk awareness score using inflation-related

terms from 10-K filings, following the same methodology used for cybersecurity awareness, I find that these alternative risk disclosures do not predict future breaches, nor do firms systematically increase inflation risk-related terms following a breach. Therefore, this pseudo test confirms that the observed associations between cybersecurity awareness and breach incidents are specific to cybersecurity, rather than driven by a general tendency to disclose risk.

This paper makes several contributions to the growing literature at the intersection of text analysis and financial disclosures. This paper shifts the focus from the consequences of breach incidents to the predictive signals and strategic responses embedded in firms' financial disclosures. While much of the existing research emphasizes the determinants and market impacts of cybersecurity events, I examine how breaches—both internal and external—inform future risk and disclosure behavior. Prior studies have found that disclosure of cybersecurity information is positively associated with firm value (Gordon et al., 2010), and that cybersecurity risk is priced in the cross-section of returns and predicts future attacks (Florackis et al., 2023). Cybersecurity risk has also been shown to reshape firms' innovation strategies (Lattanzio and Ma, 2023), with negative implications for firm innovation. While cybersecurity breaches are generally associated with negative market reactions (Gatzlaff and McCullough, 2010; Yayla and Hu, 2011; Pirounias et al., 2014; Hinz et al., 2015; Modi et al., 2015), the magnitude and direction of the market response often depend on the nature and type of the event (Campbell et al., 2003; Wang et al., 2013b; Yayla and Hu, 2011). For instance, technology firms tend to suffer greater losses from cybersecurity incidents compared to non-technology firms (Cavusoglu et al., 2004; Pirounias et al., 2014; Yayla and Hu, 2011). Pooser et al. (2018) show that early disclosures of cybersecurity risk among U.S. P&C insurers were typically smaller, more leveraged, growing faster, and less diversified, making them more vulnerable to cybersecurity breaches. Kamiya et al. (2021) indicate that cyberattacks involving the loss of personal financial information (malicious breaches) lead to significant losses in shareholder wealth. Additionally, Wang et al. (2013a) find that the level of detail in public reports of cybersecurity incidents influences market reactions, with more specific information about the breach leading to stronger negative responses. Binfarè (2019) highlights the role of firms' visibility (size) as a key

determinant of exposure to cybersecurity threats, reinforcing the idea that larger, more prominent firms may be more likely targets.

Second, this paper contributes to the growing literature on using textual analysis to extract information from financial disclosures. Prior research has shown that financial documents such as 10-K filings (Loughran and Mcdonald, 2016; Freyberger et al., 2020; Lowry et al., 2020; Konchitchki and Xie, 2023) and earnings calls (Hassan et al., 2019) provide valuable information for investors and stakeholder and can be used to construct firm-level measures of political risk, climate risk, and inflation risk. In particular, risk disclosures, the risk factor section of 10-K filings, play a critical role in helping investors understand potential threats, assess firm-specific vulnerabilities, and compare risk profiles across firms. This study builds on previous studies by developing a measure to proxy firm-level cybersecurity risk awareness based on textual analysis of 10-K filings (Gordon et al., 2010; Lawrence et al., 2018; Berkman et al., 2018; Florackis et al., 2023; Jamilov et al., 2021).

Cybersecurity breaches have spillover effects, as they not only reflect the operational weaknesses of the affected firm but also influence the strategic decisions and market perceptions of peer firms. Prior research demonstrates two types of responses to such incidents. Some studies find evidence of negative spillovers, where investors anticipate that peer firms may face similar vulnerabilities, leading to negative market reactions for those firms (Matsumura et al., 2014; Cho and Muslu, 2021; Ettredge and Richardson, 2003; Hinz et al., 2015; Kashmiri et al., 2017; Martin et al., 2017; Islam et al., 2022). In contrast, other studies report competitive spillovers, where investors expect peer firms to benefit, such as through market share gains, resulting in positive market reactions (Jeong et al., 2019; Kelton and Pennington, 2020).

Third, this study contributes to this literature by using textual analysis to examine how the disclosure of a cybersecurity breach affects both the breached firm and its industry peers. Unlike prior research that assigns cybersecurity awareness measures to a specific calendar year, this study makes

use of firm-specific 10-K filing timelines. Because firms file on varying schedules, this setup allows for a more precise examination of how cybersecurity disclosures are updated in response to breach incidents occurring between two consecutive filings. By aligning breach incidents with actual disclosure windows, the analysis captures more accurate firm responses to cybersecurity risk. As a result, this study offers new insights into how firms communicate and adapt to evolving cybersecurity risks through their 10-K fillings.

The rest of the paper is organized as follows. Section 2 describes the data and the construction of key variables. Section 3 examines the impact of breach incidents on both a firm's own cybersecurity awareness and that of its peers. Section 4 presents robustness tests. Finally, Section 5 concludes.

1.2. Data and Key Variables

1.2.1. Data on breach incidents

The primary dataset used in this analysis is constructed from the SEC's EDGAR database and the Privacy Rights Clearinghouse (PRC). All states in U.S. mandate breach notification in a timely manner under the State Security Breach Notification Laws. PRC collects breach incidents, including a brief description of each event, the date it was reported, the breach type, the type of organization, the entity's location, and the total number of records affected when available. However, the PRC data has two key limitations. First, the number of records affected is often an estimate rather than a precise count. Second, the dataset includes only those breaches that have been publicly disclosed, potentially omitting the unreported breach incidents. Despite these limitations, the PRC dataset remains the most comprehensive publicly available dataset of its kind.

Breach incidents take various forms. Some due to accidental or non-malicious actions, for example, accidental improper sharing, loss of devices. Others are caused by deliberate, malicious attempts to gain information that can be used for profit. They often target personally identifiable information,

such as customers' online accounts, credit card information, or trade secrets. Malicious breaches can lead to substantial financial losses. In this paper, I classify data breach incidents into two categories: malicious and non-malicious breaches. Malicious breaches include hacking or malware, credit card fraud, and insider breaches information. All other breach subcategories are labeled as non-malicious breaches, where records were exposed accidentally, e.g., through the loss of devices or accidental public exposure of sensitive information. Breaches of unknown type are excluded from the analysis.

This paper first focuses on breach incidents involving publicly traded firms, as these companies are required to file annual financial disclosures with the SEC. I match organization names reported in the PRC database with firm names listed in Compustat and the Center for Research in Securities Prices (CRSP). After removing duplicates, breaches reported multiple times across different states, the final sample includes 920 breach incidents involving 431 public companies between 2005 and 2022. The number of breaches per firm ranges from 0 to 11. On average, within the public sector, each firm has approximately 0.73 industry peers that have also experienced a breach, suggesting that breaches tend to cluster within the same industry.

1.2.2. Data on Cybersecurity awareness score

Starting from 2005, the SEC mandated that all firms provide a structured discussion of the risk factors they face in their annual 10-K filings. I extract the discussion on cybersecurity risk in the "Item 1A Risk Factors' section from firms' 10-K, which contains information about the most significant risk factors for each firm. Some firms do not have an "Item 1A Risk Factors" section are excluded from the sample. These firms are typically small, as defined by SEC Regulation S-K Item 10, and are not required to provide information about risk factors. The final SEC filings sample consists of 8,727 publicly traded firms and 70,105 10-K filings covering the period from 2005 to 2022. On average, each firm contributes 8 filings over the sample period. Among these, 7,764 firms have more than one filing, enabling the construction of year-over-year changes in disclosure content. A total

of 59,577 filings have a subsequent filing, which allows for the construction of firm-year observation periods and the analysis of how prior disclosures relate to future breach events. Therefore, 5.6% firms in the market experience at least one breach incident during my sample period. Then I build on recent studies and use a textual analysis of 10-K fillings to construct a proxy for firm-level cybersecurity awareness (Berkman et al., 2018).

Frist, I use a web-crawling algorithm to download all "10-K," "10-K405," and "10-KSB40" filings, excluding amended documents, from SEC Edgar. From each filing, I extract the fiscal year and the Central Index Key (CIK) to link the SEC filing data with the breach incident data.Then, I extract the cybersecurity related content from the "Item 1A Risk Factors" section based on the list of words and phrases form the National Initiative for Cybersecurity Careers and Studies (NICCS) [1] and cyber-related legislative Acts (Fischer, 2014). Next, I tokenize the text of each extracted annual report, convert all words to lowercase, and apply a stemming algorithm to obtain the root words and control for plural forms. Using RapidMiner, I process each firm's annual report by fiscal year. Using RapidMiner, I process each firm's annual report by fiscal year. Finally, for each firm and fiscal year, I compute a firm-specific cybersecurity awareness score by summing the number of identified cybersecurity realted words and phrases in the annual reports over the observation period.

Thus, I measure a firm's cybersecurity awareness by considering both the length of relevant disclosures and the specificity of the language used. The cybersecurity awareness score is higher when the language is more directly related to cybersecurity, as opposed to vague or indirect references. Accordingly, firms with higher scores place greater emphasis on cybersecurity risk in their 10-K filings compared to firms with lower scores.

In this paper, I use absolute frequency, the total number of cybersecurity-related keywords in the annual filings, to proxy a firm's cybersecurity awareness score in the following empirical design.

[1] https://niccs.cisa.gov/cybersecurity-career-resources/vocabulary

In the appendix, I also test an alternative measure based on relative frequency, which weights the total number of keywords by the length of the filing. Both versions yield consistent results. I adopt the absolute frequency measure in the main analysis, as the length of cybersecurity risk disclosure is likely to be associated with the likelihood of reported future cybersecurity incidents. Table 1.1 and Figure 1.1 presents a generally increasing level of cybersecurity awareness among public firms over time. This pattern aligns with the rising frequency and severity of cyberattacks, as well as the growing exposure of firms to cybersecurity threats.

Firms often cite potential causes of breaches, reference incidents involving competitors, or discuss the possible impact of breaches on their operations. They may mention it to signal that they are taking preventive action, or to address potential legal consequences associated with cybersecurity issues in an effort to mitigate litigation risk. In the appendix, there are some excerpts from SEC 10-K Item 1A. Risk Factors, where firms explicitly discuss cybersecurity-related concerns in their annual disclosures.

Table 1.1: Summary Statistics for Cybersecurity Awareness Scores

Score Version	Mean	Std	Q1	Median	Q3	Min.	Max.
Score (Sum)	480.1	347.9	242.0	403.0	631.0	1.0	4,499.0
Score (Ratio)	0.0572	0.0116	0.0497	0.0564	0.0637	0.0029	0.1403

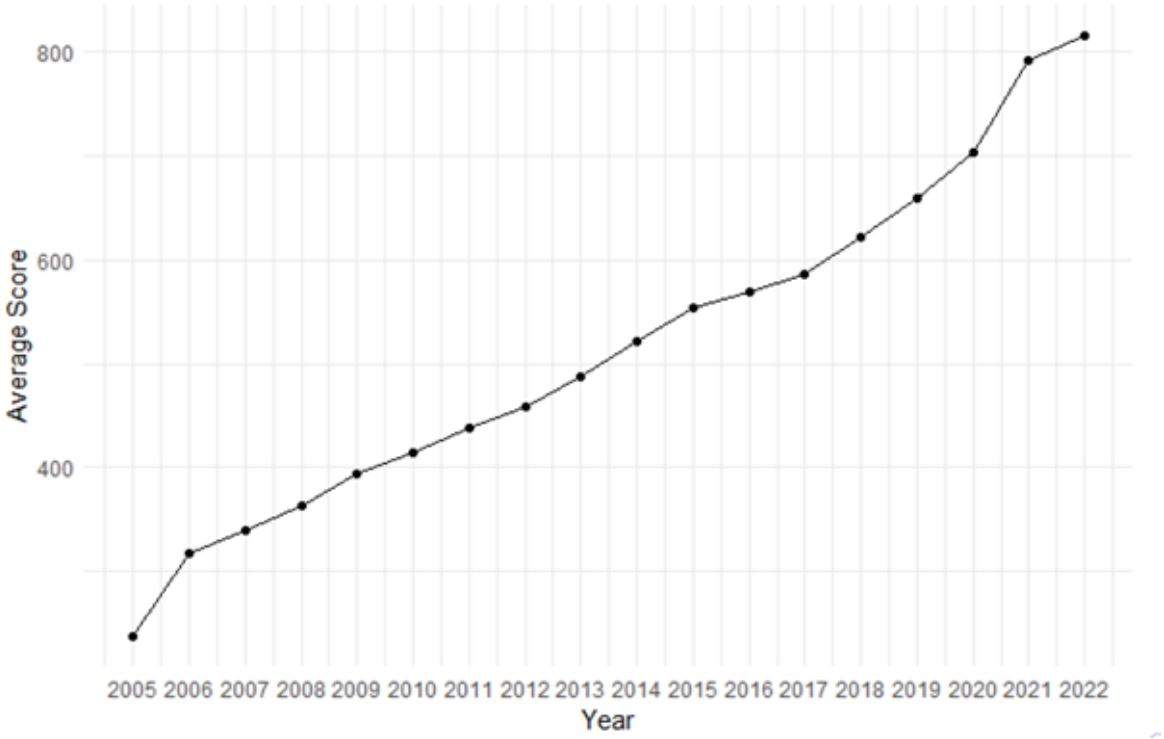

Figure 1.1: Time trends: this figure plots the cybersecurity awareness score in 10-Ks for each year, based on a sample of 8,727 publicly traded firms with 70,105 annual fillings. A higer score reflects a higher level of cybersecurity awareness.

Observation period

Typically, firms release their financial disclosures at the end of their fiscal year. However, fiscal year-ends can vary significantly across firms. While many companies align their fiscal year with the calendar year, ending on December 31, others adopt a non-calendar fiscal year. Moreover, the timing of subsequent filings is not fixed and can vary slightly from year to year. As a result, the date of the next filing is not necessarily $t+1$, but rather $t+\Delta_i t$, where Δ_i represents the time gap between filings and can differ substantially across firms.

Table 1.2 and Figure 1.2 summarize the distribution of time intervals between consecutive 10-K filings. The minimum interval is just 4 days, which likely reflects an amended or corrected filing rather than a standard annual report. On the other extreme, some intervals exceed 700 days, typically occurring when a firm files near the beginning of one fiscal year and then again near the end of the next, resulting in an unusually long gap. The histogram in Figure 2 confirms that the majority of firms file their 10-Ks approximately once per year, with most observations clustered tightly between 360 and 370 days.

To ensure consistency, I truncate the distribution of time intervals between consecutive 10-K filings by removing outliers below the 0.5th percentile and above the 99.5th percentile, focusing on the most representative range of observations. The period between two consecutive 10-K filings is treated as the observation period. This paper then focuses on identifying data breaches that occur within each observation period for each firm and calculates the corresponding cybersecurity awareness score based on the firm's annual filing at the beginning of that observation period. There are 11,232 filings where a competitor in the same industry (measured by four-digit SIC codes) among publicly traded firms experienced a breach incident during the observation period. Out of 7,764 firms, 3,909 had at least one industry peer experience a breach.

Additionally, in the appendix, I examine whether the time interval between consecutive 10-K filings affects the likelihood of a future cybersecurity breach. The results show no significant effect, suggesting that filing interval length does not confound the cybersecurity awareness score's predictive power. Therefore, I construct the cybersecurity awareness score without adjusting for filing intervals.

Table 1.2: Summary Statistics of Observation Periods Between Consecutive 10-K Filings (in Days)

Min.	**Q1**	**Median**	**Mean**	**Q3**	**Max.**	**0.5%**	**99.5%**
4.0	362.0	364.0	363.4	367.0	723.0	215.00	446.12

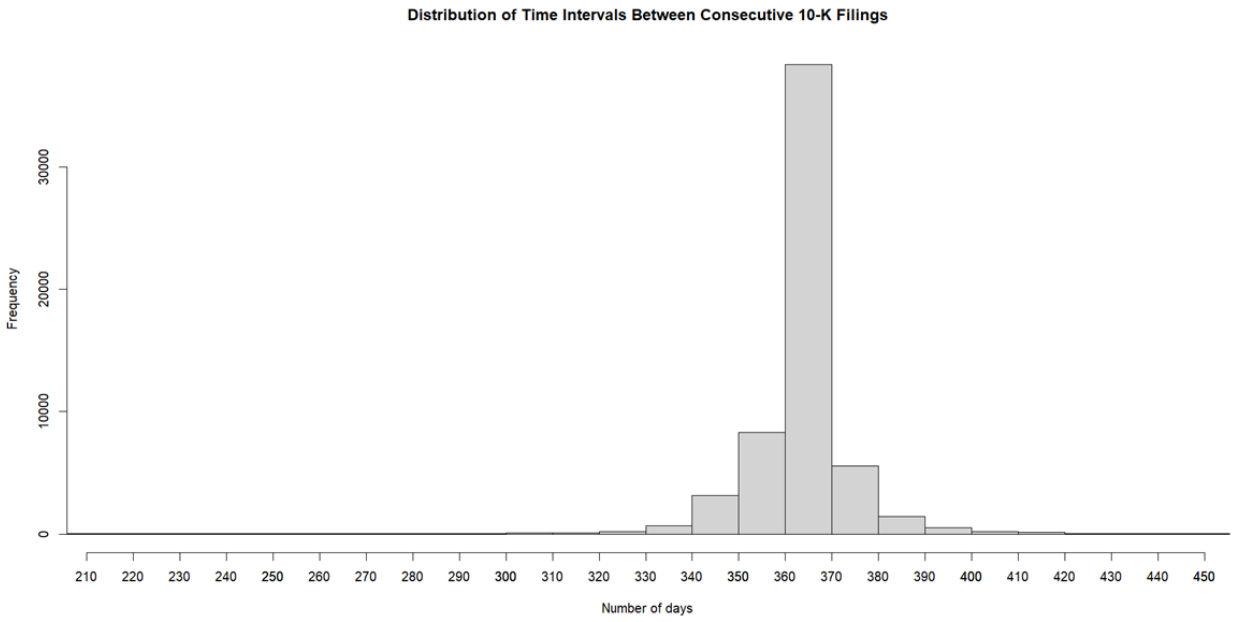

Figure 1.2: This figure displays the distribution of time intervals between consecutive 10-K filings, measured in days. The histogram uses a bin width of 10 days, which provides sufficient granularity to visualize annual reporting patterns.

1.3. Regression analysis

This paper investigates two key research questions: (1) To what extent do firms adjust their cybersecurity risk management practices in response to their own or a competitor's breach incident? and (2) Whether the cybersecurity awareness score predicts subsequent breach occurrences.

Figure 1.3 presents the timeline and empirical setup that guides the regression analysis. It depicts the sequence of 10-K filings and breach events for a focal firm i and its industry peers j, highlighting how changes in cybersecurity awareness scores are measured over time. The observation period spans from the initial filing date t to the subsequent filing date $t + \Delta_i t$, during which a breach incident occurs. Since firms release their financial disclosures on varying schedules, this setup allows us to examine how each firm updates its cybersecurity disclosures in response to breach events that occur between two consecutive filings.

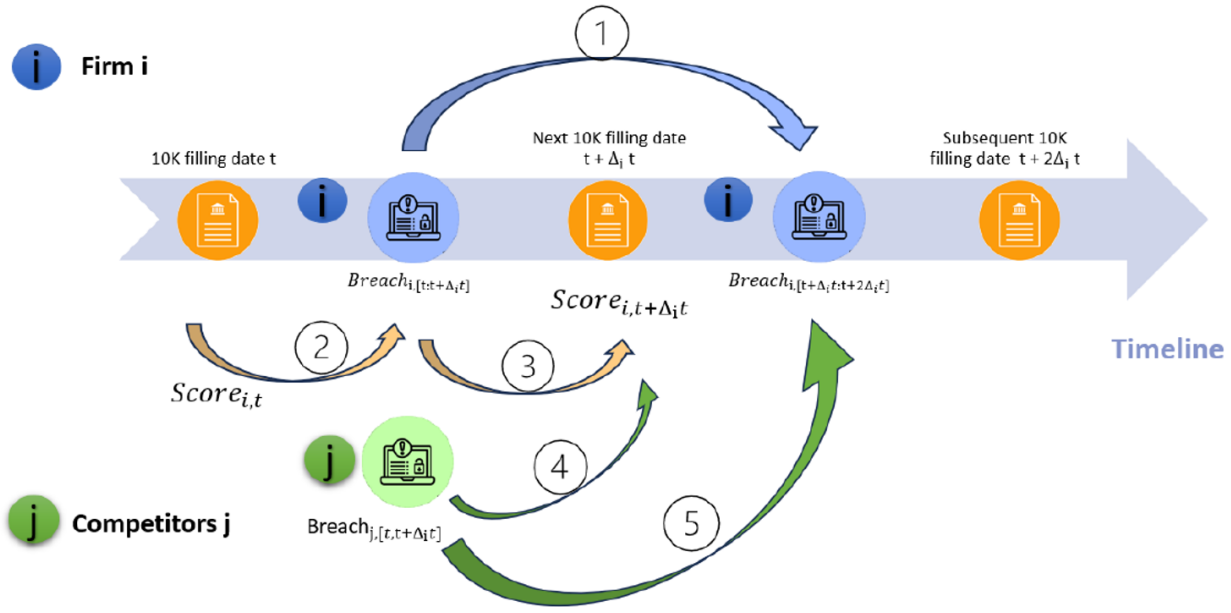

Figure 1.3: Set up

1.3.1. Own subsequent breaches

Can firms predict their own subsequent breaches?

The first part of the analysis focuses on consecutive breach incidents, as illustrated in Figure 1.4. Table 1.3 presents the results of a logistic regression estimating the likelihood that a firm will experience another data breach after already having one. The results suggest strong persistence in breach activity, a prior breach significantly increases the likelihood of a future breach. A breach incident serves as an early warning signal, suggesting the firm may have ongoing vulnerabilities or insufficient remediation efforts. This finding naturally leads to the next question: do firms anticipate and respond to this growing risk of cybersecurity breaches in their 10-K filings?

$$\log\left(\frac{P(Breach_{i,[t:t+\Delta_i t]}=1)}{1-P(Breach_{i,[t:t+\Delta_i t]}=1)}\right)=Breach_{i,t}$$

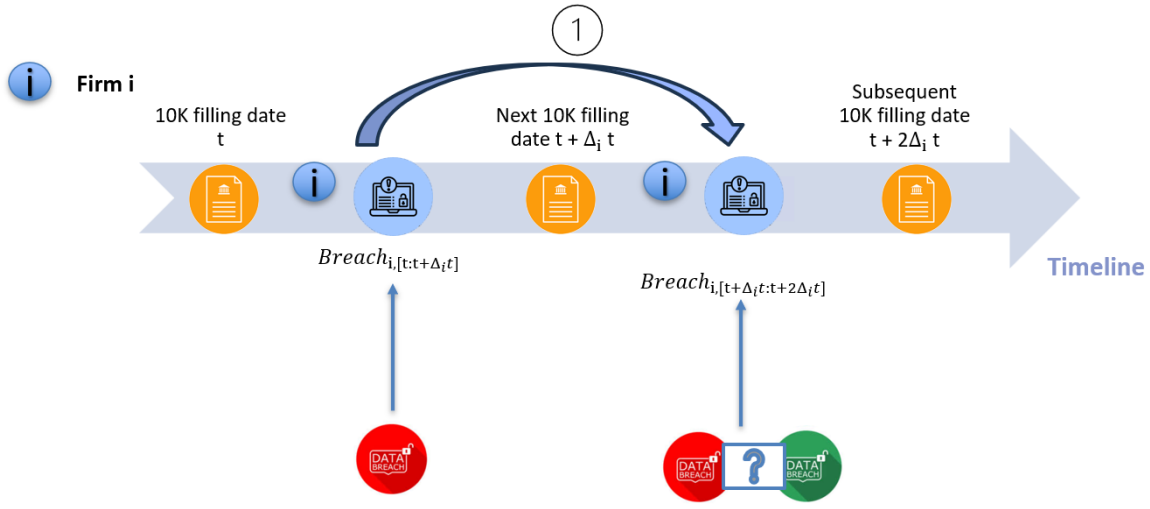

Figure 1.4: Research question 1

Table 1.3: Can firms predict their own subsequent breaches?

Dependent variable:	*Breach_lead*
Breach	3.5869***
	(0.1754)
N	59,576
Pseudo-R-squared	0.059

This table presents regression results examining whether firms can predict their own subsequent breaches. The dependent variable, *Breach_lead*, is a binary indicator equal to 1 if a breach occurs after the end of the filing's observation period. The independent variable, *Breach*, is a binary indicator equal to 1 if a breach occurs within the observation period. Both variables are dummy indicators for whether the focal firm was breached. Robust standard errors are reported in parentheses. ***, **, and * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Can firms' disclosures predict their own subsequent breaches?

The next test focuses on whether this cybersecurity awareness measure can predict future breaches. I estimate a logit regression where the dependent variable equals one if a firm experiences a breach during the observation period, and zero otherwise. The key explanatory variable is the firm's cybersecurity awareness score, measured at the beginning of the observation period.

$$Logit(P(\text{Breach}_{i,[t:t+\Delta_i t]}) = 1) = Score_{i,t}$$

As reported in Table 1.4, firms with higher cybersecurity awareness scores are significantly more likely to experience future breaches. This finding suggests that firms with greater exposure to cybersecurity threats tend to communicate these concerns to shareholders, potentially as a way to reduce litigation risk. When firms believe that the likelihood and potential consequences of a breach incident are high, they are more likely to disclose cybersecurity risks in their disclosures. This behavior can be understood through the lens of corporate risk management theory (Smith and Stulz, 1985; Froot et al., 1993). A firm's value is a concave function of future profits, meaning that increased volatility in profits, holding the mean constant, reduces firm value. Under this framework, risk management becomes valuable when it helps reduce that volatility, particularly when risk-mitigation efforts are costly and firms must balance these costs against the potential impact of adverse outcomes. In this study, the source of concavity arises from a type of operational

risk, cybersecurity risk, which can significantly disrupt business processes (Kashmiri et al., 2017). Firms facing higher expected costs from potential cybersecurity risks have stronger incentives to invest in mitigation strategies and to communicate these risks transparently in their public disclosures.

Furthermore, I include a breach indicator variable that captures whether a firm experienced a breach in the previous observation period, to control for the fact that past breaches may be strong predictors of future attacks, as demonstrated in the previous analysis. I use a one-period lead of the cybersecurity awareness score, measured at the beginning of the next observation window, to predict breach incidents occurring in the next observation period. The results in Panel B show that both the awareness score and the prior breach indicator are positively and significantly associated with the probability of a subsequent breach. This predictability supports the interpretation that the cybersecurity awareness score captures firms' exposure to cybersecurity risk.

In sum, higher cybersecurity awareness scores are associated with an increased likelihood of future breaches, confirming that the measure effectively captures forward-looking exposure to cybersecurity risk. A natural follow-up question is whether firms also respond to this elevated cybersecurity risk, especially right after experiencing a breach themselves.

$$Logit(P(\text{Breach}_{i,[t+\Delta_i t:t+2\Delta_i t]}) = 1) = Score_{i,t+\Delta_i t} + \text{Breach}_{i,[t:t+\Delta_i t]}$$

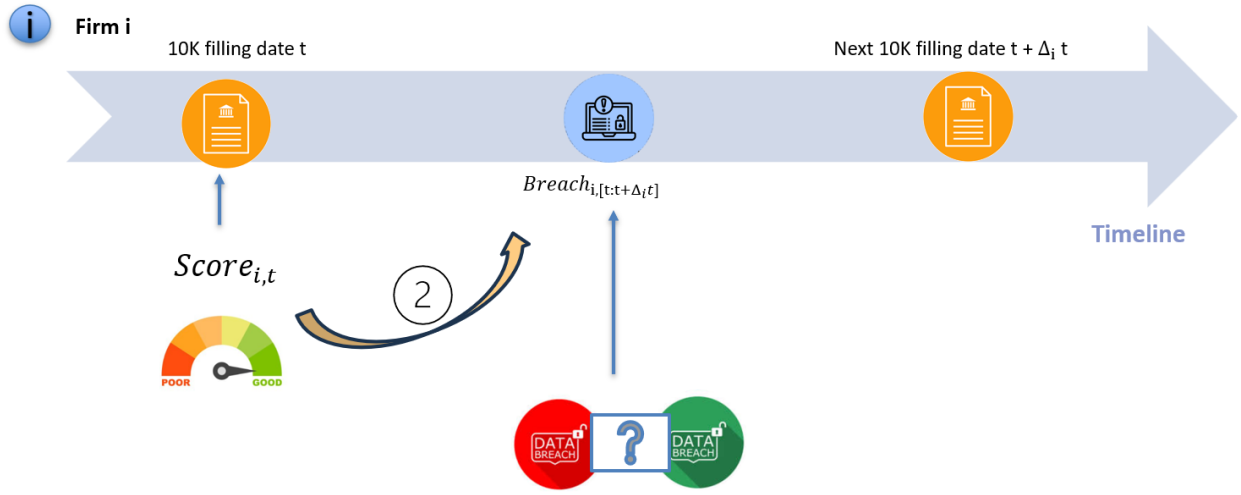

Figure 1.5: Research question 2

Table 1.4: Can firms' disclosures predict their own subsequent breaches?

Panel A: Dependent variable: Breach	(1)	(2)	(3)	(4)
Score	0.0008***	0.0008***	0.0009***	0.0008***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Quarter FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	59,576	59,576	59,576	59,576
Pseudo-R-squared	0.008	0.032	0.208	0.235
Panel B: Dependent variable: Breach_lead	(1)			
Score_lead	0.0008***			
	(0.0001)			
Breach	3.4889***			
	(0.1777)			
Observations	59,576			
Pseudo-R-squared	0.059			

This table presents regression results examining whether firms that disclose more about cybersecurity risks in their 10-K filings are more or less likely to experience subsequent breaches. The dependent variable *Breach*, is a binary indicator equal to 1 if a breach occurs with the observation period. The independent variable, *Score*, refers to the firm's cybersecurity awareness score at the beginning of the observation period. *Score_lead*, is the firm's cybersecurity awareness score at the end of the observation period. *Breach_lead*, is a breach indicator equal to 1 if the firm experiences a data breach during the next observation period, between the 10-K filing date $t + \Delta_i t$ and $t + 2\Delta_i t$. Quarter FE and Industry FE refer to the inclusion of quarter and industry fixed effects, respectively, to control for time-specific and industry-specific heterogeneity. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Do firms' disclosures respond to breaches?

Building on the prior evidence that firms' disclosures convey cybersecurity risk, I then examine whether these disclosures also respond to observed breach incidents. Table 1.5 presents regression results showing that firms significantly increase their cybersecurity awareness disclosures after experiencing a data breach. Specifically, breached firms mention approximately 23 more cybersecurity-related terms in their subsequent 10-K filings item 1A Risk factors compared to non-breached firms, holding their prior cybersecurity awareness measure constant. This suggests that breached firms may take visible actions—such as strengthening cybersecurity infrastructure, enhancing compliance with data protection regulations, or improving risk communication—to restore stakeholder trust and reduce the likelihood of future incidents. This result aligns with prior findings by Akey et al. (2023), who indicate that breached firms tend to invest more in cybersecurity following a breach to rebuild their reputation and prevent recurrence. Similarly, Pooser et al. (2018) show that some firms explicitly disclose the purchase of cyber insurance in their filings as a risk mitigation strategy. This is a strategic response aimed at managing risk and reassuring investors of their commitment to future breach prevention.

Overall, the results suggest that breach incidents drive firms to strategically adjust their 10-k fillings as part of a broader effort to manage reputation and mitigate risk.

$$Score_{i,t+\Delta t} = Breach_{i,[t:t+\Delta_i t]} + Score_{i,t}$$

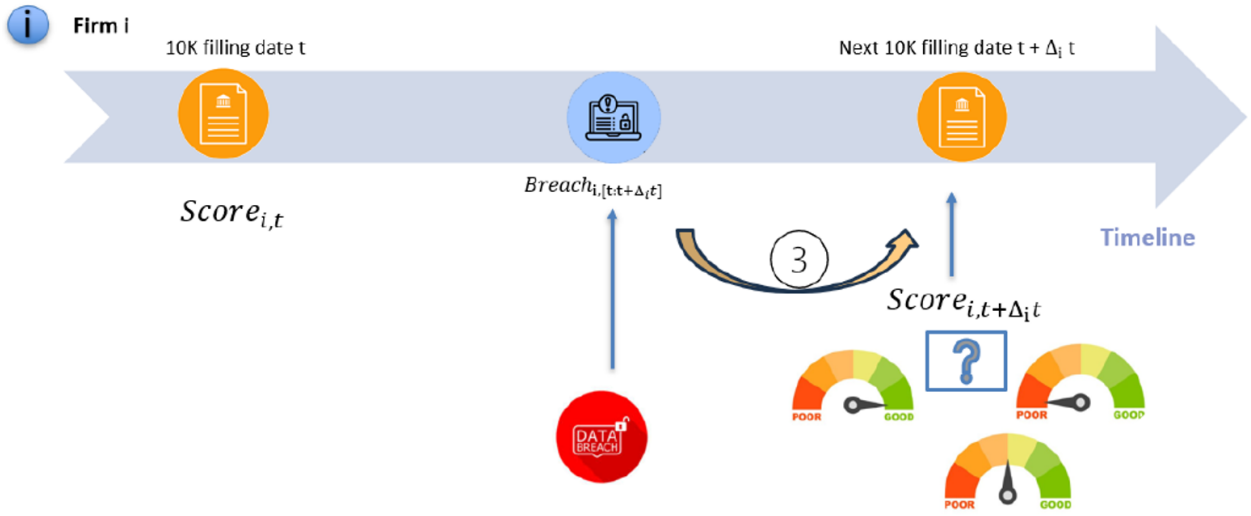

Figure 1.6: Research question 3

Table 1.5: Do firms' disclosures respond to breaches?

Dependent var: Score_lead	(1)	(2)	(3)	(4)
Breach	23.2075**	21.9752**	25.0363***	22.9593**
	(7.4784)	(7.4232)	(7.5149)	(7.4555)
Score	0.9322***	0.9273***	0.8975***	0.8891***
	(0.0017)	(0.0019)	(0.0021)	(0.0023)
Quarter FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	59576	59576	59576	59576
Adj R-squared	0.8341	0.8368	0.8441	0.8469

This table presents regression results examining whether firms increase their attention to cybersecurity risk management after experiencing a breach in the previous year. The dependent variable *Score_lead*, is the firm's cybersecurity awareness score at the end of the observation period. The independent variable, *Breach*, a binary indicator equal to 1 if a breach occurs with the observation period. Quarter FE and Industry FE refer to the inclusion of quarter and industry fixed effects, respectively, to control for time-specific and industry-specific heterogeneity. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

1.3.2. Spillover effect: Competitors breaches

Do firms' disclosures respond to their competitors past breaches?

To capture risk exposure with higer precision, this paper uses four-digit SIC codes for industry classification, which is an improvement over prior studies that typically rely on two- or three-digit codes. This level of granularity enables a more accurate assessment of how breaches at one firm may impact others within the same industry.

Analyzing the effects of competitors' breach incidents is important for several reasons. First, such incidents provide insight into the potential costs and vulnerabilities associated with cybersecurity risks. This information can help firms strengthen their own cybersecurity strategies, improve incident response plans, and ultimately reduce the likelihood and impact of future breaches and their negative consequences. In addition, it provides valuable insights into the effectiveness of cybersecurity risk management and can help companies learn from the experiences of others to prevent similar breaches.

When a breach conveys information to a focal firm and its stakeholders, it may also transmit signals to the broader market, particularly to firms in the same industry. Such signals may indicate that cybersecurity risks are not just firm-specific but systemic across an industry. As discussed in earlier sections, prior literature shows mixed evidence on the nature of these spillover effects. Some studies suggest that a breach at one firm may be interpreted as good news for its competitors. For instance, if a breach signals poor management or weak governance at the focal firm, peer firms might benefit by comparison, especially in highly competitive industries or when the attacked firm suffers repeated breaches. On the other hand, breaches may also reveal industry-wide vulnerabilities. For example, if multiple firms rely on the same cybersecurity providers, or if the breach demonstrates higher-than-expected costs and risks, then investors may reassess the entire sector's exposure to cybersecurity risk.

The regression results in Table 1.6 show that both a firm's own breaches and those of its competitors increase cybersecurity awareness in 10-K disclosures. This suggests that breach incidents trigger industry-wide contagion, raising broader concerns about cybersecurity risks, not only firm-specific but also industry-wide. Firms affected by a competitor's breach increase their use of cybersecurity-related terms by approximately 7 terms in the subsequent 10-K filing. Although this peer effect is statistically significant in baseline models, it becomes insignificant once industry fixed effects are included, suggesting that spillovers may be concentrated within particular industries or vary by sector-specific risk. In contrast, firms that experience their own breach exhibit a much stronger response, adding roughly 23 more cybersecurity-related terms in Item 1A Risk Factors. This finding also align with prior research on contagion effects in financial markets. When investors view a negative event as symptomatic of a systemic issue, they may revise expectations for peer firms accordingly, resulting in spillover effects. For instance, contagion has been observed in response to management forecasts (Pyo and Lustgarten, 1990), accounting restatements (Gleason et al., 2008), earnings announcements (Freeman and Tse, 1992; Ramnath, 2002), dividend changes (Firth, 1996), and financial misconduct (Kang, 2008). Similarly, in the context of cybersecurity, breach events convey adverse information about future risks that affect the entire sector.

$$Score_{i,t+\Delta_i t} = Breach_{j,[t,t+\Delta_j t]} + Breach_{i,[t,t+\Delta_i t]} + Score_{i,t}$$

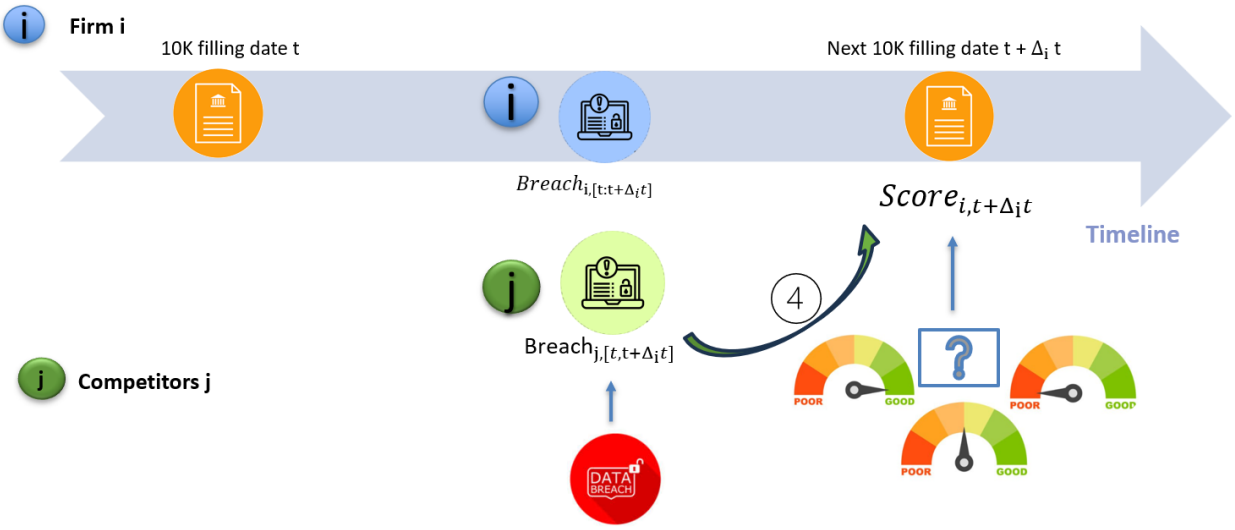

Figure 1.7: Research question 4

Table 1.6: Do firms' disclosures respond to their competitors past breaches?

Dependent var: Score_lead	(1)	(2)	(3)	(4)
Breach_competitor	7.9987***	6.9075***	-0.2910	-1.4780
	(1.4325)	(1,4667)	(1.7312)	(1.79795)
Breach_own	23.9976**	22.7661**	25.1553***	22.9661**
	(7.4546)	(7.3924)	(7.5154)	(7.4558)
Score	0.9375***	0.9328***	0.8980***	0.8896***
	(0.0018)	(0.0019)	(0.0021)	(0.0023)
Quarter FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes
Observations	59576	59576	59576	59576
Adj R-squared	0.8414	0.8443	0.8444	0.8472

This table presents regression results examining whether firms pay more attention to cybersecurity risk management after their competitors experience breaches in the previous year. The dependent variable *Score_lead*, is the firm's cybersecurity awareness score at the end of the observation period. The key independent variable, *Breach_competitor*, is a binary indicator equal to 1 if any other firm in the same industry, defined based on the same Standard Industrial Classification (SIC) code as the focal firm, experiences a breach within the observation period. The variable *Breach_own* is another binary indicator equal to 1 if the focal firm itself experiences a breach within the same observation period. *Score*, refers to the firm's cybersecurity awareness score at the beginning of the observation period. Quarter FE and Industry FE refer to the inclusion of quarter and industry fixed effects, respectively, to control for time-specific and industry-specific heterogeneity. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Do competitors past breaches predict firms' future breaches?

The final test evaluates whether competitors' breach incidents have predictive power on a firm's own future breaches. Table 1.7 shows that while a firm's own past breach is a strong and statistically significant predictor of future breach incidents, the effect of a competitor's breach is not statistically significant. This suggests that although firms appear to respond to competitor breaches in their disclosures (as shown in prior sections), for example, by increasing their cybersecurity awareness, enhancing their risk management strategies, and insulating themselves from similar vulnerabilities. However, such peer incidents do not systematically predict future breaches for the focal firm. Therefore, firms may be heeding the warnings from industry peers. The quantified effect of competitor breaches, as captured by changes in the cybersecurity awareness score, is relatively small though statistically significant. Thus, it is not surprising that competitor breaches lack the direct predictive power of a focal firm's own future breaches.

$$Logit(P(Breach_{i,[t+\Delta_i t:t+2\Delta_i t]}) = 1) = Score_{i,t} + Breach_{i,[t,t+\Delta_i t]} + Breach_{j,[t,t+\Delta_j t]}$$

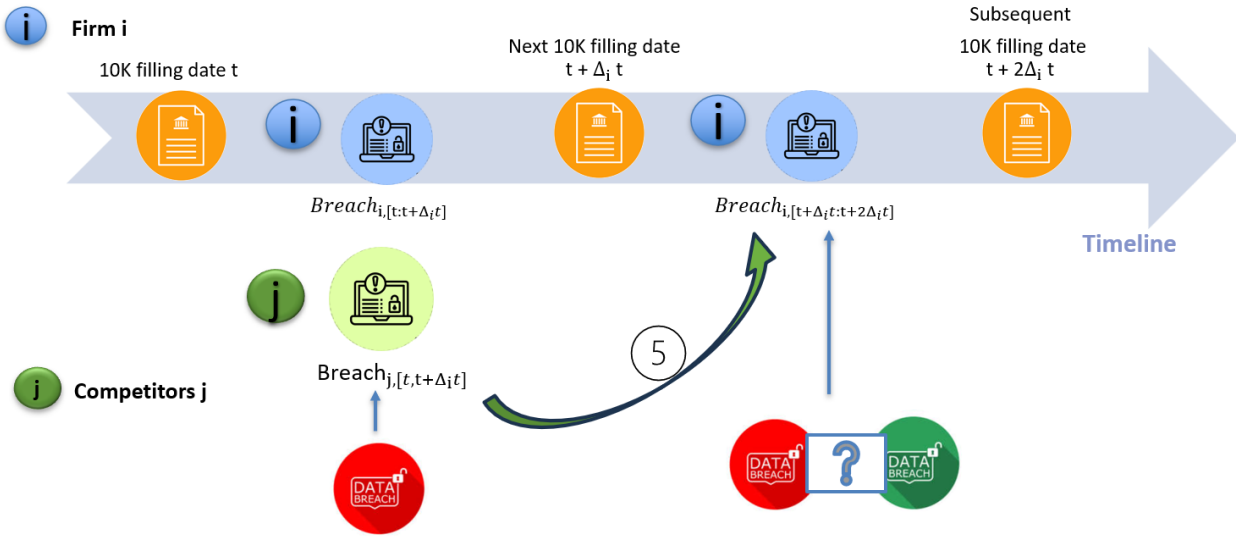

Figure 1.8: Research question 5

Table 1.7: Do competitors past breaches predict firms' future breaches?

Dependent var:	Breach_own_lead
Score	0.0007***
	(0.0002)
Breach_competitor	0.0782
	(0.1415)
Breach_own	3.4639***
	(0.1788)
Observations	59576
Pseudo-R-squared	0.0653

This table presents regression results examining whether competitors' past breaches predict a firm's future breach incidents. The dependent variable, *Breach_own_lead*, is a binary indicator equal to 1 if the focal firm experiences a breach after the end of the observation period. The independent variable, *Score*, measures the firm's cybersecurity awareness at the beginning of the observation period. *Breach_competitor* is a binary indicator equal to 1 if any other firm in the same industry (defined by SIC code) experienced a breach within the observation period. *Breach_own* is a binary indicator equal to 1 if the focal firm itself was breached within the observation period. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

1.4. Robustness tests

1.4.1. Pseudo test: Inflation Risks

To validate the robustness of the main findings, I conduct a pseudo-test using climate risk disclosures since inflation risk should not be systematically associated with cybersecurity breaches. Inflation risk is chosen as a comparison because it is subject to the same underlying SEC regulation. Beginning in 2005, the SEC mandated risk factor disclosures in firms' 10-K filings, requiring the discussion of material risks, including market-wide risks.

Following the same methodology used to construct the cybersecurity awareness score, I build an inflation risk awareness score by considering the presence of 19 inflation-related terms from Konchitchki and Xie (2023), including "inflation," "deflation," "inflationary," "hyperinflation," and "hyperinflationary," "product price," "CPI," "consumer price index," "PPI," "producer price index," "output price," "sale price," "service price," "input price," "commodity price," "raw materials price," "purchase price," "supplier price," and "manufacturer price."

I again use absolute frequency, the total count of inflation-related keywords in each annual 10-K filing, as a proxy for a firm's inflation risk awareness score. Table 1.8 reports the summary statistics for the constructed inflation risk awareness score. The results indicate substantial variation across firms in the extent to which inflation-related risks are discussed in their filings. In comparison, the cybersecurity awareness score reflects more extensive and variable attention in firm disclosures. It exhibits a higher mean (480.1), median (403.0), and greater dispersion (standard deviation = 347.9), suggesting that firms generally allocate more disclosure effort to cybersecurity risk than to inflation risk.

Table 1.8: Summary Statistics for Inflation Risk Awareness Scores

Min.	**Q1**	**Median**	**Mean**	**Q3**	**Max.**
1.0	56.0	120.0	168.6	229.0	1992.0

To evaluate whether the observed relationship between cybersecurity awareness and breach likelihood is unique to cybersecurity disclosures, I repeat the regression in research question 2 and 3 by replacing the original cybersecurity awareness score with inflation risk awareness score.

Pseudo test for Research Question 2: Can firms' inflation risk awareness score predict future cybersecurity breaches?

Table 1.9: Pseudo test for Research Question 2

Dependent var:	*Breach*
Score (Inflation)	-0.0005
	(0.0004)
N	49,994

This table presents pseudo regression results examining whether firms that disclose more about inflation risk in their annual filings are more or less likely to experience subsequent breaches. The dependent variable, *Breach*, is a binary indicator equal to 1 if a breach occurs with the observation period. The independent variable, *Score (Inflation)*, refers to the firm's inflation risk awareness score at the beginning of the observation period. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table 1.9 presents the results of the pseudo test that firms with higher inflation risk awareness scores are no more likely to suffer a subsequent data breach. Thus, only cybersecurity-related disclosures are meaningfully associated with future breach events. The cybersecurity awareness score reflects firms' actual exposure to cybersecurity risk, rather than a general risk disclosure tendency.

Pseudo test for Research Question 3: Do firms increase their attention to inflation risk in response to experiencing a cybersecurity breach?

Table 1.10 shows this placebo test suggests that firms do not systematically increase their inflation risk disclosures in response to a cybersecurity breach. This reinforces the interpretation that the firm responses observed in cybersecurity language are specific to the breach event rather than a broad-based increase in all types of risk communication.

Table 1.10: Pseudo test for Research Question 3

Dependent var:	*Score_lead (Inflation)*
Breach	-2.9275
	(2.7596)
Score (Inflation)	0.9756^{***}
	(0.0012)
N	57,350

This table presents pseudo regression results examining whether firms increase their attention to inflation risk management after experiencing a breach in the previous year. The dependent variable *Score_lead (Inflation)*, is the firm's inflation risk awareness score at the end of the observation period. The independent variable, *Breach*, a binary indicator equal to 1 if a breach occurs with the observation period. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

This pseudo tests support the validity of the main findings by confirming the lack of predictive power for inflation risk awareness score and reinforcing the interpretation that firms' cybersecurity risk disclosures contain information relevant to their actual breach exposure, whereas similarly constructed but unrelated risk measure does not. Firms appear to respond to breaches in a topic-specific manner, rather than through a generic increase in all risk disclosures.

1.4.2. More tests

Additionally, appendix results show that cybersecurity breaches are associated with increases in cash flow, both contemporaneously and in the subsequent period, suggesting possible efficiency improvements or resource reallocation following an incident. However, breaches do not have significant effects on other firm fundamentals such as ROA, leverage, or R&D expenditures.

Furthermore, to capture the spillover effect more accurately, I refine the cybersecurity awareness score using ChatGPT to identify disclosure components that focus specifically on risk management. The enhanced score confirms that firms emphasize risk management strategies more after both own and peer breach incidents. This suggests that firms do internalize the signals from competitor incidents—even when those incidents do not directly raise their own cybersecurity risk and proactively strengthen their risk management.

1.5. Conclusion

Awareness of risk is a vital prerequisite for managing it effectively (Berkman et al., 2018). To measure how firms respond to cybersecurity threats, whether their own or those of competitors, it is essential to first assess whether firms recognize these risks. Firms that acknowledge cybersecurity risks are expected to have comprehensive strategies in place to prevent, mitigate, and respond to breach incidents. Evidence of this cybersecurity risk awareness and strategic plan is often reflected in their annual disclosures to the SEC.

This study contributes to the literature by applying a novel measure of cybersecurity risk based on firm-level 10-K disclosures. Unlike prior research that assigns awareness scores to fixed calendar years, this study uses firm-specific 10-K filing timelines to more accurately align disclosure behavior with breach incidents. This improves the precision of measuring how firms update their cybersecurity disclosures in response to breach incidents, offering deeper insights into the timing, magnitude,

and intent of cybersecurity risk disclosure behavior.

First, this study demonstrates that firms are more likely to experience future breaches if they have been breached in the past, highlighting the persistence of cybersecurity vulnerabilities. Both a firm's own breaches and those of its competitors serve as warning signals, conveying valuable information about future cybersecurity risks. Firms respond to these events by strategically increasing their cybersecurity awareness in subsequent 10-K disclosures, reflecting heightened concern and proactive risk management engagement. Moreover, higher cybersecurity awareness scores are positively associated with the likelihood of future breaches, indicating that firms recognize and communicate their heightened exposure to cybersecurity risks.

In summary, this paper provides novel evidence on how firms use disclosures to signal cybersecurity awareness, respond strategically to breach events, and potentially benefit from regulatory transparency. The findings emphasize the value of mandated disclosures in shaping firm behavior and fostering a more resilient corporate response to rising cybersecurity risks. Future research could expand this approach by incorporating data on cyber insurance coverage, detailed cyber loss records, and specific incident characteristics that most affect firm behavior. Enhanced reporting and aggregation of cybersecurity risk regulations and incidents data would aid future researchers in assessing, comparing, and mitigating cybersecurity risks across firms and industries.

CHAPTER 2

CAN CHATGPT REVEAL MORE? ASSESSING CYBERSECURITY RISK ANALYSIS IN CORPORATE DISCLOSURES

2.1. Introduction

AI is affecting our lives in multiple aspects, from supporting user interactions on digital platforms to improving deep learning models in research. As AI capabilities become widely available for public use, the integration of natural language processing tasks with AI poses new opportunities and challenges. Traditional dictionary-based approaches in sentiment analysis often fall short in capturing the syntactical relationships between words. However, the Large Language Models (LLMs) for natural language processing can capture both lexical meanings of words and syntactical interactions among words to recognize subtle emotions and contexts. GPT models have demonstrated their capabilities in a wide range of natural language processing tasks, including text generation, question answering, language translation, and sentiment analysis (Ray, 2023). ChatGPT has demonstrated effectiveness as a sentiment analyzer, particularly in understanding the opinions, sentiments, and emotions expressed in text (Wang et al., 2023).

While new and powerful, LLMs also have limitations and present new challenges. Additionally, research exploring the potential of LLMs for in-depth qualitative data analysis is still in its early stages. This reality motivates this paper's broad research question: Can LLMs potentially offer significant potential for improving business decision-making by allowing both firms and investors to learn better and faster from vast quantities of data? This paper explores the applications of LLMs for cybersecurity research and compares the performance of different LLMs models to the traditional dictionary model in sentiment analysis.

The first chapter has shown that cybersecurity awareness scores, quantified by thefrequency of cybersecurity-related terms in SEC Form 10-K filings item 1A, have predictive power in identifying

firms at higher risk of breaches. This paper extends that analysis by evaluating the ability of LLMs to predict cybersecurity risks through sentiment analysis of the same text. I evaluate the performance of various LLMs, including GPT-4, Llama3, and finBERT, against traditional dictionary methods in analyzing the sentiment of cybersecurity-related disclosures in Item 1A of firm filings to assess firms' future cybersecurity risk-taking behavior. This comparison examines LLMs' capacity for processing complex textual information and extracting new information about cybersecurity compared to traditional dictionary methods. This study is the first to systematically evaluate the effectiveness of LLMs in analyzing sentiment analysis related to cybersecurity risks, exploring their potential to predict the likelihood of future breaches. This paper starts by constructing LLM-based sentiment scores from firms' annual disclosures, focusing specifically on the risk-related documents found in Item 1A of the SEC Form 10-K. Among the models, GPT-based sentiments are sensitive to textual variations scores and accurately reflect major economic trends, demonstrating GPT's capability to capture economic uncertainty and the associated impacts on financial disclosures.

Then I compare the results of sentiment analysis performed by various LLMs using different prompts, as well as the traditional dictionary method, to the cybersecurity awareness analysis in the first chapter. This study investigates similar reseach questions: whether the sentiment of firms' risk disclosures, measured using various sentiment analysis tools, can predict future cybersecurity breaches and whether firms adjust this tone in response to breach incidents. Using sentiment scores derived from models such as GPT, Llama3, finBERT, and the VADER dictionary, the analysis provides mixed evidence. First, firms that use a more positive tone in their disclosures, particularly as measured by GPT sentiment, are more likely to experience future breaches. This suggests that overly optimistic language may unintentionally signal underlying vulnerability, potentially due to underreporting of risk, overconfidence in cybersecurity risk management, or efforts to reassure investors at the expense of transparency. Second, despite this predictive power, firms do not appear to significantly adjust the tone of their disclosures after experiencing a breach.

Furthermore, this study examines how firms adjust the sentiment in their risk disclosures in response to competitors' breaches. There is no significant change in sentiment in response to breaches by industry peers, consistent with the posture they maintain regarding their own post-breach behavior. This cautious approach is likely adopted to avoid triggering investor concern or revealing potential security weaknesses. When addressing cybersecurity risks in 10-K fillings, firms may strategically highlight their cybersecurity infrastructure investments, enhanced protection policies, and adherence to industry regulations rather than acknowledging heightened vulnerability. Risk perception in financial disclosure is also conveyed through the relationships between sentences and paragraphs, poses significant challenges for current sentiment analysis techniques. Both LLM-based and traditional dictionary approaches struggle decode these strategic business communications, which go beyond simple positive or negative words and syntactical relationships. This is consistent with findings in legal studies. Ash et al. (2022) demonstrate that sentiment in judicial rulings remains relatively stable over time and across different court circuits, despite variations in intensity. Similarly, in financial disclosures, the strategic and standardized nature of corporate language makes it difficult for sentiment analysis models to detect subtle variations in tone, limiting their ability to extract meaningful insights. The finding is also consistent with the inherent caution in legal writing, which often intentionally obscures or neutralizes sentiment. When firms disclose only mundane or boilerplate tax-audit information, there is no clear impact on stock prices. However, announcements of audits by tax authorities lead to a decline in cumulative abnormal returns (Choi and Kim, 2024).

Cybersecurity awareness scores increase with perceived risk, while sentiment scores may reflect deliberate understatements. Firms may expand risk discussions while keeping tone positive, underscoring the importance of jointly considering both the volume and tone of disclosure when assessing risk communication. By comparing sentiment scores with disclosure volume (awareness scores), this study highlights the importance of jointly analyzing tone and content to better understand how firms signal and manage cybersecurity risks.

This paper contributes to the growing literature on the impact of artificial intelligence (AI) technologies on firms. Prior researches have explored AI's influence across a variety of domains, including AI investments (Babina et al., 2024), AI-inventing firms (Alderucci et al., 2019), robo-advising (D'Acunto et al., 2019), fintech innovation (Chen et al., 2019), loan underwriting (Jansen et al., 2023; Fuster et al., 2022), financial analysts (Grennan and Michaely, 2020; Abis and Veldkamp, 2024; Cao et al., 2024), and entrepreneurship (Gofman and Jin, 2024). AI encompasses three major subfields: machine learning (ML), natural language processing (NLP), and computer vision (CV). A more recent stream of research centers on NLP, particularly large language models (LLMs), to analyze unstructured financial textual data at scale. Due to their ability to distill redundant narrative information and identify firm-level risks (Kim et al., 2024), LLMs have been increasingly applied in finance to evaluate corporate strategy (Jha et al., 2025), measure corporate culture (Li et al., 2021), and forecast stock returns (Jiang et al., 2023; Murray et al., 2024; Lopez-Lira and Tang, 2023). Considering that financial disclosures are often verbose and repetitive, recent studies highlight the potential of LLMs to enhance investor understanding and support more efficient decision-making by extracting meaningful signals from lengthy and complex financial documents.

This study also contributes to the growing literature at the intersection of large language models (LLMs) and financial disclosures. It is the first to examine the intersection of large language model (LLM) sentiment and cybersecurity risk. The primary goal of this study is to explore and compare the capabilities of various LLMs in evaluating sentiment related to 10-K fillings. Traditionally, sentiment analysis in finance has relied on dictionary-based methods, most notably the Loughran-McDonald Master Dictionary (Loughran and McDonald, 2022), which classifies words as positive, negative, or neutral. However, these methods are limited by their inability to consider the context of words. LLMs offer a powerful alternative by embedding words in context, enabling more accurate sentiment interpretation. For instance, Araci (2019) demonstrates that fine-tuning BERT on financial text significantly improves sentiment prediction accuracy compared to traditional models. Lopez-Lira and Tang (2023) provide evidence that firm-specific sentiment scores generated by LLMs (e.g., GPT-3 and GPT-4) are significantly correlated with stock returns, highlighting their strong

contextualization capabilities. This study offers a comprehensive comparison of multiple LLMs (GPT-4, Llama3, finBERT) against traditional dictionary sentiment analysis methods, evaluating their ability to capture subtle tone differences in firms' cybersecurity-related fillings. It provides new evidence on the relative effectiveness of advanced language models in capturing context-specific sentiment in financial risk communication.

However, applying LLMs to predictive tasks inevitably introduces challenges, including lookahead bias, limited replicability, and concerns regarding interpretability and generalizability. These issues are discussed and addressed in the following section. Admittedly, the underlying mechanisms by which LLMs enhance predictive power are not yet fully understood and remain an active area of research.

The rest of the paper is organized as follows. Section 2 describes the data, variable construction, and summary statistics. Section 3 investigates the impact of breach incidents on a firm's risk disclosure sentiment and that of its industry peers. Section 4 concludes.

2.2. Data and Variable Construction

2.2.1. Data on firm cybersecurity breaches

All states require that breach incidents be reported. The Privacy Rights Clearinghouse (PRC) collects these notifications, including a brief description of the incident, the public disclosure date, the type of breach, the type of organization, and the number of records affected when available. This study specifically focuses on breaches involving publicly traded companies, as they are required to file annual financial disclosures with the SEC. After removing duplicates, where the same breach was reported multiple times across different states, a total of 920 breach incidents involving publicly traded companies from 2005 to 2022 were identified.

2.2.2. Data on firm annual SEC disclosures

The sample of firms' annual disclosures comprise the universe of electronic 10-K fillings on SEC's EDGAR database for the period 2005-2024. For each 10-K filing, I apply a keyword-based search algorithm to identify discussions of cybersecurity for two main reasons. First, narrow down the context using the glossary of common cybersecurity words and phrases form NICCS [2] and two additional two terms (act and laws) helps the model identify pertinent information (Berkman et al., 2018). Second, there are token limits for GPT models. GPT-4 has a token limit of 8,191, while GPT-3.5-turbo-16k has a token limit of 16,384. Instead of splitting one disclosure into multiple segments for analysis, analyzing the entire disclosure with a single prompt reduces complexity and avoids potential errors arising from fragmented analysis. Since I am not processing the entire annual filings but a subset of cybersecurity-related sentences, it is better in dealing with the token limit. If the input text length exceeds the limit, I process the expert before exceeding the tokens limits. There are 71,608 observations from 8,784 firms.

2.2.3. Constructing Measures of Cybersecurity Sentiment Scores

An LLM is a machine learning model trained to understand, generate, and interact with human language. LLMs demonstrate efficient information processing capabilities across various fields. Researchers in finance increasingly utilize LLM applications, recognizing their capacity to process and analyze extensive textual and picture information. Sentiment analysis is a standout application of LLMs since it presents several benefits over existing textual analysis methods relied on simple word counts.

Recent work has made progress in examining the LLMs in textual analysis in various contexts: Fed speak(Hansen and Kazinnik, 2024), Corporate Policies (Jha et al., 2025), accounting research (de Kok, 2024), future interest rate decisions (Woodhouse and Charlesworth, 2023), AI readership (Cao et al., 2024) and stock price movement (Lopez-Lira and Tang, 2023).

[2] https://niccs.cisa.gov/cybersecurity-career-resources/vocabulary

Bag-of words

The bag-of words methods use lists of words (dictionaries) with associated tone (sentiment) and corresponding sentiment strength. For example, dictionaries, such as Harvard's General Inquirer tag categories, Loughran and Mcdonald (2011), and Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015), have been extensively used to measure the sentiment of documents. As noted by Loughran and Mcdonald (2016), bag-of words methods prioritize word counts over the sequence of words in a document. Henry and Leone (2016) indicate that word-frequency tone measures are as effective as Naïve Bayesian machine-learning tone measure from Li (2010)) in predicting the market reaction to earning announcement.

This paper uses dictionary-based sentiment tool, VADER (Valence Aware Dictionary and Sentiment Reasoner), which assigns sentiment scores by combining a pre-defined lexicon of words rated for their positive, neutral, or negative sentiment, along with rules that account for punctuation, capitalization, degree modifiers and conjunctions. Unlike traditional bag-of-words dictionaries, VADER is sensitive to both polarity (positive/negative) and intensity (strength) of emotion, making it well-suited for texts that rely on subtle cues to convey tone. Although originally designed for general-purpose text, VADER's lexicon-based approach has been adapted for a variety of domains. In this paper, the VADER dictionary sentiment score is calculated as the difference between the total negative and positive word intensities. To account for variation in filing length and sentiment word coverage, the raw score is normalized by dividing it by the number of covered tokens (i.e., total tokens minus uncovered tokens) to ensure comparability across fillings with varying lengths and proportions of sentiment-related words. As a result, higher values indicate more negative sentiment.

A critical challenge for bag-of-words methods lies in their limited ability to extract nuanced information by analyzing contextual meaning. This is particularly crucial in the sentiment analysis of financial documents. First, the frequently used industry-specific jargon and terms with unique

financial connotations require sentiment analysis to be correctly interpreted within their specific context. Second, unlike news reports, financial writing expresses sentiment subtly and indirectly and avoid direct emotional language. Even slight changes in language when discussing risks can significantly impact sentiment. Third, financial documents are often structured according to a consistent annual format and use formal, cautious language in compliance with regulatory requirements. This consistent format and standardization make it challenging to extract incremental information from year to year. Thus, the potential benefits of LLMs for texture analysis of financial documents remain to be fully explored.

GPT

GPT (Generative Pre-training Transformer), is an autoregressive language generation model that predicts text sequentially, one word at a time. It employs a unidirectional training approach where each new word is predicted based on the preceding text. This method, known as causal language modeling, is designed to emulate the natural human process of constructing sentences in a forward-flowing manner. In sentiment analysis, GPT's predictive capabilities enable it to effectively interpret the emotional tone of a text based on its sequential structure. Designed to generate coherent and contextually relevant text, GPT excels at tasks such as text completion, creative writing, and code generation. This study mainly uses a specific version of gpt-4-turbo for evaluation, given its lower cost and improved performance (as stated in the OpenAI documentation [3]). Considering the API usage costs and the length of 10-K disclosures, GPT-based sentiment analysis is conducted on all breached firms and a 10% random sample of non-breached firms.

Other LLMs

BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer), and Llama (Large Language Model by Meta) are all based on the transformer architecture, which leverages attention mechanisms to process text in a way that captures deep contextual relationships within data. Thus, to explore the effectiveness and precision of LLM sentiment analy-

[3] https://platform.openai.com/docs/models

sis, I also employed Llama3 and finBERT. FinBERT, accessible via Hugging Face, is a BERT-based model that has been specifically fine-tuned for sentiment analysis in the financial domain. While its training on financial corpora makes it well-suited for understanding the complex jargon and nuanced context within 10-K filings, finBERT's context window limitation of 512 tokens presents a challenge. This token limit requires truncating longer texts, which is often the case after extracting cybersecurity-related excerpts from Item 1A risk factors. Consequently, this truncation may lead to the loss of critical contextual information at the end of documents, which might affect the accuracy of the sentiment analysis. Llama3 is available for free and capable of analyzing large volumes of text. Thus, it can carry the comprehensive analysis of all firms' annual cybersecurity-related excerpts from Item 1A risk factors in 10-K filings. By utilizing three LLM models as well as the traditional dictionary sentiment analysis model, this paper provides a well-rounded approach to analyzing the tone of cybersecurity-related reporting in corporate disclosures.

After pre-processing the textual data of firms' annual disclosure, I use predefined prompts to all LLMs to obtain the LLM-based sentiment score. The hyperparameter, temperature is set to be 0 to maximize the reproducibility of the results and limit computational rumination. Finally, this paper employs three categories— 1 for negative, 2 for neutral, and 3 for positive—to label sentiments, using the strategies of analysis from financial experts. Detailed descriptions of the prompts and examples of LLM results applied to selected excerpts are provided in the Appendix.

Look ahead bias

Pretrained language models are trained on extensive datasets of historical text—including content from the broader internet, which may embed information about past and future events (Radford et al., 2018; Devlin et al., 2019). As a result, analyses that rely on these model may exhibit lookahead bias, where the model unintentionally incorporates knowledge of events that occur after the text being analyzed. Recent studies show that prompt-based approaches do not fully eliminate this bias (Glasserman and Lin, 2023; Lopez-Lira and Tang, 2023; Halawi et al., 2024).

Glasserman and Lin (2023) also highlight a related challenge in sentiment analysis known as the distraction effect, in which an LLM's general knowledge about a firm may interfere its sentiment evaluation. This interference could arise from the model's knowledge of future information or historical context. Importantly, the direction of this effect is ambiguous.

In this study, to mitigate potential biases when assessing firm-level sentiment from 10-k fillings, I adopt an anonymization strategy as suggested by Glasserman and Lin (2023). Specifically, I extract only cybersecurity-related terms from the 10-K risk factor sections and omit identifiable firm information such as company names, CIK codes, and filing dates. While this approach cannot fully eliminate lookahead bias and the distraction effect, it limits the model's access to firm-specific context, helping improve the reliability of sentiment analysis.

2.2.4. Summary statistics

To create lag and lead sentiment scores, firms lacking three-year data periods and observations missing these scores were excluded. The final dataset including all firms in the market consists of 59,531 observations from 7,762 firms. The Llama 3-based sentiment analysis and dictionary-based sentiment analysis are based on this dataset. The GPT-based sentiment analysis is performed on a sample that includes all breached firms (2,763 observations from 241 firms) and 10% of randomly selected firms with no breach history (855 firms). After removing missing observations from the combined dataset, the analysis includes 6,594 observations from 803 firms. In Table 2.1 , I report on descriptive statistics of the variables used in the analyses.

I compared the GPT based sentiment score with Cybersecurity term frequency score (Cybersecurity Awareness Score in Chapter 1). The mean for Cybersecurity term frequency score show an increasing trend from lag (452.72) to current (477.78) and lead (495.65), which suggests a gradual

improvement in cybersecurity awareness over time. The large standard deviation indicates significant variability in cybersecurity awareness among firms, likely reflecting differing levels of adoption or emphasis on cybersecurity practices. However, the GPT based sentiment scores are remarkably consistent, with all median and quartile values tightly clustered around 3, on a scale. There is a generally neutral or mildly negative sentiment persistently maintained across the sample.

Panel A of Table 2.1 shows sentiment patterns derived from various text analysis methods. The sentiment scores from large language models (LLMs), GPT-4, Llama3, and finBERT, are based on a 3-point ordinal scale where higher values reflect more negative tone. All three models show similar patterns, with median scores of 3, indicating that cybersecurity risk disclosures tend to be cautious or negatively framed. The VADER dictionary-based sentiment score is a continuous, normalized measure constructed as the net tone, negativity minus positivity, divided by the number of covered tokens. While not directly comparable to the ordinal LLM scales, VADER scores range from –0.48 to 0.59 in the whole sample, with higher values indicating more negative sentiment. The distribution is tightly centered around a slightly negative mean (–0.09), suggesting again limited variation in tone across disclosures. It is worth noting that the overall variation in sentiment scores, captured from boilerplate corporate disclosures, may be limited. This raises concerns about the extent to which meaningful sentiment signals can be extracted, particularly from firms that might recycle similar language across years when reporting the fillings.

Panel B of Table 2.1 summarizes the breach-related activity across firms. While the average number of breaches per firm across the full sample is low (0.04), indicating that most firms do not experience any breaches, 241 firms experience breaches. Among those breached firms, the average number of years with breaches is 1.39, suggesting that breaches are not isolated events but may recur over time. Moreover, in breach years, firms experience an average of 1.14 breaches, showing that while most breach years involve a single incident, multiple breaches within the same year are not uncommon. These patterns underscore the persistence and concentration of cybersecurity risk among affected

firms since they tend to cluster in multiple years or involve multiple incidents once they occur.

Table 2.1: Summary statistics

Variables	Mean	Std	Q1	Median	Q3	N
Panel A. Textural Aanalysis Data Variables						
Cybersecurity term frequency score	463	328	236	392	610	59,531
GPT sentiment score	2.72	0.53	3	3	3	6,594
Llama sentiment score	2.88	0.36	3	3	3	59,531
finBERT sentiment score	2.67	0.54	2	3	3	59,531
VADER dictionary sentiment score	-0.09	0.06	-0.13	-0.09	-0.06	59,531
Panel B: Breach Summary — 7,762 Firms (3% Experienced at Least One Breach)						
Total breaches per firm	0.04	0.30	0	0	0	7,762
Years with breach per breached firm	1.39	0.99	1	1	1	241
Years to first breach per breached firm	5.42	4.63	1	4.02	8	241

This table reports the summary statistics of key variables. *Cybersecurity term frequency score* is the count of firm mentions of cybersecurity risk phrases (NICCS-defined) in SEC 10K Item 1A, which is the *cybersecurity awareness score* in the first chapter. *GPT sentiment score* reflects the sentiment (positive = 1, neutral = 2, negative = 3) expressed in SEC 10K Item 1A, as analyzed by GPT. Similarly, *Llama sentiment score* and *finBert sentiment score* are obtained from Llama and finBert, respectively. *VADER dictionary sentiment score* is normalized by text length (based on covered tokens) and reflect net tone (negative minus positive); higher values indicate more negative sentiment. The table reports the mean, median, standard deviation, 25th percentile (Q1), 75th percentile (Q3), and number of observations (N).

2.2.5. Time trends

Major economic disruptions likely influenced the tone of corporate financial disclosures. Figure 1 reports the time-series trend in GPT sentiment scores (with 3 being most negative and 1 being most positive). Two distinct troughs in GPT sentiment scores are apparent.

First, following the 2008 financial crisis, subsequent annual financial disclosures exhibited a pronounced negative shift in GPT sentiment scores. This downturn reflects a more constrained tone in the disclosures, corresponding with the pervasive economic anxieties and uncertainties of that period. Second, a similar trend is observed during the COVID-19 pandemic in 2020 in the United State. Figure 1 provides strong evidence that the tone in financial disclosures does reflect the economic environment, and firms increasingly use concerned language as they address the challenges and disruptions following crises.

However, notable differences emerge in the trends of sentiment scores by each sentiment measurement. Llama remains relatively stable and consistently reports the highest sentiment scores (most negative). finBERT, which is specifically tuned to understand financial language, presents the lowest sentiment scores (most positive). This suggests that finBERT may be less sensitive to negative tones in financial texts compared to the other models, possibly due to its specialized training on financial lexicon and contexts, leading to interpretations that differ from those of general language models.

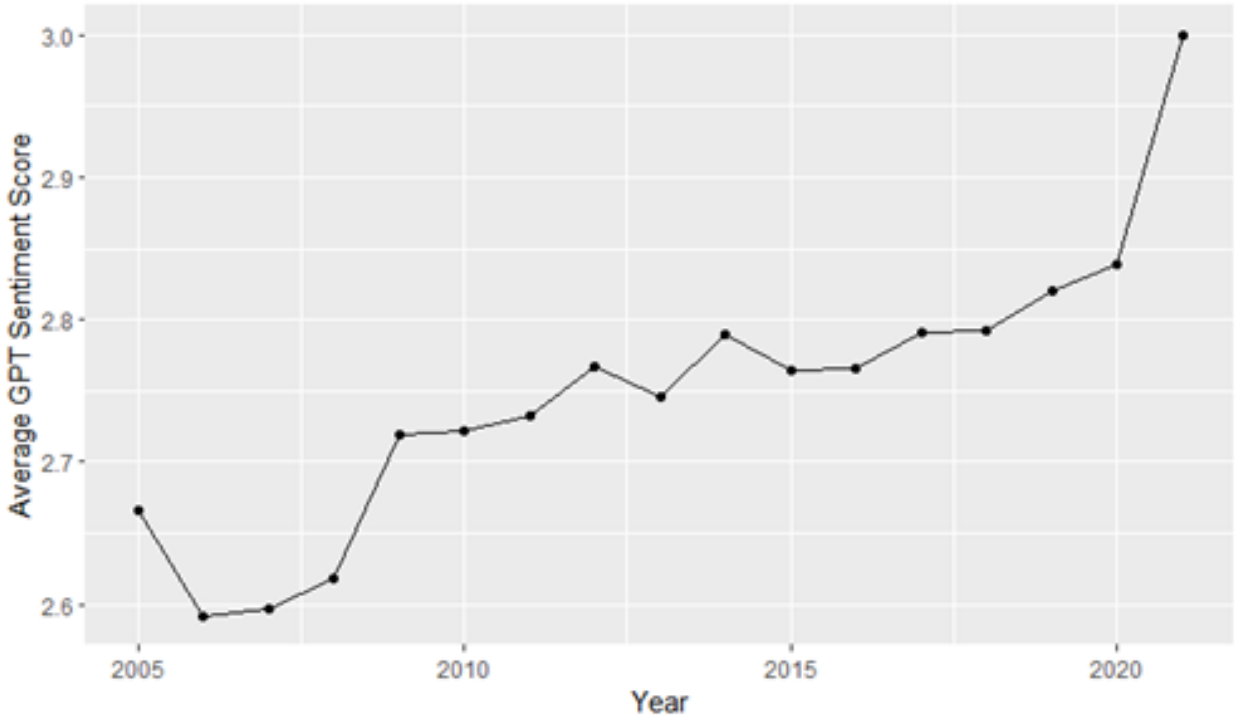

Figure 2.1: Time trends: This figure plots the average GPT sentiment score in 10-Ks for each year, based on a sample of 6,594 observations from 803 firms. A score closer to 3 reflects a more negative tone, whereas a score closer to 1 indicates a more positive tone.

2.3. Regression analysis

This paper begins the empirical analysis by evaluating the predictive capacity of various sentiment measures for future cybersecurity breach incidents. Initial results from chapter one indicate that firms with higher cybersecurity awareness scores, measured by the frequency of cybersecurity-related terms in their 10-K fillings, are more likely to experience future breaches. Building on the informativeness of this cybersecurity awareness score, the analysis is extended to assess whether the sentiment embedded in these same 10-k fillings also serves as a meaningful predictor of future breach incidents.

2.3.1. Own subsequent breaches

Do firms' sentiment in 10-K fillings predict their own subsequent breaches?

While previous analysis from chapter one shows that firms signal their cybersecurity exposure by mentioning relevant terms more frequently, it remains unclear whether the tone of those disclosures adds meaningful insight. By extending the previous analysis and applying the same regression framework to sentiment scores, I test whether tone provides incremental predictive power over word-count-based measures.

$$\log\left(\frac{P(Breach_{i,[t:t+\Delta_i t]}=1)}{1-P(Breach_{i,[t:t+\Delta_i t]}=1)}\right)=Score_{i,t}$$

Table 2.2 presents a comparative summary of four logistic regression analyses assessing whether different sentiment and frequency-based scores can predict future breach incidents. The results show that the GPT sentiment score significantly predicts breach likelihood in the matched sample (Panels A and B), with more negative sentiment associated with a lower probability of future breaches. Additionally GPT may better capture subtle warning signals embedded in the text compared to other models. Therefore, firms that speak more cautiously or conservatively about cybersecurity risk may be more proactive in risk management or more transparent, possibly helping them reduce

their exposure.

Awareness scores reflect the volume and focus of disclosure, which tends to increase when firms recognize or anticipate heightened cybersecurity risk. In contrast, sentiment scores reflect the framing of that risk, a more positive tone may indicate that firms are underplaying threats, potentially exposing themselves to future incidents. Thus, firms may increase their cybersecurity discussion (increasing awareness) while adopting a more cautious tone (higher sentiment score). A negative tone in 10-K fillings does not necessarily signal panic or bad news, it may signals awareness, preparedness, or proactive communication. Firms that acknowledge risk clearly,rather than gloss over it, may be more likely to strengthen their cybersecurity risk management in response. This additional insights underscores the value of using both frequency and tone to capture the full nuance of corporate risk communication.

Additionally, GPT was not applied to the full sample. Among the remaining sentiment measures, finBERT shows statistically significant results in the full sample regression (Panel A, B, C), indicating that its sentiment classifications carry predictive value in broader settings. In contrast, sentiment scores derived from Llama3 and the VADER dictionary are not statistically significant in any specification, suggesting limited predictive utility. The negative coefficients of finBERT sentiment score observed in Panel C are consistent with those in Panels A and B.

Table 2.2: Do firms' sentiment in 10-K fillings predict their own subsequent breaches?

Dependent Var: Breach					
Score Types:	Frequency		Sentiment		
Score Names:	Cybersecurity term	GPT	Llama3	finBERT	VADER dictionary
Panel A: Sample — 10% Non-Breached Firms and All Breached Firms (Regression with Weights)					
Past Score Coefficient	0.00075***	-0.46***	0.26	-0.27*	0.07
	(0.00015)	(0.10)	(0.24)	(0.11)	(1.10)
N	6,593	6,593	6,593	6,593	6,593
Panel B: Sample — 10% Non-Breached Firms and All Breached Firms (Regression without Weights)					
Past Score Coefficient	0.00075***	-0.46***	0.26	-0.28*	0.07
	(0.00015)	(0.10)	(0.24)	(0.11)	(1.12)
N	6,593	6,593	6,593	6,593	6,593
Panel C: Sample — All SEC Filings					
Past Score Coefficient	0.00082***		0.04	-0.21*	1.26
	(0.00013)		(0.16)	(0.09)	(0.87)
N	59,526		59,526	59,526	59,526
Panel D: Sample — Breached Firms Over 17 Years					
Past Score Coefficient	0.00043***	-0.04	0.03	-0.09	-0.37
	(0.00015)	(0.09)	(0.16)	(0.11)	(0.92)
N	2,763	2,763	2,763	2,763	2,763

This table presents regression results examining whether evaluating whether firms' cybersecurity-related disclosures can predict their own future breach incidents. The dependent variable *Breach*, an indicator for whether the focal firm was breached. Past Score is the value from the firm's most recent filing before the current breach, based on either cybersecurity term frequency or sentiment scores generated using GPT, Llama3, finBERT, and the VADER sentiment dictionary. Panel A and Panel B restrict the sample to all breached firms and a 10% stratified sample of non-breached firms, with and without regression weights, respectively. Panel C uses the full set of SEC 10-K filings across all firms and years. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Do firms' sentiment in 10-K fillings respond to their past breaches?

However, a key question remains: if a firm experiences a breach, does it subsequently convey greater concern in its 10-K fillings? In other words, do breach incidents lead to a change in the tone of future disclosures? The next analysis examines the effect of breach incidents on the sentiment in firms' risk-related fillings. Prior findings using cybersecurity awareness scores suggest that firms increase their disclosure volume following a breach, signaling heightened attention to cybersecurity risk. However, contrary to expectations, the results here show that firms do not significantly alter the tone of their public risk communications after experiencing a breach, at least not in a manner detectable by the various sentiment analysis models. This may be due to the overall nature of financial disclosures, which are typically written in a consistently cautious or concerned tone. This makes it difficult to discern any subtle shifts in tone following a breach. Thus, while the content of disclosures may change (reflected in higher awareness scores), the tone remains relatively stable.

These results from Table 2.3 highlight a critical distinction between how firms signal risk and how they respond to it. While firms with more negative sentiment (higher sentiment scores) are less likely to experience future breaches, they do not significantly change the sentiment of their disclosures after a breach occurs. Instead, firms respond to breaches by increasing the volume of cybersecurity-related language, as captured by the awareness score. This suggests that firms tend to maintain a consistent, cautious tone across disclosures, possibly to manage investor perception or comply with regulatory expectations, while adjusting the depth or emphasis of content to reflect evolving risks. Interestingly, the GPT-based sentiment score is statistically significant in the unweighted matched sample (Panel B), but not in the weighted sample (Panel A). Results in Panel B, which uses an unweighted matched sample, show that the GPT-based sentiment score exhibits a statistically significant and negative coefficient following breach incidents. This suggests that firms adopt a more cautious or conservative tone in their 10-K disclosures after experiencing a breach. However, this result should be interpreted with caution. The unweighted model treats all observations equally, despite the sampling design that includes all breached firms and only a 10% random sample of non-

breached firms. This overrepresents breached firms relative to their true population proportion, potentially exaggerating firm-level effects that are not generalizable. Overall, the findings suggest that firms tend to maintain the structure of their risk-related 10-K fillings but may refine their tone in more nuanced ways after experiencing a breach.

$$Sentiment\ Score_{i,t+\Delta t} = Breach_{i,[t:t+\Delta t]} + Sentiment\ Score_{i,t} + Cybersecurity\ term\ frequency_{i,t}$$

Table 2.3: Do past breaches affect how firms communicate about cybersecurity risks in 10-K fillings

Dependent Var: Score					
Score Types:	Frequency		Sentiment		
Score Names:	Cybersecurity term	GPT	Llama3	finBert	VADER dictionary
Panel A: 10% Non-Breached Firms + All Breached Firms (Weighted)					
Past Breach	26.96	-0.039	-0.0092	0.0047	0.002
	(23.02)	(0.089)	(0.041)	(0.077)	(0.0055)
Past Sentiment Score		0.43***	0.62***	0.65***	0.84***
		(0.011)	(0.0088)	(0.0094)	(0.0061)
Past Cybersecurity Term	0.96***	0.00006***	0.000037***	-0.000016	-0.0000021*
	(0.0045)	(0.000017)	(0.0000082)	(0.000015)	(0.000001)
N	6,593	6,593	6,593	6,593	6,593
Panel B: Same Sample (Unweighted)					
Past Breach	27.08***	-0.04***	-0.009	0.0047	0.002
	(7.78)	(0.029)	(0.014)	(0.025)	(0.0018)
Past Sentiment Score		0.43***	0.62***	0.66***	0.84***
		(0.011)	(0.0088)	(0.0094)	(0.0061)
Past Cybersecurity Term	0.96***	0.00006***	0.000035***	-0.000016	-0.000002
	(0.0046)	(0.000017)	(0.000008)	(0.000015)	(0.000001)
N	6,593	6,593	6,593	6,593	6,593
Panel C: Sample — All SEC Filings					
Past Breach	23.21**		0.0076	0.0097	0.0027
	(7.48)		(0.015)	(0.025)	(0.0021)
Past Sentiment Score			0.45***	0.62***	0.72***
			(0.0031)	(0.0035)	(0.0024)
Past Cybersecurity term	0.93***		0.000042***	0.0000097	-0.0000086***
	(7.48)		(0.000036)	(0.0000006)	(0.0000005)
N	59,526		59,526	59,526	59,526
Panel D: Breached Firms Over 17 Years					
Past Breach	17.79*	0.012	0.0044	0.027	-0.00062
	(8.90)	(0.034)	(0.015)	(0.025)	(0.0021
Past Sentiment Score		0.39***	0.42***	0.69***	0.66***
		(0.017)	(0.013)	(0.014)	(0.01)

Dependent Var: Score					
Score Types:	Frequency	Sentiment			
Score Names:	Cybersecurity term	GPT	Llama3	finBert	VADER dictionary
Past Cybersecurity Term	0.93**	0.000044	0.000018	0.000012	-0.000011***
	(0.0080)	(0.000031)	(0.000014)	(0.000022)	(0.0000019)
N	2,763	2,763	2,763	2,763	2,763

This table presents regression results examining whether firms following own sentiment scores respond to their current cybersecurity breaches. The dependent variable, Past Score refers to the value from the firm's most recent filing before the current breach, while Score refers to the value from the most recent filing after the breach. Both scores are based on cybersecurity term frequency and sentiment scores generated using GPT, Llama3, finBERT, and the VADER sentiment dictionary. The key independent variables are *Breach*, an indicator for whether the focal firm was breached. Panel A and Panel B restrict the sample to all breached firms and a 10% stratified sample of non-breached firms, with and without regression weights, respectively. Panel C uses the full set of SEC 10-K filings across all firms and years. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

2.3.2. Spillover effect: Competitors' breaches

Do firms' sentiment in 10-K fillings responds to their competitors' past breaches?

Furthermore, the analysis examines whether firms respond to peer breaches by changing how they communicate cybersecurity risks in their 10-K filings. Prior literature suggests that breach incidents convey valuable information to the market and may motivate firms to bolster their own cybersecurity posture. However, the results from Table 2.5 reveal mixed evidence. Just as firms do not significantly alter the tone of their disclosures after experiencing their own breaches, they also show mixed sentiment shift in response to competitor breaches depending on the model. While GPT and finBERT sentiment scores show no significant tone change, Llama3 and VADER scores detect subtle yet statistically significant shifts, suggesting that some models capture more positive tone adjustments in language following peer breaches. These changes are modest, indicating that tone spillovers might exist but are limited and highly model-dependent.

Such adjustments may reflect strategic communication behavior, where firms aim to maintain stakeholder confidence while subtly signaling broader risks. These results also highlight the importance of using multiple sentiment models to detect subtle shifts, as traditional and transformer-based models differ in sensitivity to context and tone.

$$\text{Score}_{i,t+\Delta_{i,t}} = \textit{Breach}_{j,[t,t+\Delta_{i,t}]} + \textit{Breach}_{i,[t,t+\Delta_{i,t}]} + \text{Score}_{i,t}$$

Table 2.5: Do firms' sentiment in 10-K fillings responds to their competitors' past breaches?

Dependent Var: Score					
Score Types:	Frequency	Sentiment			
Score Names:	Cybersecurity term	GPT	Llama3	finBERT	VADER dictionary
Panel A: 10% Non-Breached + All Breached Firms (Weighted)					
Past Rival Breach	1.57	-0.02	-0.016*	0.005	-0.0017*
	(3.64)	(0.015)	(0.0063)	(0.013)	(0.00083)
Past Breach	30.36	-0.033	-0.007	0.008	0.0018
	(22.38)	(0.09)	(0.039)	(0.078)	(0.0051)
Past Score	0.96***	0.44***	0.64***	0.66***	0.85***
	(0.0045)	(0.011)	(0.009)	(0.01)	(0.0061)
N	6,593	6,593	6,593	6,593	6,593
Panel B: Same Sample (Unweighted)					
Past Rival Breach	0.87	-0.023	-0.017**	0.009	-0.0017*
	(3.75)	(0.015)	(0.0063)	(0.013)	(0.00082)
Past Breach	30.47***	-0.033	-0.007	0.0075	0.0018
	(7.63)	(0.03)	(0.013)	(0.026)	(0.0017)
Past Score	0.97***	0.44***	0.64***	0.67***	0.85***
	(0.0046)	(0.011)	(0.009)	(0.0097)	(0.0061)
N	6,593	6,593	6,593	6,593	6,593
Panel C: All SEC Filings					
Past Rival Breach	8.00***		-0.00060	-0.0019	-0.003***
	(1. 43)		(0.0027)	(0.0050)	(0.00039)
Past Breach	24.00**		0.0072	0.018	0.0013
	(7.46)		(0.014)	(0.026)	(0.0020)
Past Score	0.94***		0.44***	0.63***	0.72***
	(0.0018)		(0.0031)	(0.0038)	(0.0025)
N	59,526		59,526	59,526	59,526
Panel D: Breached Firms Over 17 Years					
Past Rival Breach	-0.062	-0.050	-0.019	0.037*	-0.004*
	(0.028)	(0.026)	(0.011)	(0.019)	(0.0016)
Past Breach	-0.00025	0.014	0.0072	0.028	-0.0014
	(0.038)	(0.036)	(0.015)	(0.025)	(0.002)

Dependent Var: Score					
Score Types:	Frequency	Sentiment			
Score Names:	Cybersecurity term	GPT	Llama3	finBERT	VADER dictionary
Past Score	0.000081	0.39***	0.43***	0.69***	0.65***
	(0.000035)	(0.018)	(0.013)	(0.015)	(0.011)
N	2,763	2,763	2,763	2,763	2,763

This table presents regression results examining whether firms adjust how they discuss cybersecurity in 10-K fillings after breaches occur at peer firms. Past Score refers to the value from the firm's most recent filing before the current breach, while Score refers to the value from the most recent filing after the breach. Both scores are based on cybersecurity term frequency and sentiment scores generated using GPT, Llama3, finBERT, and the VADER sentiment dictionary. The key independent variables are *Rival Breach*, an indicator for whether any industry peer experienced a breach in year t, and *Breach*, an indicator for whether the focal firm was breached. Panel A and Panel B restrict the sample to all breached firms and a 10% stratified sample of non-breached firms, with and without regression weights, respectively. Panel C uses the full set of SEC 10-K filings across all firms and years. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

2.4. Conclusion

This paper reveals nuanced insights into the predictive capabilities of various sentiment measures for future cybersecurity breaches and examines how firms adjust their sentiment of risk disclosures in response to such events. Consistent with prior researches that LLMs are better at capturing contextual meaning, the results suggest that GPT-based sentiment scores more effectively predict breach incidents than alternatives such as Llama, finBERT, or traditional dictionary-based methods. This study provides mixed evidence on the relationship between firms' sentiment in their risk disclosures and cybersecurity breach incidents. First, sentiment scores, particularly using GPT and finBERT, are predictive of future breaches. Firms conveying a more negative tone are less likely to experience own subsequent breaches. Second, however, firms do not appear to meaningfully adjust the tone of their risk disclosures after experiencing a breach. Finally, there is mixed evidence that firms change the sentiment of their disclosures in response to peer breaches.

In summary, these findings highlight that sentiment scores capture the tone of disclosure, while awareness scores reflect the volume and extent of disclosure.Firms may increase cybersecurity discussions yet adopt a more positive tone to manage perception. Using both measures offers complementary insights into how firms communicate risk. Moreover, although the tone of disclosures may unintentionally signal vulnerability, firms appear reluctant to explicitly express cybersecurity concerns in the sentiment of their public disclosures. The tone of financial disclosures remains relatively stable following a breach. This consistent tone of inherent cautiousness across financial disclosures making it challenging to detect any subtle shifts in sentiment post-breach. This cautious

approach extends to how firms sentiment in the risk disclosures react to competitor breaches. This reinforces the overall pattern in the firm's annual discourse, where firms appear to maintain a steady, strategically cautious communication approach. This strategic communication, while potentially effective in managing public perception, poses a significant challenge for sentiment analysis. Both the LLM-based sentiment measures and the traditional dictionary sentiment measure fall short in capturing the nuanced communication tactics used in firms' financial filings.

Future research could delve deeper into these subtle business communication strategies—such as keyword choices, rhetorical framing, and variations in tone, to better understand how firms frame risk-related events and position themselves relative to competitors. Additional insights could be gained by analyzing alternative communication channels, including investor presentations, press releases, earnings calls, and social media posts. These may offer a more complete picture of how firms manage their sentiment about cybersecurity risks. Finally, Doshi et al. (2025) propose an innovative approach to aggregating predictions from artificial evaluators, such as various LLMs, roles, and prompts, to evaluate strategic decisions. Their findings suggest that aggregated outputs from multiple LLM evaluations closely resemble those of human experts. Applying such aggregation methods in LLM sentiment analysis could enhance our understanding of how firms communicate and make strategic decisions in response to cybersecurity risks and other forms of operational risk.

CHAPTER 3

THE IMPACT OF DATA BREACHES ON STOCK RETURNS

3.1. Introduction

Digital technologies have driven transformative changes across all industry over the past several decades. More and more organizations and individuals rely on the internet to conduct business transactions. On the one hand, digitalization enhances productivity and improves management practices. On the other hand, it exposes businesses and households to significant cybersecurity threats, with cybersecurity risks emerging as the most common and severe risk. Cybersecurity risks are a form of operational risk to information and technology assets that have consequences for the confidentiality, availability, and integrity of information and information systems (Cebula and Young, 2010). In 2011, the U.S. Securities and Exchange Commission (SEC) issued guidance urging public companies to disclose cybersecurity risks to investors. According to a 2020 report by the Identity Theft Resource Center,[4] there were 1,108 publicly reported breach events that exposed over 300 million exposed records containing sensitive information. Although the number of breaches and affected individuals declined compared to the peak years of 2017–2018, the threat remains substantial. Companies are increasingly prioritizing cybersecurity risk management, not only due to the exposure of confidential information but also because of the significant financial consequences of data breaches. The 2020 Cost of a Data Breach Report,[5] conducted by the Ponemon Institute for IBM, indicated that the global average total cost of a data breach was $3.86 million. Industries subject to stricter regulatory standards,such as healthcare, energy, financial services, and pharmaceuticals, tend to incur even higher breach-related costs.

Data breaches have serious and lasting consequences. They erode consumer trust, damage a company's reputation, and most importantly, the impact could last for years. In this chapter, I investigate whether and how data breaches affect stock market volatility, and examine both inter-industry and intra-industry spillover effects. Specifically, this study addresses three key questions: (1) Are firms that experience data breaches more likely to face future breaches? (2) How do abnormal stock returns and volatility respond to breach disclosures? (3) Do data breaches produce spillover effects across firms within the same industry or across different industries?

Hypothesis 1: Data breaches are autocorrelated.

If a firm had been attacked before, then hackers may already possess critical information that facilitates future breach incident. A subsequent incident, particularly a malicious data breach, is more likely to occur if the firm fails to address underlying cybersecurity weaknesses. In such cases, attackers or insiders have already identified system vulnerabilities and gained unauthorized access. Additionally, fraudsters may use the exposure records from past

[4] https://notified.idtheftcenter.org/s/

[5] https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/

breaches to execute various schemes. Even more concerning, stolen data can be bought and sold through a network of illicit, secretive, and easily accessible mechanisms (Holt et al., 2010).

Hypothesis 2: The stock market reaction to data breach announcements will be negative.

Anecdotal and empirical evidence suggests that firms incur both tangible and intangible costs following a cyber incident (Deloitte, 2018). Data breaches often lead to significant financial losses from system repairs, data destruction or deletion, business disruptions, theft of intellectual property, loss of customers, regulatory fines, litigation expenses, and post-breach remediation. In addition to these direct costs, breaches can severely damage a company's reputation and erode customer trust, potentially resulting in abnormal customer turnover, loss of pricing power,declining sales and profits, and higher future insurance premiums. Romanosky et al. (2014) examine how data breaches influence litigation risk and find that the number of compromised records is positively associated with the likelihood of being sued. Similarly, Kamiya et al. (2018) show that breached firms experience higher cash flow volatility, higher probabilities of credit rating downgrades, and an increased probability of bankruptcy. Overall, these adverse outcomes are expected to have a material impact on share prices, consistent with a range of traditional factor models, rational valuation theory, and arbitrage pricing theory.

Data breaches are a special type of bad news. Prior literature (Fostel and Geanakoplos, 2012) shows that bad news is associated with an increase in volatility. The flow of information is also related to the variance of returns (Ross, 1989). Therefore, the probable reason for an increase in volatility may be heightened investor uncertainty resulting from the direct consequences of reported data breach incidents. In addition, stakeholders are concerned about companies' levels of IT security and information privacy (Schmidt et al., 2016). There is also evidence indicating that investors perceive the occurrence of data breaches negatively (Spanos and Angelis, 2016). Data breaches are likely to place downward pressure on firms' future cash flows. Since the stock market reflects unexpected changes in firm cash flows, I expect to observe negative abnormal returns and increased stock volatility for firms and industries that report data breach incidents.

Extant literature has shown that investment returns and equity markets react to cyberattack events. However, the findings are mixed regarding the impact of data breaches on the stock market. Acquisti et al. (2006) show that data breaches have a negative and statistically significant effect on a company's market value on the announcement day. The cumulative impact increases in magnitude over the days following the announcement but then declines and loses statistical significance. Goel and Shawky (2009) find that the announcement of a corporate security breach reduces the firm's market value by about 1% in the days surrounding the event. In contrast, other studies find no statistically significant financial impact from data breach incidents (Campbell et al., 2003; Hovav et al., 2004; Ko and Dorantes, 2006). Chichernea et al. (2018) report that when a firm is successfully attacked, it experiences negative abnormal

returns once the news becomes public.

However, the impact of data breach events on stock volatility and the spillover effects into financial markets have not been well explored. Equity investors are particularly concerned with stock returns, trading volumes, and volatility. While volatility can signal increased investment risk, it also presents opportunities for substantial returns and plays a critical role in the operation of financial markets. Therefore, it is essential to examine events that may influence volatility. Although extensive literature has investigated the causes and consequences of volatility in the stock market, relatively little attention has been given to industry-level volatility. Attempts to understand the effect of data breaches on market volatility include a study by Tweneboah-Koduah et al. (2020), which identifies that data breach announcements can significantly influence equity volatility, based on an event study of 96 listed firms.

This study adds to the growing literature on cybersecurity risks. Extant literature This study contributes to the growing literature on cybersecurity risks. Existing research on data breach events primarily focuses on their direct effects on the breached firms. One attempt to examine spillover effects is Garg (2020), who study corporate cash holdings behavior. Their findings show that firms significantly increase cash holdings after experiencing a cybersecurity attack and that breaches have a spillover effect, peer firms also increase cash holdings in response to another firm's breach. Prior literature on cyberattacks also explores the determinants of such incidents. Sen et al. (2015) identify several factors that increase the risk of a data breach occurrence and find that greater investment in information technology (IT) security is associated with a higher risk of breach. Wheatley et al. (2016) report that both the frequency and severity of data breaches increase with firm size and vary significantly across sectors.

Cybersecurity threats are rapidly evolving, with new vulnerabilities being disclosed on a daily basis (Kuypers et al., 2016). As data breaches become increasingly widespread, this study examines their impact not only on the directly targeted firms but also on their industries and closely related peers. I use data from the Privacy Rights Clearinghouse (PRC) to identify firms that experienced data breaches between January 2015 and December 2018, focusing on six specific industries. The initial sample includes 8,474 breach incidents recorded during this period.

To address the first research question, I estimate a logistic regression with lagged breach indicator variables of the firm and the lagged breach indicator variables of peer firms in the same industry as the dependent variables regressed on the current data breach indicator variable. The results indicate that data breaches are autocorrelated, suggesting that cybercriminals often reuse compromised data rather than seeking entirely new targets. Consequently, firms that have previously experienced breaches have a higher likelihood of being breached again. Recent breaches increase the likelihood of future incidents within the same industry. This may be due to certain sectors or firms being more frequently targeted. Moreover, the analysis reveals both inter-industry and intra-industry spillover effects from data breaches. To better understand the nature of these breach incidents, I further group data breaches into two distinct

categories, malicious, where attackers actively target personal information, and non-malicious, which occur when private information is exposed accidentally. Preliminary findings suggest that malicious breaches behave differently than non-malicious breaches. Thus, I would further examine the difference between the two types of data breaches in the next section.

To address the second and third research question, I estimate fixed-effect regressions in which current stock volatility and abnormal returns are regressed on lagged firm-level breach indicators, lagged peer breach indicators, lagged volatility, and lagged abnormal returns. Overall, the results indicate that data breaches lead to an increase in volatility but have no significant effect on abnormal returns. Firm-level analysis suggests that contemporaneous data breaches significantly impact contemporaneous volatility. The stock volatility increases by 2.3% during the week of a firm's own breach and by 0.6% during the week of a peer firm's breach. These findings confirm that the stock market reaction to data breaches is generally negative in terms of volatility. Moreover, the analysis reveals a volatility spillover effect from data breaches, which is mainly driven by malicious breaches. Interestingly, peer firms' malicious breaches are associated with a decrease in a firm's volatility and an increase in its abnormal return, suggesting nuanced market perceptions based on breach characteristics. However, at the industry level, the presence of past breaches is not jointly significant in explaining variation in volatility or abnormal returns. This may be due in part to the PRC dataset including many private firms whose breach events do not highly influence publicly traded industry stock return. As a result, the impact of data breaches appears to be more pronounced at the firm level than at the industry level.

The rest of the paper is organized as follows. Section 2 describes the data and summary statistics. Section 3 explores the spillover effects on data breaches. Section 4 examines the impact of data breaches on stock abnormal returns and volatility. Section 5 concludes the paper.

3.2. Data and Basic Characteristics

3.2.1. Data on firm cybersecurity breaches

I construct the main dataset used in my analysis from the Center for Research in Securities Prices (CRSP) and the Privacy Rights Clearinghouse (PRC). Data breach information from 2005 to 2019 is obtained from the Chronology of Data Breaches provided by PRC, which collects data on public announcements made by breached firms in the U.S. For each breach, the dataset includes: the date the breach was made public, the name of the entity responsible for the data, the type of entity breached, a classification of the breach type, the total number of records breached, the location where the entity operates, information on the source of the data, and a brief description of the breach.

There are two limitations to the PRC data. First, the number of records listed for each breach is only an estimate of the total records exposed. Second, the dataset contains only those breaches that have been publicly acknowledged. It is possible that many breaches go unreported. Nevertheless, the PRC dataset remains the largest and most comprehensive public dataset of its kind, containing 9,015 data breaches and 10,387,398,893 total records exposed.

For public companies, I match organization names reported in the PRC database with firm names listed in Compustat and CRSP. For private companies, I match each record by organization name using ReferenceUSA to identify the primary SIC code. This study focuses on firms with identified industries. I exclude data breach incidents in 2019, as they pertain solely to the medical services sector. As a result, the final sample includes 8,747 data breaches with 10,386,084,705 records exposed, spanning six industries—banking and insurance, government entities, educational institutions, medical services, retail/merchants, and other businesses and nonprofit organizations—from January 2005 to December 2018.

For the firm-level analysis, this paper focuses only on publicly listed firms. Thus, I restrict the sample to data breaches for which stock return data are available in CRSP. The resulting firm-level sample comprises 393 publicly traded firms with 697 incidents, exposing 6,629,557,327 records from January 2005 to December 2018.

Data breaches take many different forms. Some result from accidental or non-malicious actions, such as improper sharing or the loss of devices. Others are caused by calculated, malicious acts intended to obtain information for profit. These types of attacks commonly target personally identifiable information, such as customers' online accounts, credit card details, or trade secrets. Malicious breaches may cause significant monetary losses.

I categorize data breach incidents into two groups: malicious breaches and non-malicious breaches. Malicious breaches include hacking or malware, credit card fraud, and insider-related information breaches. The remaining types are labeled as non-malicious breaches, where records are exposed accidentally—for example, through the loss of devices or unintended public disclosure of sensitive information. I exclude breaches with an unknown classification.

Table 3.1 reports the descriptive statistics for the data breach dataset. Panel A shows that most data breach incidents are caused by hacking or malware. There are 3,175 malicious data breaches and 5,575 non-malicious data breaches. Panel B presents summary statistics on the number of data breach events and the total number of exposed records by sector. More data breach incidents occurred in 2012 and 2013 than in other years. The number of exposed records has increased dramatically over the past decade. The medical and business sectors have consistently been the most affected by data breaches. The MED (medical service) category is the most affected, likely because healthcare companies store more sensitive and valuable information than firms in other industries. The BSO (other business) category is the second most affected, as this sector is often targeted due to the potentially high financial rewards

from a successful attack. The NGO (nonprofit organization) category reports the fewest data breach incidents among all categories. Therefore, I create a new category, BSONGO, by combining the BSO and NGO sectors. Figure 3.1 shows the time pattern of weekly total data breach events. The number of breaches has declined in recent years, likely because companies are paying more attention to cybersecurity issues.

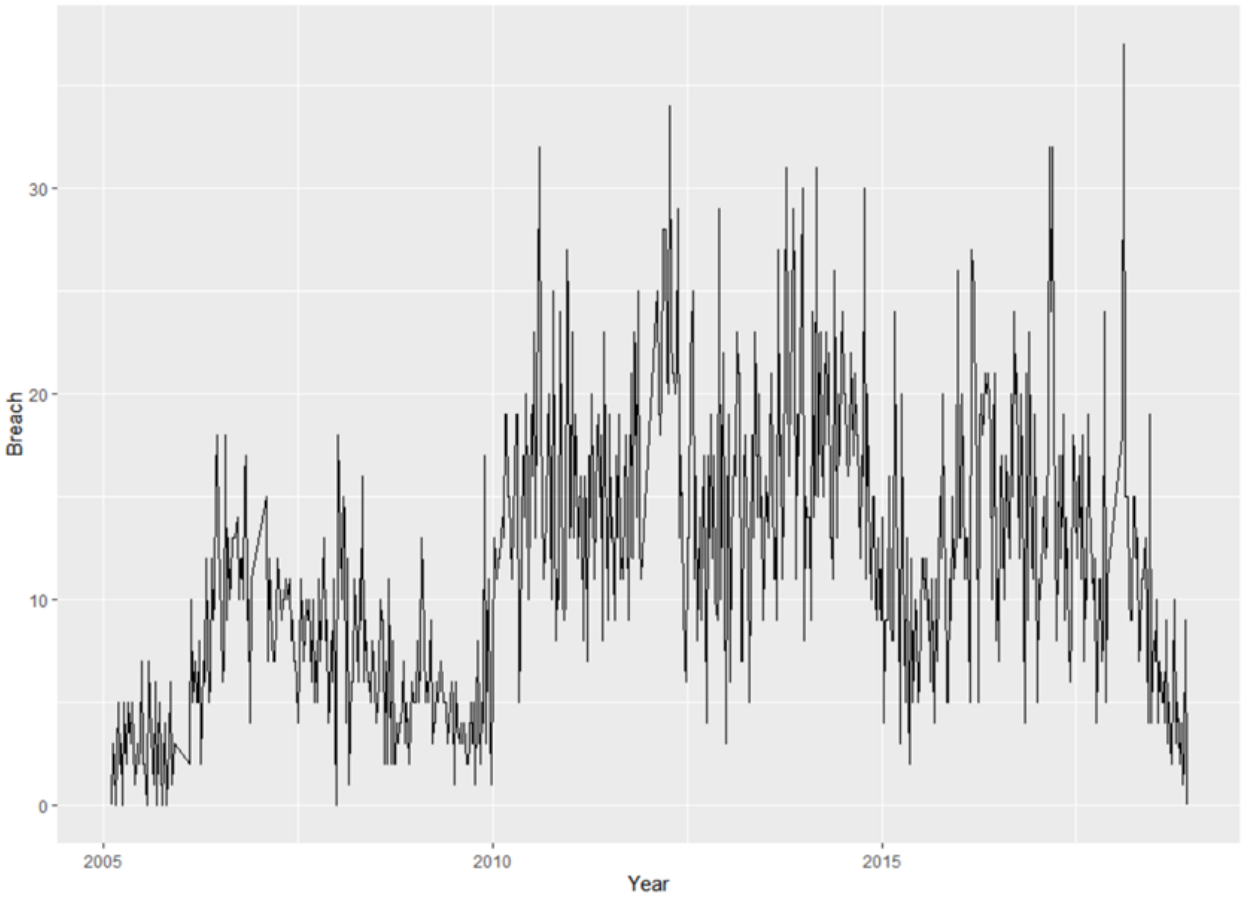

Figure 3.1: Weekly breaches from January 2005 to December 2018.

3.2.2. Data on stock return

The daily U.S. stock returns data from 2004 to 2019 come from the Center for Research in Security Prices (CRSP). I closely follow the Kenneth R. French industry portfolio definitions and match SIC codes to the six sectors defined by the Privacy Rights Clearinghouse (PRC). Table A1 presents the industry classifications based on SIC codes.

Table 3.1: Descriptive Statistics

Panel A. Type of data breach	
Type of data breach	Number of data breach events
Hacking or Malware	2,501
Credit card fraud	68
Insider breaches information	606
Physical loss of information	1,721
Loss of portable devices	1,172
Loss of stationary devices	249
Unintended Disclosure Not Involving Hacking	1,833
Unknown	235
Total	**8,474**

Panel B. Number of breaches by industry									
Year	BSF	BSO	BSR	EDU	GOV	MED	NGO	TOTAL	Total records
2005	22	12	10	64	15	11	2	136	55,101,241
2006	106	61	31	102	115	58	9	482	68,580,749
2007	76	62	48	107	89	60	13	456	149,957,921
2008	38	50	27	104	69	56	11	355	130,896,900
2009	25	36	12	72	53	65	8	271	251,576,460
2010	98	47	87	75	103	385	11	806	140,980,938
2011	50	94	87	68	86	401	16	797	447,913,238
2012	72	87	107	86	86	438	20	897	298,819,016
2013	62	87	86	46	56	554	7	903	158,898,528
2014	43	84	45	38	28	632	9	872	318,379,341
2015	43	98	11	39	22	352	2	554	319,440,416
2016	36	155	5	21	27	577	3	832	4,626,238,665
2017	73	160	33	29	22	446	6	874	2,051,973,997
2018	43	12	34	28	7	236	2	708	1,367,327,295
Total	**787**	**1,045**	**623**	**848**	**781**	**4,271**	**119**	**8,943**	**10,386,084,705**

This table presents descriptive statistics for the data breach dataset. The PRC data categorizes organizations into eight groups: BSF (Businesses—Financial and Insurance Services), BSO (Businesses—Other), BSR (Businesses—Retail/Merchant, including Online Retail), EDU (Educational Institutions), GOV (Government and Military), MED (Healthcare, Medical Providers, and Medical Insurance Services), NGO (Nonprofits), and UNKN (Unknown).

Constructing the measures of volatility

The method for calculating the idiosyncratic volatility of a firm's return follows the approach of Campbell et al. (2001). Industries are denoted by the subscript i, and individual firms are indexed by j. The simple excess return of firm j in industry i during period t is denoted as R_{jit}. This excess return is measured relative to the Treasury bill rate. The weight of firm j in industry i is denoted as w_{jit}, with weights based on market capitalization. The excess return of industry i in period t is given by

$$R_{it} = \sum_{j \in i} w_{jit} R_{jit}.$$

Industries are aggregated correspondingly. The weight of industry i in the total market is denoted by w_{it}, and the excess market return is

$$R_{mt} = \sum_{i} w_{it} R_{it}.$$

Consider the following simplified industry return decomposition that drops the industry beta coefficients β_{im}, then the decomposition based on the CAPM:

$$R_{it} = R_{mt} + \epsilon_{it}$$

where ϵ_{it} is the difference between the industry return R_{it} and the market return R_{mt}.

Let s denote the interval at which returns are measured. I use daily returns for most estimates but also consider weekly and monthly returns to check the sensitivity of our results with respect to the return interval. Using returns of interval s, I construct volatility estimates at intervals t.

For volatility in industry i, I sum the squares of the industry-specific residual in the above equation within a period t:

$$\text{Vol}_{it} = \sum_{s \in t} \epsilon_{is}^2.$$

Constructing the measures of abnormal return

The abnormal return (AR) on a given day within the event window is computed as the difference between the actual return and the estimated return from the model. The estimation window spans from day 170 to day 30 before the data breach disclosure to avoid parameter contamination by the event under study.

To estimate abnormal stock return, a market return model is used with an estimation window of 180 daily returns of each stock:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}$$

where R_{mt} is the return of a market portfolio for period t and ϵ_{it} is the normally distributed residual.

Using the beta and alpha estimates from the above equation, the expected stock return is measured by:

$$\mathbb{E}(R_{it}) = \hat{\alpha}_i + \hat{\beta}_i R_{mt}$$

For industry i and period t, the abnormal return is:

$$\mathrm{AR}_{it} = R_{it} - \mathbb{E}(R_{it})$$

Figure 3.2 shows daily stock market volatility during the sample period. The highest levels of volatility occur during the financial crisis. Figure 3.3 presents daily stock volatility by sector. The "Other Business" and "Medical Services" sectors exhibit higher volatility than other industries. As expected, the "Government" sector has the lowest volatility. Table 3.2 reports the descriptive statistics of daily stock returns and volatility across different sectors.

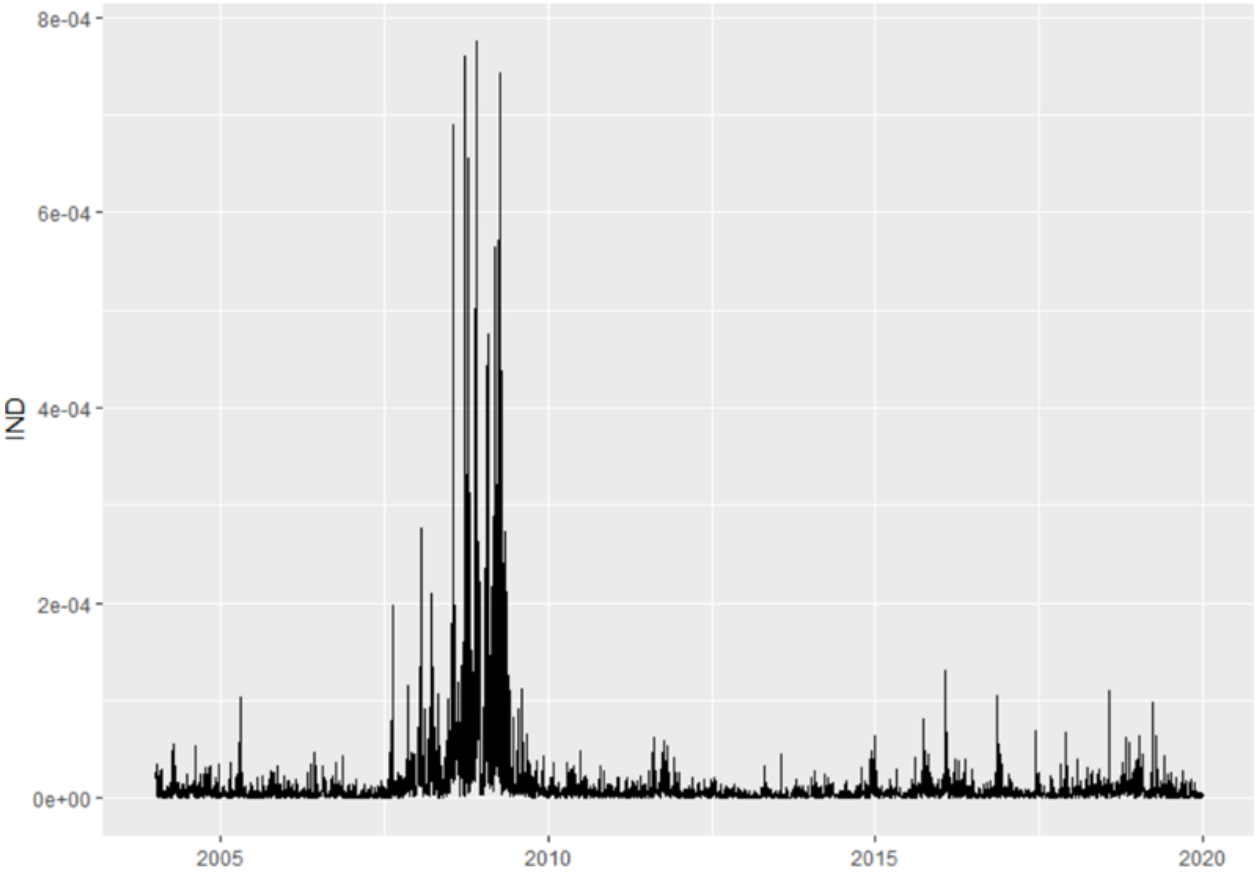

Figure 3.2: Daily volatility of the value-weighted market volatility for the period 2004 through 2019.

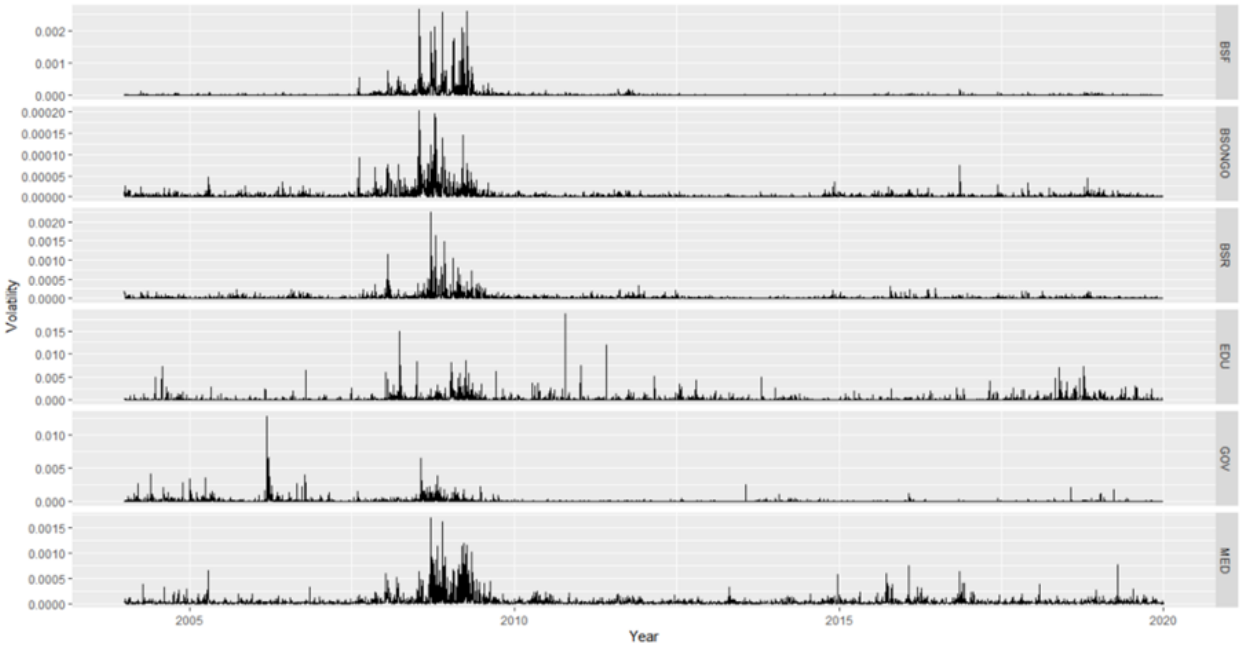

Figure 3.3: Daily value-weighted industry volatility

Table 3.2: Industry-level Summary Statistics for Volatility and Abnormal Return

Industry	Weight	Volatility		Abnormal return	
		Mean	Std	Mean	Std
Banking and insurance	24.72%	0.000141	0.000464	-0.000570	0.010690
Other business	57.36%	0.000019	0.000038	0.000042	0.004445
Retail/merchant	7.10%	0.000135	0.000238	0.000341	0.010189
Educational institution	0.13%	0.001196	0.001927	0.000374	0.034104
Governmental entity	1.69%	0.000507	0.001158	0.002120	0.022532
Medical service	8.90%	0.000181	0.000304	0.000605	0.011606

This table presents the descriptive statistics for stock returns of each industry by weight. All volatilities are measured weekly. Weight is computed as the ratio of the average market value of industry to the total market value of all industries.

3.3. Regression analysis

3.3.1. Data breaches autocorrelation and spillover effects

In this section, I examine whether data breaches are autocorrelated. If a firm has been attacked before, hackers may already possess the information they need. A similar incident might occur in the future due to poor cybersecurity risk management. Alternatively, fraudsters might use exposed records from previous breaches to carry out a variety of schemes.

To capture the serial dependence and spillover effects of data breach events, I employ the following fixed-effects logistic regression model for the industry-level analysis.

$$P(Breach_{i,t} = 1 \mid \mathbf{x}_i) = \frac{\exp(\tau_i + \mu_t + \mathbf{x}_i'\beta)}{1 + \exp(\tau_i + \mu_t + \mathbf{x}_i'\beta)}$$

Where β' represents the vector of regression coefficients, and $\mathbf{x}_i' = (Breach_{i,t-1}, Breach_{i,t-2}, Breach_{-i,t}, Breach_{-i,t-1}, Breach_{-i,t-2})$ is the covariate vector representing the lagged and contemporaneous data breach indicator. $Breach_{i,t}$ is the dummy variable indicating the existence of data breaches of industry i in period t, $Breach_{-i,t-1}$ is the dummy variable indicating the existence of data breaches of other industry in period $t-1$, τ_i is the industry fixed effect, μ_t is the time fixed effect.

For firm-level analysis, I use the following fixed effects logistic regression.

$$P(Breach_{i:j,t} = 1 \mid \mathbf{x}_i) = \frac{\exp(\tau_i + \mu_t + \mathbf{x}_i'\beta)}{1 + \exp(\tau_i + \mu_t + \mathbf{x}_i'\beta)}$$

Where $\mathbf{x}_i' = (Breach_{i:j,t-1}, Breach_{i:j,t-2}, Breach_{i:-j,t}, Breach_{i:-j,t-1}, Breach_{i:-j,t-2})$ is the covariate vector representing the lagged and contemporaneous firm data breach indicators. Where $Breach_{j,t}$ is the dummy variable indicating the existence of data breaches of firm j that belongs to industry i in period t, $Breach_{i:-j,t-1}$ is the dummy variable indicating the existence of data breaches of peer firms in industry i in period $t-1$, τ_j is the firm fixed effect, μ_t is the time fixed effect.

Malicious breaches are the most common form of data breach, and the underlying causes of malicious breaches differ from those of non-malicious breaches. Malicious breaches result from deliberate actions targeting private information, whereas non-malicious breaches typically arise from negligence, where sensitive data are unintentionally exposed rather than actively pursued by an attacker. Malicious breaches may occur when hackers or insiders exploit system vulnerabilities or inject malicious code to gain unauthorized access. In contrast, non-malicious breaches are more difficult to predict, often stemming from accidents or procedural failures. Malicious breach risk is also closely tied to

the evolution and sophistication of technology. Firms with weak cybersecurity infrastructure are particularly vulnerable. These distinctions motivate a closer examination of the differences between malicious and non-malicious breaches.

As shown in the regression results in Table 3.3, firms that have experienced a prior data breach are significantly more likely to suffer another breach in the future. This effect is especially strong for non-malicious breaches. Additionally, there is evidence of spillover effects within industries. When firms in the same industry deploy similar cybersecurity solutions, it is plausible that a breach at one firm could increase vulnerability at others, particularly to malicious attacks. The analysis shows that peer firms are found to experience fewer breaches immediately after a breach and more breaches two weeks after a data breach, suggesting a short-term alertness effect followed by delayed exposure. Spillover effects are more pronounced for malicious breaches than for non-malicious ones. Incidents like device loss or unintentional exposure, common in non-malicious breaches, are often random. For example, an employee might leave a laptop on public transportation, potentially granting unauthorized access to sensitive information if the device is misused.

At the industry level, the analysis reveals that if one industry experiences a breach, it faces an elevated risk of additional breaches in the following two weeks. Panel B further indicates a significant intra-industry spillover effect immediately following a breach. If a breach occurs in a different industry, the focal industry faces a higher probability of contemporaneous breaches but a lower probability of future data breaches.

Overall, the results from both firm-level and industry-level analyses support the hypotheses that data breaches are autocorrelated and exhibit significant inter-industry and intra-industry spillover effects.

Table 3.3: Fixed effects logistic regression

Panel A [Firm level]					
Variable	$Breach_{i:j,t}$	**Variable**	$Breach^{M}_{i:j,t}$	**Variable**	$Breach^{NM}_{i:j,t}$
$Breach_{i:j,t-1}$	1.315***	$Breach^{M}_{i:j,t-1}$	0.943	$Breach^{NM}_{i:j,t-1}$	1.833***
	(3.538)		(1.231)		(3.536)
$Breach_{i:j,t-2}$	-0.984	$Breach^{M}_{i:j,t-2}$	-11.280	$Breach^{NM}_{i:j,t-2}$	-11.422
	(-0.967)		(-0.034)		(-0.045)
$Breach_{i:-j,t}$	-1.414***	$Breach^{M}_{i:-j,t}$	-0.838***	$Breach^{NM}_{i:-j,t}$	-1.557***
	(-9.149)		(-4.068)		(-7.552)
$Breach_{i:-j,t-1}$	0.452***	$Breach^{M}_{i:-j,t-1}$	0.675***	$Breach^{NM}_{i:-j,t-1}$	0.359*
	(2.973)		(3.162)		(1.804)
$Breach_{i:-j,t-2}$	0.283*	$Breach^{M}_{i:-j,t-2}$	0.242	$Breach^{NM}_{i:-j,t-2}$	0.202
	(1.856)		(1.114)		(1.003)
Time f.e.	Y		Y		Y
Firm f.e.	Y		Y		Y

Panel B [Industry level]					
Variable	$Breach_{i,t}$	**Variable**	$Breach^{M}_{i,t}$	**Variable**	$Breach^{NM}_{i,t}$
$Breach_{i,t-1}$	0.960***	$Breach^{M}_{i,t-1}$	0.812***	$Breach^{NM}_{i,t-1}$	0.889***
	(12.816)		(10.930)		(11.462)
$Breach_{i,t-2}$	0.787***	$Breach^{M}_{i,t-2}$	0.701***	$Breach^{NM}_{i,t-2}$	0.718***
	(10.457)		(9.410)		(9.234)
$Breach_{i,t}$	1.000***	$Breach^{M}_{i,t}$	0.285**	$Breach^{NM}_{i,t}$	0.643***
	(3.925)		(2.456)		(3.834)
$Breach_{-i,t-1}$	-0.067	$Breach^{M}_{-i,t-1}$	0.157	$Breach^{NM}_{-i,t-1}$	-0.182
	(-0.261)		(1.409)		(-1.088)
$Breach_{-i,t-2}$	-0.577**	$Breach^{M}_{-i,t-2}$	0.169	$Breach^{NM}_{-i,t-2}$	-0.149
	(-2.037)		(1.483)		(-0.895)
Industry f.e.	Y		Y		Y

This table presents the fixed effects logistic regression result for both industry level and firm level. Values of z-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

3.3.2. The effect of data breaches on stock abnormal return and volatility

Assessing the change in volatility resulting from data breaches remains an open issue. This paper empirically examines the effects of data breaches on stock abnormal returns and volatility, as well as the potential contagion effects of such incidents. As specified in my hypotheses, if a firm or industry experiences a data breach, it is expected to lead to a negative abnormal return and an increase in stock volatility. In this section, I present both firm-level and industry-level empirical analyses of data breaches.

Firm level

Volatility To investigate how stock volatility reacts to the data breach disclosure, I estimate the following fixed effect regression:

$$\begin{aligned} Vol_{i,j,t} = \tau_j + \mu_t + \beta_{j1} Breach_{i,j,t-1} + \beta_{j2} Breach_{i,j,t-2} + \beta_{j3} Breach_{i,j,t} \\ + \beta_{j4} Breach_{i,-j,t-1} + \beta_{j5} Breach_{i,-j,t-2} + \eta_{j1} Vol_{i,j,t-1} + \eta_{j2} Vol_{i,j,t-2} \\ + \gamma_{i1} AR_{i,j,t-1} + \gamma_{i2} AR_{i,j,t-2} + \delta_j X_{i,j,t} + \varepsilon_{i,j,t} \end{aligned}$$

Where $Vol_{i,j,t}$ is the weekly volatility of firm j that belongs to industry i in period t, $Breach_{i,j,t}$ is the dummy variable indicating the existence of data breaches of firm j that belongs to industry i in period t, $Breach_{i:-j,t}$ is the dummy variable indicating the existence of data breaches of peer firms that belong to industry i in period t. $Vol_{i,j,t-1:t-2}$ is weekly lagged volatility of firm j that belongs to industry i for week $t-1:t-2$, the vector X is a set of controls, $\varepsilon_{i,j,t}$ is a vector of error terms. $AR_{i,j,t-1:t-2}$ is the lagged abnormal return of firm j that belongs to industry i for week $t-1:t-2$. τ_j is the firm fixed effect, μ_t is time fixed effect. This model includes controls for firm characteristics (the size of the firm measured as the natural logarithm of market capitalization) and data breach characteristics (the total number of records exposed in data breaches).

Abnormal return

$$\begin{aligned} AR_{i,j,t} = \alpha_j + \mu_t + \beta_{j1} Breach_{i,j,t-1} + \beta_{j2} Breach_{i,j,t-2} + \beta_{j3} Breach_{i,j,t} + \beta_{j4} Breach_{i:-j,t-1} \\ + \beta_{j5} Breach_{i:-j,t-2} + \eta_{j1} Vol_{i,j,t-1} + \eta_{j2} Vol_{i,j,t-2} + \gamma_{i1} AR_{i,j,t-1} + \gamma_{i2} AR_{i,j,t-2} + \delta_j X_{i,j,t} + \varepsilon_{i,j,t} \end{aligned}$$

Where $AR_{i,j,t}$ is the abnormal return of firm j that belongs to industry i in period t.

Table 3.4 displays the results of these estimations. I begin the analysis with two contemporaneous data breach indicator variables, as shown in Column 1. The contemporaneous breach indicators are jointly significant. Stock volatility

increases by 2.3% immediately after a data breach incident and by 0.6% following a peer firm's breach. A probable reason for this increase in volatility is heightened investor uncertainty stemming from the direct consequences of reported data breaches, such as loss of customer trust, system repair costs, trading disruptions, business losses, and increased litigation risk. To further examine the duration of this effect, I include both contemporaneous and lagged firm data breach indicators. Column 2 shows that a firm's own breach indicators are jointly insignificant, whereas all peer firm breach indicators are jointly significant in predicting stock volatility, based on the F-test. This finding confirms the existence of an inter-firm spillover effect on volatility that persists for two periods. Finally, Columns 3 and 4 incorporate lagged volatility and abnormal returns. The results indicate that stock volatility is autocorrelated.

For abnormal returns, Column 5 presents results using two contemporaneous data breach indicator variables. Only the peer firm data breach indicator is marginally significant. Abnormal returns decrease by 0.2% following a firm's own data breach and increase by 0.3% if a peer firm experiences a breach. To further examine the duration of this effect, I include all contemporaneous and lagged data breach indicators. Column 6 shows the results including all lagged firm-level and peer-level breach indicators. The firm's own breach indicators are jointly insignificant in explaining abnormal returns. However, the peer firms' breach indicators are jointly significant in predicting abnormal returns, based on the F-test. Specifically, abnormal returns increase in the contemporaneous period and decrease over the following two-week period. A possible explanation for this pattern is a shift in investor focus regarding the informational content of breach announcements, which may have driven the observed market reactions during and after the event period. This suggests that the stock market's response to data breach announcements—in terms of abnormal returns—differs between the contemporaneous and post-event periods. Finally, I include lagged volatility and abnormal returns. Columns 7 and 8 indicate that abnormal returns are negatively related to both lagged volatility and lagged abnormal returns.

The findings are consistent with the hypothesis that a data breach announcement increases stock volatility, although this effect is only evident during the week of the event. Additionally, a data breach announcement marginally decreases abnormal returns two weeks after the incident. There are also inter-industry spillover effects on both volatility and abnormal returns when a peer firm in the same industry is attacked. Interestingly, a firm's own data breach announcements do not have significant predictive power in explaining its abnormal returns, whereas peer firms' breach announcements do. Specifically, a firm's abnormal return increases following a peer firm's data breach within the same industry. Taken together, these results suggest that data breach incidents negatively affect the financial performance of the breached firm. To identify which type of breach primarily drives this effect, I next examine the impact of malicious data breaches on stock returns.

Table 3.4: Market Reactions to Data Breaches (firm-level)

VARIABLES	Volatility				Abnormal return			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Vol_{i,j,t}$	$Vol_{i,j,t}$	$Vol_{i,j,t}$	$Vol_{i,j,t}$	$AR_{i,j,t}$	$AR_{i,j,t}$	$AR_{i,j,t}$	$AR_{i,j,t}$
$Breach_{i,j,t}$	0.023**	0.023**	0.020*	0.020*	-0.002	-0.001	-0.002	-0.001
	(0.012)	(0.012)	(0.011)	(0.011)	(0.003)	(0.002)	(0.002)	(0.002)
$Breach_{i,j,t-1}$		0.005	-0.001	-0.001		0.000	0.000	0.000
		(0.009)	(0.009)	(0.009)		(0.002)	(0.002)	(0.002)
$Breach_{i,j,t-2}$		0.007	0.000	0.002		-0.004**	-0.004**	-0.004**
		(0.009)	(0.009)	(0.009)		(0.002)	(0.002)	(0.002)
$Breach_{i,-j,t}$	0.006***	0.004***		0.002*	0.003***	0.003***		0.003***
	(0.001)	(0.001)		(0.001)	(0.000)	(0.000)		(0.000)
$Breach_{i,-j,t-1}$		0.008***		0.006***				-0.005***
		(0.001)		(0.001)				(0.000)
$Breach_{i,-j,t-2}$		0.014***		0.014***				-0.001***
		(0.001)		(0.001)				(0.000)
$Vol_{i,j,t-1}$			0.209***	0.210***			-0.005***	-0.005***
			(0.002)	(0.002)			(0.000)	(0.000)
$Vol_{i,j,t-2}$			0.107***	0.107***			-0.004***	-0.004***
			(0.002)	(0.002)			(0.000)	(0.000)
$AR_{i,j,t-1}$			0.190***	0.184***			-0.060***	-0.059***
			(0.008)	(0.008)			(0.002)	(0.002)
$AR_{i,j,t-2}$			-0.017**	-0.019**			-0.017***	-0.018***
			(0.008)	(0.008)			(0.002)	(0.002)
Time f.e.	Y	Y	Y	Y	Y	Y	Y	Y
Firm f.e.	Y	Y	Y	Y	Y	Y	Y	Y
Adj. R squared	99.97%	99.97%	99.97%	99.97%	28.16%	28.25%	29.10%	29.21%

Standard errors are given in parentheses. Time and firm fixed effects and intercept are not tabulated. Independent variables are lagged by a week or two weeks. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level, respectively.

Malicious and non-malicious breaches I find that the existence of past data breaches explains volatility but not abnormal returns, which is consistent with the results from the full data breach sample. Given that malicious breaches exhibit distinct characteristics, I further examine their predictive impact on stock abnormal returns and volatility.

Regression results are presented in Table 3.5. Column 1 shows strong evidence that both malicious and non-malicious firm-level data breaches influence stock volatility during the week of the event. Stock volatility increases by 5.1% immediately after a malicious breach and by 4.3% immediately after a non-malicious breach. These findings are consistent with the results from the full sample of firm-level data breaches. Column 3 shows that neither malicious nor non-malicious data breach announcements have predictive power in explaining abnormal returns. However, peer firms' malicious data breach announcements do. Specifically, a firm's abnormal return increases after a peer firm in the same industry experiences a malicious breach. Columns 2 and 4 indicate significant spillover effects of both malicious and non-malicious data breaches on stock abnormal returns and volatility. When a peer firm experiences a malicious breach, the focal firm's stock volatility decreases by 0.9% and its abnormal return increases by 0.6% immediately. In contrast, when a peer firm experiences a non-malicious breach, the focal firm's stock volatility increases by 0.8%, while the change in abnormal return is negligible. These results suggest that malicious and non-malicious breaches have distinct effects on market behavior.

Overall, the findings of this section suggest that a peer firm's data breach may benefit an individual firm, and this effect is primarily driven by malicious data breaches. Malicious breaches involve the intentional release of confidential information for profit. The significant monetary losses caused by such breaches may lead to inter-industry spillover effects on volatility.

Table 3.5: Market Reactions to Malicious and Non-malicious Data Breaches (firm-level)

VARIABLES	Volatility		Abnormal return	
	(1)	(2)	(3)	(4)
	$Vol_{i,j,t}$	$Vol_{i,j,t}$	$AR_{i,j,t}$	$AR_{i,j,t}$
$Breach^{M}_{i,j,t}$	0.051***	0.045***	-0.001	-0.001
	(0.015)	(0.015)	(0.003)	(0.003)
$Breach^{NM}_{i,j,t}$	0.043***	0.043***	-0.003	-0.003
	(0.015)	(0.015)	(0.003)	(0.003)
$Breach^{M}_{i,-j,t}$	-0.009***	-0.009***	0.006***	0.006***
	(0.001)	(0.001)	(0.000)	(0.000)
$Breach^{NM}_{i,-j,t}$	0.008***	0.007***	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
$Vol_{i,j,t-1}$		0.210***		-0.005***
		(0.002)		(0.000)
$Vol_{i,j,t-2}$		0.107***		-0.004***
		(0.002)		(0.000)
$AR_{i,j,t-1}$		0.189***		-0.059***
		(0.008)		(0.002)
$AR_{i,j,t-2}$		-0.017**		-0.017***
		(0.008)		(0.002)
Time f.e.	Y	Y	Y	Y
Firm f.e.	Y	Y	Y	Y
Adj. R squared	99.97%	99.97%	28.25%	29.21%

This table reports the regression results distinguishing between malicious and non-malicious data breaches. Standard errors are in parentheses. Time and firm fixed effects are included but not reported. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level, respectively.

Industry level

The PRC dataset includes only those breaches that have been publicly reported. It is possible that many data breaches go unreported. Therefore, I further examine the effects of data breaches on stock abnormal returns and volatility, as well as the potential contagion effects of these incidents at the industry level.

Volatility and abnormal return To investigate how industry volatility reacts to the data breach disclosure, I estimate the following regression equation:

$$\begin{aligned} Vol_{i,t} = \beta_0 &+ \beta_1 Breach_{i,t} + \beta_2 Breach_{i,t-1} + \beta_3 Breach_{i,t-2} + \beta_4 Breach_{-i,t-1} \\ &+ \beta_5 Breach_{-i,t-2} + \eta_1 Vol_{i,t-1} + \eta_2 Vol_{i,t-2} + \gamma_1 AR_{i,t-1} + \gamma_2 AR_{i,t-2} \\ &+ \delta' X_{i,t} + \tau_i + \mu_t + \varepsilon_{i,t} \end{aligned} \tag{3.1}$$

Where Vol_{it} is the weekly volatility of industry i in period t, $Breach_{it}$ is a dummy variable indicating the existence of data breaches of industry i in period t, $Breach_{-i,t}$ is a dummy variable indicating the existence of data breaches of other industry in period t, $Vol_{i,t-1:t-2}$ is weekly lagged volatility of industry i for week $t-1:t-2$, the vector X is a set of controls, ε_{it} is a vector of error terms. $AR_{i,t-1:t-2}$ is the lagged abnormal return of industry i for period $t-1:t-2$. α_i is the industry fixed effect, μ_t is the time fixed effect. This model includes controls for industry characteristics (the size of the industry) and data breach characteristics (the total number of records exposed in data breaches).

To investigate how abnormal return reacts to the data breach disclosure, I use the following fixed effect regression:

$$\begin{aligned} AR_{i,t} = \alpha_i + \mu_t &+ \beta_1 Breach_{i,t} + \beta_2 Breach_{i,t-1} + \beta_3 Breach_{i,t-2} + \beta_4 Breach_{-i,t-1} \\ &+ \beta_5 Breach_{-i,t-2} + \eta_1 Vol_{i,t-1} + \eta_2 Vol_{i,t-2} + \gamma_1 AR_{i,t-1} + \gamma_2 AR_{i,t-2} \\ &+ \delta_i X_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{3.2}$$

Where $AR_{i,t}$ is the abnormal return for period t in industry i.

Table 3.6 reports the estimation results. Columns 1 and 5 show that all contemporaneous data breach indicators are insignificant in explaining both volatility and abnormal return. Columns 2 and 6 show results with all data breach indicators and lagged data breach indicators of other industry on the market. The existence of data breaches is not significant in explaining both volatility and abnormal return because all lagged data breach indicators are jointly insignificant. The industry volatility marginal increases 0.008% after 2 weeks of a data breach incident. In additional, there is no evidence of the intra-industry spillover effect on volatility and abnormal return. Volatility is autocorrelated, which is consistent with the firm-level analysis. The abnormal return is closely related to the lagged volatility but not lagged abnormal return.

To further examine the effect of data breaches on stock volatility and abnormal return across industries, I extend the baseline model by including interaction terms between breach indicators and industry categories. The regression results, reported in Table C.1 in the Appendix, suggest that while breaches are common in some sectors, their average market impact is limited. I next explore whether distinguishing between malicious and non-malicious breaches reveals more pronounced effects.

Table 3.6: Market Reactions to Data Breaches (Industry-level)

VARIABLES	Volatility				Abnormal return			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Vol_{i,t}$	$Vol_{i,t}$	$Vol_{i,t}$	$Vol_{i,t}$	$AR_{i,t}$	$AR_{i,t}$	$AR_{i,t}$	$AR_{i,t}$
$Breach_{i,t}$	0.00004	0.00002	0.00001	0.00002	0.00012	0.00020	0.00021	0.00013
	(0.00004)	(0.00005)	(0.00004)	(0.00005)	(0.00087)	(0.00085)	(0.00084)	(0.00086)
$Breach_{i,t-1}$		0.00006	0.00003	0.00004		0.00043	0.00054	0.00055
		(0.00004)	(0.00004)	(0.00004)		(0.00081)	(0.00080)	(0.00081)
$Breach_{i,t-2}$		0.00008**	0.00006	0.00007*		-0.00042	-0.00037	-0.00042
		(0.00004)	(0.00004)	(0.00004)		(0.00080)	(0.00079)	(0.00080)
$Breach_{-i,t}$	0.00016	0.00005		-0.00003	-0.00191	-0.00179		-0.00163
	(0.00015)	(0.00016)		(0.00016)	(0.00362)	(0.00337)		(0.00339)
$Breach_{-i,t-1}$		0.00026		0.00021		0.00020		0.00053
		(0.00016)		(0.00016)		(0.00331)		(0.00331)
$Breach_{-i,t-2}$				0.00021		-0.00097		-0.00083
				(0.00016)		(0.00333)		(0.00334)
$Vol_{i,t-1}$			0.18120***	0.18020***			-0.33297	-0.32368
			(0.01879)	(0.01880)			(0.40054)	(0.40104)
$Vol_{i,t-2}$			0.14970***	0.14870***			-0.87808**	-0.87756**
			(0.01761)	(0.01762)			(0.37547)	(0.37583)
$AR_{i,t-1}$			-0.00176**	-0.00175**			0.01978	0.01938
			(0.00084)	(0.00084)			(0.01795)	(0.01798)
$AR_{i,t-2}$			-0.00057	-0.00050			-0.01152	-0.01152
			(0.00083)	(0.00083)			(0.01775)	(0.01778)
Time f.e.	Y	Y	Y	Y	Y	Y	Y	Y
Industry f.e.	Y	Y	Y	Y	Y	Y	Y	Y
Adj. R squared	25.38%	25.54%	30.40%	30.43%	0.00%	0.00%	0.06%	0.00%

This table presents regression results examining industry-level market reactions to data breaches. Volatility and abnormal return are the dependent variables. Standard errors are given in parentheses. Industry and time fixed effects and intercept are not tabulated. Independent variables are lagged by a week or two weeks. ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Malicious and Non-Malicious breaches Since the existence of past data breaches is not significant in explaining both industry abnormal return and volatility, I further examine the predictive performance of malicious and non-malicious data breaches on abnormal return and stock volatility. Regression results are presented in Table 3.7. Panel A presents the effect of malicious breaches on abnormal return and volatility. It indicates that both contemporaneous and past data breaches have no predictive power on both abnormal return and volatility. But there is a marginally significant intra-industry spillover effect on volatility after two weeks of a malicious breach. Panel B indicates that the industry data breaches have a significant impact on the prediction ability for stock return volatility on the week of the event.

Overall, the existence of a past malicious and non-malicious data breach is not significant in explaining industry abnormal return and volatility. Both malicious and non-malicious data breaches have no intra-industry spillover effect on abnormal return and volatility. This finding is consistent with the full sample of industry data breaches.

Table 3.7: Market Reactions to Malicious and Non-malicious Data Breaches (Industry-level)

Panel A: Malicious breaches					**Panel B: Non-Malicious breaches**				
	Volatility		**Abnormal return**			**Volatility**		**Abnormal return**	
Variables	(1)	(2)	(3)	(4)	**Variables**	(1)	(2)	(3)	(4)
	$Vol_{i,t}$	$Vol_{i,t}$	$AR_{i,t}$	$AR_{i,t}$		$Vol_{i,t}$	$Vol_{i,t}$	$AR_{i,t}$	$AR_{i,t}$
$Breach^{M}_{i,t}$	0.00006	0.00005	-0.00015	0.00015	$Breach^{NM}_{i,t}$	0.00007**	-0.00003	0.00008	0.00011
	(0.00004)	(0.00004)	(0.00083)	(0.00085)		(0.00004)	(0.00004)	(0.00081)	(0.00086)
$Breach^{M}_{i,t-1}$		0.00006		-0.00013	$Breach^{NM}_{i,t-1}$		0.00001		-0.00024
		(0.00004)		(0.00083)			(0.00004)		(0.00083)
$Breach^{M}_{i,t-2}$		0.00004		-0.00088	$Breach^{NM}_{i,t-2}$		0.00009**		0.00035
		(0.00004)		(0.00083)			(0.00004)		(0.00083)
$Breach^{M}_{-i,t}$	0.00006	0.00004	0.00113	0.00133	$Breach^{NM}_{-i,t}$	-0.00003	0.00003	-0.00163	-0.00115
	(0.00009)	(0.00009)	(0.00190)	(0.00193)		(0.00010)	(0.00010)	(0.00201)	(0.00211)
$Breach^{M}_{-i,t-1}$		0.00016*		0.00020	$Breach^{NM}_{-i,t-1}$		0.00007		0.00157
		(0.00009)		(0.00193)			(0.00010)		(0.00210)
$Breach^{M}_{-i,t-2}$		0.00026***		-0.00205	$Breach^{NM}_{-i,t-2}$		-0.00003		-0.00232
		(0.00009)		(0.00193)			(0.00010)		(0.00210)
Time f.e.	Y	Y	Y	Y	Time f.e.	Y	Y	Y	Y
Industry f.e.	Y	Y	Y	Y	Industry f.e.	Y	Y	Y	Y
Adj. R^2	25.40%	25.57%	0.00%	0.00%	Adj. R^2	25.44%	25.38%	0.00%	0.00%
N	1582	1582	1582	1582	N	1811	1811	1811	1811

This table presents industry-level regression results examining how malicious and non-malicious data breaches influence stock volatility and abnormal return. Panel A includes only malicious breaches, while Panel B includes only non-malicious breaches. Industry and time fixed effects and intercept are not tabulated. Independent variables are lagged by a week or two weeks. T statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

3.4. Conclusion

Cybersecurity risks have become one of the top challenges faced by companies today. The analysis reveals that data breaches are not isolated shocks. I find that data breach incidents are autocorrelated and exhibit both intra-industry and inter-industry spillover effects. This study further shows that stock volatility increases significantly during the week of a breach announcement, confirming that investors respond to the heightened uncertainty surrounding a firm's operations, reputation, and future performance.

One of the most notable findings is that the spillover effects on volatility and abnormal returns from data breach incidents are primarily driven by malicious breaches. The distinction between malicious and non-malicious breaches proves critical. Malicious breaches, resulting from intentional attacks for profit, have more severe financial consequences and are the primary drivers of spillover effects. In contrast, non-malicious breaches tend to occur randomly and generate weaker market reactions. Furthermore, while firms do not consistently experience abnormal returns following their own breaches, peer firms often benefit from malicious breaches within the same industry, exhibiting higher abnormal returns and reduced volatility, suggesting competitive effects rather than contagion effects. However, this effect is limited to the firm level. At the industry level, past breaches do not jointly explain variations in volatility or returns, possibly because most firms in the PRC dataset are private and therefore do not influence industry-level stock performance.

There are several potential limitations to this study. First, the firm-level analysis includes only publicly listed firms due to the lack of financial performance data for private companies. The results could be further improved by distinguishing between parent and subsidiary firms, where possible. Second, the analysis is limited to firms included in the PRC database. Other firms may have experienced data breaches but are not captured in this dataset, which could introduce bias in estimating data breach risk. Finally, one could argue that events other than data breaches (e.g., other forms of negative news) may have contributed to changes in firm performance. However, this limitation is less relevant for the industry-level analysis, as it is unlikely that unrelated events would systematically affect all firms within an industry in the same way.

APPENDIX A

APPENDIX FOR CHAPTER 1

A.0.1. Cybersecurity awareness score: relative frequency version

Research Question 2: Do extensive cybersecurity risk discussions in 10-K filings increase a firm's likelihood of suffering a future breach?

Table A.1: Research Question 2

Dependent var:	Breach
Score (ratio)	45.6957***
	(3.8981)
N	59574

This table presents regression results examining whether firms that disclose more about cybersecurity risks in their annual filings are more or less likely to experience subsequent breaches. The dependent variable *Breach*, is a binary indicator equal to 1 if a breach occurs with the observation period. The independent variable, *Score (ratio)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with weights the total number of keywords by the length of the filing. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Research Question 3: Do firms increase their attention to cybersecurity risk management after experiencing breaches in the previous year?

Research Question 4: Do firms increase their attention to cybersecurity risk management after their competitors experience breaches in the previous year?

Table A.2: Research Question 3

Dependent variable:	Score_lead (ratio)
Breach	0.0016***
	(0.0003)
Score (ratio)	0.8807***
	(0.0018)
N	51,810

This table presents regression results examining whether firms increase their attention to cybersecurity risk management after experiencing a breach in the previous year. The dependent variable *Score_lead (ratio)*, is the firm's cybersecurity awareness score at the end of the observation period, with weights the total number of keywords by the length of the filing. The independent variable, *Breach*, a binary indicator equal to 1 if a breach occurs with the observation period. *Score (ratio)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with weights the total number of keywords by the length of the filing. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table A.3: Spillover effect:Research Question 4

Dependent var:	Score_lead (ratio)
Breach_competitor	0.0004***
	(0.00005)
Breach_own	0.0015***
	(0.0003)
Score (ratio)	0.8890***
	(0.0019)
N	46109

This table presents regression results examining whether firms pay more attention to cybersecurity risk management after their competitors experience breaches in the previous year. The dependent variable *Score_lead (ratio)*, is the firm's cybersecurity awareness score at the end of the observation period, with weights the total number of keywords by the length of the filing. The key independent variable, *Breach_competitor*, is a binary indicator equal to 1 if any other firm in the same industry, defined based on the same Standard Industrial Classification (SIC) code as the focal firm, experiences a breach within the observation period. The variable *Breach_own* is another binary indicator equal to 1 if the focal firm itself experiences a breach within the same observation period. *Score (ratio)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with weights the total number of keywords by the length of the filing. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

A.0.2. Excerpts from SEC 10-K item 1A Risk Factors

Company Name: *IPG Photonics Corp*

Filing Date: 2018-02-28

> Our security measures may be breached as a result of third-party action, including intentional misconduct by computer hackers, employee error, malfeasance, or otherwise.

Company Name: *Cryoport Inc*

Filing Date: 2020-03-10

> We attempt to protect trade secrets by entering into confidentiality agreements with employees, consultants, and third parties, although, in the past, we have not always obtained such agreements. It is possible that these agreements may be breached, invalidated, or rendered unenforceable, and if so, our trade secrets could be disclosed to our competitors.

A.0.3. Decomposing Cybersecurity Awareness Score

I measure firms' cybersecurity awareness using the linguistic relevance of a dictionary of common cybersecurity terminology and related laws. To explore firms' post-breach responses in greater depth, I decompose the original cybersecurity awareness score by refining it into more granular categories with the assistance of ChatGPT. Specifically, I focus on the domain of Risk Management and Compliance. After excluding the broad and general term "risk," this category consists 12 specific terms. Using the same methodology applied to the cybersecurity awarness score, I construct a specialized cybersecurity awareness score with a focus on cybersecurity risk management based on these 12 risk management-related terms. The regression results indicate that when a firm or its industry peer experiences a breach, the firm tends to increase its emphasis on cybersecurity risk management, reflected by higher scores within this focused category, thereby reinforcing the earlier findings.

Table A.4: Summary Statistics for Cybersecurity Awareness Score with a focus on cybersecurity risk management

Score Version	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Min.**	**Max.**
Score_rm	363.2	261.18	182	307.0	479	0	3256

Research Question 2: Do extensive cybersecurity risk discussions in 10-K filings increase a firm's likelihood of suffering a future breach?

Table A.5: Research Question 2

Dependent var:	Breach
Score (RM)	0.0006***
	(0.0002)

This table presents regression results examining whether firms that disclose more about cybersecurity risks in their annual filings are more or less likely to experience subsequent breaches. The dependent variable *Breach*, is a binary indicator equal to 1 if a breach occurs with the observation period. The independent variable, *Score (RM)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with a focus on cybersecurity risk management. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Research Question 3: Do firms increase their attention to cybersecurity risk management after

experiencing breaches in the previous year?

Table A.6: Research Question 3

Dependent variable: Score_lead (RM)	
Breach	16.4955**
	(5.6625)
Score (RM)	0.9271***
	(0.0018)

This table presents regression results examining whether firms increase their attention to cybersecurity risk management after experiencing a breach in the previous year. The dependent variable *Score_lead (RM)*, is the firm's cybersecurity awareness score at the end of the observation period, with a focus on cybersecurity risk management. The independent variable, *Breach*, a binary indicator equal to 1 if a breach occurs with the observation period. *Score (RM)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with a focus on cybersecurity risk management. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Research Question 4: Do firms increase their attention to cybersecurity risk management after their competitors experience breaches in the previous year?

Table A.7: Spillover effect:Research Question 4

Dependent var:	Score_lead (ratio)
Breach_competitor	8.8250***
	(1.0808)
Breach_own	16.2717**
	(5.6182)
Score (ratio)	0.9314***
	(0.0018)

This table presents regression results examining whether firms pay more attention to cybersecurity risk management after their competitors experience breaches in the previous year. The dependent variable *Score_lead (RM)*, is the firm's cybersecurity awareness score at the end of the observation period,with a focus on cybersecurity risk management. The key independent variable, *Breach_competitor*, is a binary indicator equal to 1 if any other firm in the same industry, defined based on the same Standard Industrial Classification (SIC) code as the focal firm, experiences a breach within the observation period. The variable *Breach_own* is another binary indicator equal to 1 if the focal firm itself experiences a breach within the same observation period. *Score (RM)*, refers to the firm's cybersecurity awareness score at the beginning of the observation period, with a focus on cybersecurity risk management. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Research Question 5 :Do past competitor breaches have predictive power on firms' future breaches?

Table A.8: Research Question 5

Dependent var:	Breach_own_lead
Score	0.0006**
	(0.0002)
Breach_competitor	0.0625
	(0.1361)
Breach_own	3.4629***
	(0.1734)

This table presents regression results examining whether competitors' past breaches predict a firm's future breach incidents. The dependent variable, *Breach_ own_ lead*, is a binary indicator equal to 1 if the focal firm experiences a breach after the end of the observation period. The independent variable, *Score (RM)*, measures the firm's cybersecurity awareness at the beginning of the observation period, with a focus on cybersecurity risk management. *Breach_ competitor* is a binary indicator equal to 1 if any other firm in the same industry (defined by SIC code) experienced a breach within the observation period. *Breach_ own* is a binary indicator equal to 1 if the focal firm itself was breached within the observation period. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

A.0.4. The Effect of Time Intervals Between Consecutive 10-K Filings

To ensure that variation in the time intervals between consecutive 10-K filings does not bias the construction or interpretation of the cybersecurity awareness score, I examine whether the time difference between two filings influences the likelihood of a future cybersecurity breach.

$$\text{Logit}\left(P(\text{Breach}_{i,[t:t+\Delta_i t]} = 1)\right) = \text{Score}_{i,t} + \text{time diff}_{i,t} + \text{Score}_{i,t} \times \text{time diff}_{i,t}$$

As shown in Table A.9, none of the coefficients on time-related variables or their interaction with the score are statistically significant. This suggests that the time difference between filings does not have a meaningful effect on breach likelihood, nor does it alter the relationship between awareness scores and breach probability. Therefore, I proceed with constructing and interpreting the cybersecurity awareness score without adjusting for filing interval length.

Table A.9: Logistic Regression: Breach Likelihood and Interaction with Time Difference

Dependent var: Breach	
Score	-0.0008
	(0.0024)
Time diff	0.0006
	(0.0042)
Score × Time diff	0.0000044
	(0.0000066)

This table presents regression results examining whether the time interval between consecutive 10-K filings influences the likelihood of a future cybersecurity breach. The dependent variable, *Breach*, is a binary indicator equal to 1 if a breach occurs. *Score* measures the firm's cybersecurity awareness level at the beginning of the observation period. *Time diff* is the number of days between two consecutive 10-K filings. *Score* × *Time diff* is an interaction term testing whether the predictive power of the cybersecurity awareness score varies with the length of the filing interval. Robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

A.0.5. The financial impacts of a breach incident on a firm

To examine the effect of cybersecurity breaches on firm fundamentals, I merge breach data with Compustat based on the finalization date (fdate) of each firm's financial report. A Compustat observation is linked to two breach indicators: one occurring before and one after the financial data is finalized.

As shown in Table A.10, breaches are significantly associated with increases in cash flow, both contemporaneously and in the following period. This suggests that firms may respond to breaches by improving operational efficiency or reallocating resources. However, no statistically significant effects are observed for ROA, leverage, or R&D expenditures, indicating that breaches do not consistently affect these aspects of firm fundamentals.

Table A.10: Effect of Cybersecurity Breaches on Firm Fundamentals

Dependent variable: ROA	
Breach_lead	0.1792
	(0.3293)
Breach	0.2088
	(0.03312)
Dependent variable: Leverage	
Breach_lead)	-214.8
	(462.1)
Breach	648.89
	(464.7)
Dependent variable: Cash Flow	
Breach_lead	2158 ***
	(2158)
Breach	1278 ***
	(221.5)
Dependent variable: R&D	
Breach_lead	-5.1866
	(17.4183)
Breach	-5.6333
	(17.5163)

This table presents regression results evaluating the impact of cybersecurity breaches on firm fundamentals including ROA, leverage, cash flow, and R&D. *Breach_lead* is a binary indicator equal to 1 if a breach occurs after the financial data is finalized. *Breach* equals 1 if a breach occurs before the financial data is finalized. Robust standard errors are reported in parentheses. ***, **, * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

APPENDIX B

APPENDIX FOR CHAPTER 2

B.0.1. Prompt: Sentiment score

Prompts are critical in guiding LLM's responses to tasks. I explore multiple prompts to optimize performance and accuracy in sentiment analysis. Assigning numerical values to different sentiment categories identified by GPT on a scale lacks transparency. This obscurity arises from the challenge of understanding how LLMs calculate specific numerical scores for a designated range, whether broad or narrow. In Table B.1, I present examples of GPT outputs generated from different prompts used to analyze the sentiment score of SYNOPSYS INC's 2009 10-K Item 1A excerpt. To prevent GPT models from retaining information related to previous company filings, instructions to forget previous commands have been added. The following is the prompt used to analyze the sentiment scores of SEC 10K fillings.

Forget all your previous instructions. Pretend you are a linguistic expert. Determine whether the sentiment of this text is positive, neutral, or negative. Answer only with a number: 1 if positive, 2 if neutral, and 3 if negative. Here is the text:"

Table B.2 reports a summary of sentiment scores made by LLMs and traditional dictionary method for an excerpt from a company's SEC 10-K filing Item 1A.

Table B.1: ChatGPT Sentiment Analysis Outputs

Prompts	**GPT sentiment score**
Rate the sentiment of the text below using a scale from 0 to 5, where 0 represents extremely negative sentiment and 5 represents extremely positive sentiment.	The sentiment of the text provided can be rated as 1 on a scale from 0 to 5. The text expresses concerns about potential negative outcomes (such as breaches of agreements and inadequate remedies) that could harm the business if its proprietary technology is not adequately protected. This reflects a predominantly negative sentiment.
Rate the sentiment of the text below using a scale from 0 to 10, where 0 represents extremely negative sentiment and 10 represents extremely positive sentiment.	The sentiment of the provided text is generally negative as it discusses potential risks and concerns regarding the protection of proprietary technology. I would rate the sentiment of this text as 2 on a scale from 0 to 10.
Is the sentiment of this text positive, neutral, or negative? Answer only with a number: 1 if positive, 2 if neutral, and 3 if negative.	3
The following text is an excerpt from a company's 10-k filing. You are a finance expert. Based on this text only, please answer the following question. Determine whether the sentiment of this text is positive, neutral, or negative. Answer only with a number: 1 if positive, 2 if neutral, and 3 if negative. Here is the text:"	3
Forget all your previous instructions. Pretend you are a linguistic expert. Determine whether the sentiment of this text is positive, neutral, or negative. Answer only with a number: 1 if positive, 2 if neutral, and 3 if negative. Here is the text:"	3

Table B.2: LLM Sentiment Analysis Results

Excerpt from a company's SEC 10-K filing Item 1A	**GPT**	**Llama3**	**finBERT**	**VADER dictionary**
Company: SYNOPSYS INC Filing Date: 2009-12-18 *If we fail to protect our proprietary technology our business will be harmed. Our success depends in part upon protecting our proprietary technology. We rely on agreements with customers, employees and others and on intellectual property laws worldwide to protect our proprietary technology. These agreements may be breached, and we may not have adequate remedies for any breach and our trade secrets may otherwise become known or be independently developed by competitors.*	3 negative	3 negative	3 negative	1.795 (sum of all sentiment word scores)
Company:SPARTANNASH CO Filing Date: 2016-03-02 *Despite the Company compliance with these standards and other information security measures, the Company cannot be certain that all of its IT systems are able to prevent, contain or detect any cyber attacks or security breaches from known malware or malware that may be developed in the future. To the extent that any disruption results in the loss, damage or misappropriation of information, the Company may be adversely affected by claims from customers, financial institutions, regulatory authorities, payment card associations and others. In addition, the cost of complying with stricter privacy and information security laws and standards could be significant to the Company. Threats to security or the occurrence of severe weather conditions, natural disasters or health pandemics could harm the Company business. The Company business could be severely impacted by wartime activities, threats or acts of terrorism, severe weather conditions, natural disasters, or widespread health pandemics.*	3 negative	3 negative	1positive	-4.487 (sum of all sentiment word scores)

APPENDIX C

APPENDIX FOR CHAPTER 3

C.0.1. The effect of data breaches on volatility and abnormal return: adding interaction terms

To further examine the effect of data breaches on volatility and abnormal return for each industry, I add several interaction terms to the regression model:

$$\begin{aligned} Vol_{i,t} = {} & \tau_i + \mu_t + \beta_1 Breach_{i,t} + \beta_2 Breach_{i,t-1} + \beta_3 Breach_{i,t-2} + \beta_4 Industry_i * Breach_{i,t-1} \\ & + \beta_5 Industry_i * Breach_{i,t-2} + \eta_1 Vol_{i,t-1} + \eta_2 Vol_{i,t-2} \\ & + \gamma_1 AR_{i,t-1} + \gamma_2 AR_{i,t-2} + \delta X_{i,t} + \varepsilon_{i,t} \end{aligned}$$

$$\begin{aligned} AR_{i,t} = {} & \alpha_i + \mu_t + \beta_1 Breach_{i,t} + \beta_2 Breach_{i,t-1} + \beta_3 Breach_{i,t-2} + \beta_4 Industry_i * Breach_{i,t-1} \\ & + \beta_5 Industry_i * Breach_{i,t-2} + \eta_1 Vol_{i,t-1} + \eta_2 Vol_{i,t-2} \\ & + \gamma_1 AR_{i,t-1} + \gamma_2 AR_{i,t-2} + \delta_i X_{i,t} + \varepsilon_{i,t} \end{aligned}$$

Where $Industry_i$ is the dummy variable labeled for each industry.

Regression results are presented in Table C.1. Columns 1 and 2 shows that the existence of data breaches is not significant in explaining both stock volatility and abnormal return for each industry. Even though the frequency of data breaches is very high in the medical and other business sectors, but the overall effect of data breaches on the stock market is not significant. Next, I turn my attention to the impact of malicious and non-malicious data breaches on the stock market.

Table C.1: Market Reactions

VARIABLES	(1) $Vol_{i,t}$	(2) $AR_{i,t}$
$Breach_{i,t-4}$	-0.00004	-0.00292
	(0.484)	(1.823)
$Breach_{i,t-2}$	-0.00004	-0.00001
	(0.558)	(0.006)
$Breach_{i,t}$	0.00002	0.00044
	(0.460)	(0.595)
$Vol_{i,t-4}$	0.12970***	0.02214
	(7.673)	(0.787)
$Vol_{i,t-2}$	0.17270***	-1.35700**
	(10.215)	(3.793)
$AR_{i,t-4}$	-0.00065	0.00685
	(0.804)	(0.403)
$AR_{i,t-2}$	-0.00113	-0.01328
	(1.319)	(0.777)
$Breach_{BSONGO,t-4}$	0.00005	0.00351
	(0.437)	(1.517)
$Breach_{BSR,t-4}$	-0.00006	0.00324
	(0.608)	(1.447)
$Breach_{EDU,t-4}$	0.00004	0.00419
	(0.363)	(1.862)
$Breach_{GOV,t-4}$	0.00004	0.00318
	(0.375)	(1.862)
$Breach_{MED,t-4}$	0.00004	0.00303
	(0.367)	(1.098)
$Breach_{BSONGO,t-2}$	0.00012	-0.00058
	(1.071)	(0.253)
$Breach_{BSR,t-2}$	0.00010	0.00129
	(0.973)	(0.776)
$Breach_{EDU,t-2}$	0.00006	-0.00285
	(0.592)	(1.266)
$Breach_{GOV,t-2}$	0.00018	-0.00089
	(1.750)	(0.398)
$Breach_{MED,t-2}$	0.00011	0.00127
	(0.809)	(0.455)
Adj. R squared	29.07%	0.38%
F-statistic	3.340	1.022

This table presents regression results analyzing the impact of various breach types on volatility and abnormal returns by adding more interaction terms. Industry and time fixed effects and intercept are not tabulated. Independent variables are lagged by a week or two weeks. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

C.0.2. Industry classification

Table C.2 provides the industry classification description. The PRC data divides all companies into 8 groups, BSF, BSO, BSR, EDU, GOV, MED, NGO, and UNKN. I closely follow the Kenneth R. French Industry Portfolios definitions and matches the SIC codes to PRC 8 groups.

Table C.2: Industry Classification by Type and SIC Code

TYPE	**NAME**	**NUM**	**SIC**
BSF	Businesses (Financial and Insurance Services)	891	45 *Banks, Banking* 6000–6099 DEPOSITORY INSTITUTIONS 6100–6199 NONDEPOSITORY CREDIT INSTITUTIONS 46 *Insur, Insurance* 6300–6399 INSURANCE CARRIERS 6400–6499 INSURANCE AGENTS, BROKERS, AND SERVICE 48 *Fin, Trading* 6200–6299 Security and commodity brokers, dealers, exchanges & services 6700–6799 HOLDING AND OTHER INVESTMENT OFFICES

Continued on next page

Table C.2 – continued from previous page

TYPE	NAME	NUM	SIC
BSO+ NGO	Businesses (Other) + Non-profits	1226 +138	1–32 *Agric., Mines, Oil, Stone, Const.*, Food, Smoke, *Textile, Apparel, Wood*, Chair, Paper, Print, Chems, *Rubber, Leather*, Glass, Metal, *Mach., Elec.*, Cars, Junk, Water, Trans., *Utilities, Garbage*, Steam, Water 3400–3829 2840–3839 3852–4979 47 *RlEst, Real Estate* 6500–6599 44 Hotels, Motels 7000–7099 Hotels & other lodging places 33 PerSv Personal Services and 34 BusSv Business Services 7200–7299 PERSONAL SERVICES 7300–7399 BUSINESS SERVICES 7500–7599 AUTOMOTIVE REPAIR, SERVICES, AND PARKING 7600–7699 MISCELLANEOUS REPAIR SERVICES 7 Fun *Entertainment* 7800–7999 8100–8199 Services - legal 8300–8399 Services - social services 8400–8499 Services - museums, art galleries, botanical and zoological gardens 8600–8699 Services - membership organizations 8700–8799 ENGINEERING, ACCOUNTING, RESEARCH, MANAGEMENT, AND RELATED SERVICES 8800–8899 Services - private households 8900–8999 MISCELLANEOUS SERVICES

Table C.2 – continued from previous page

TYPE	NAME	NUM	SIC
BSR	Businesses (Retail/Merchant including Online Retail)	655	42 *Whlsl , Wholesales* 5000–5199 43 *Rtail, Retail Stores* 5200–5999
EDU	Educational Institutions	861	32 *PerSv, Personal Services* 8200–8299 Services - educational
GOV	Government & Military	801	37 *Govt, Public Administration* 9000–9999 (EXECUTIVE OFFICES 9100-9199 COURTS 9200-9299)

Continued on next page

Table C.2 – continued from previous page

TYPE	NAME	NUM	SIC
MED	Healthcare, Dedicated and Medical Insurance Services	4435	Healthcare, Medical Equipment, and Drugs 11 *Hlth* Healthcare 8000–8099 Services – health 12 *MedEq* Medical Equipment 3840–3849 Surgical, medical, and dental instruments and supplies 3850–3851 Ophthalmic goods 13 *Drugs* Pharmaceutical Products 2830–2839 2830–2830 Drugs 2831–2831 Biological products 2833–2833 Medicinal chemicals 2834–2834 Pharmaceutical preparations 2835–2835 In vitro, in vivo diagnostic substances 2836–2836 Biological products, except diagnostic substances
UNKN	Unknown	8	N/A

Bibliography

Abis, S., Veldkamp, L., 2024. The Changing Economics of Knowledge Production. The Review of Financial Studies 37, 89–118.

Acquisti, A., Friedman, A., Telang, R., 2006. Is There a Cost to Privacy Breaches? An Event Study.

Akey, P., Lewellen, S., Liskovich, I., Schiller, C., 2023. Hacking Corporate Reputations. Tech. rep., Rotman School of Management Working Paper No. 3143740, European Corporate Governance Institute – Finance Working Paper No. 948/2024.

Alderucci, D., Branstetter, L. G., Hovy, E. H., Runge, A., Ryskina, M., Zolas, N., 2019. Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata 1.

Araci, D., 2019. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.

Ash, E., Chen, D. L., Galletta, S., 2022. Measuring Judicial Sentiment: Methods and Application to US Circuit Courts. Economica 89, 362–376.

Babina, T., Fedyk, A., He, A., Hodson, J., 2024. Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics 151, 103745.

Berkman, H., Jona, J., Lee, G., Soderstrom, N., 2018. Cybersecurity awareness and market valuations. Journal of Accounting and Public Policy 37, 508–526.

Binfarè, M., 2019. The Real Effects of Operational Risk: Evidence from Data Breaches.

Campbell, J. Y., Lettau, M., Malkiel, B. G., Xu, Y., 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. The Journal of Finance 56, 1–43.

Campbell, K., Gordon, L. A., Loeb, M. P., Zhou, L., 2003. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. J. Comput. Secur. 11, 431–448.

Cao, S., Jiang, W., Wang, J., Yang, B., 2024. From Man vs. Machine to Man + Machine: The art and AI of stock analyses. Journal of Financial Economics 160, 103910.

Cavusoglu, H., , Birendra, M., , Raghunathan, S., 2004. The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers. International Journal of Electronic Commerce 9, 70–104.

Cebula, J., Young, L., 2010. A taxonomy of operational cyber security risks. Tech. rep., Carnegie Mellon University.

Chen, M. A., Wu, Q., Yang, B., 2019. How Valuable Is FinTech Innovation? The Review of Financial Studies 32, 2062–2106.

Chichernea, D., Holder, A., Petkevich, A., Robin, A., 2018. Better audits, better cybersecurity?

Cho, H., Muslu, V., 2021. How Do Firms Change Investments Based on MD&A Disclosures of Peer Firms? The Accounting Review 96, 177–204.

Choi, G.-Y., Kim, A. G., 2024. Firm-Level Tax Audits: A Generative AI-Based Measurement .

de Kok, T., 2024. ChatGPT for Textual Analysis? How to Use Generative LLMs in Accounting Research .

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Doshi, A. R., Bell, J. J., Mirzayev, E., Vanneste, B. S., 2025. Generative artificial intelligence and evaluating strategic decisions. Strategic Management Journal 46, 583–610.

D'Acunto, F., Prabhala, N., Rossi, A. G., 2019. The Promises and Pitfalls of Robo-Advising. The Review of Financial Studies 32, 1983–2020.

Ettredge, M. L., Richardson, V. J., 2003. Information Transfer among Internet Firms: The Case of Hacker Attacks. Journal of Information Systems 17, 71–82.

Firth, M., 1996. Dividend Changes, Abnormal Returns, and Intra-Industry Firm Valuations. The Journal of Financial and Quantitative Analysis 31, 189–211.

Fischer, E. A., 2014. Federal Laws Relating to Cybersecurity: Overview of Major Issues, Current Laws, and Proposed Legislation.

Florackis, C., Louca, C., Michaely, R., Weber, M., 2023. Cybersecurity Risk. The Review of Financial Studies 36, 351–407.

Fostel, A., Geanakoplos, J., 2012. Why does bad news increase volatility and decrease leverage? Journal of Economic Theory 147, 501–525.

Freeman, R., Tse, S., 1992. An earnings prediction approach to examining intercompany information transfers. Journal of Accounting and Economics 15, 509–523.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting Characteristics Nonparametrically. The Review of Financial Studies 33, 2326–2377.

Froot, K. A., Scharfstein, D. S., Stein, J. C., 1993. Risk Management: Coordinating Corporate Investment and Financing Policies. The Journal of Finance 48, 1629–1658.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. Predictably Unequal? The Effects of Machine Learning on Credit Markets. The Journal of Finance 77, 5–47.

Garg, P., 2020. Cybersecurity Breaches and Cash Holdings: Spillover Effect. Financial Management 49, 503–519.

Gatzlaff, K. M., McCullough, K. A., 2010. The Effect of Data Breaches on Shareholder Wealth. Risk Management and Insurance Review 13, 61–83.

Glasserman, P., Lin, C., 2023. Assessing Look-Ahead Bias in Stock Return Predictions Generated By GPT Sentiment Analysis.

Gleason, C. A., Jenkins, N. T., Johnson, W. B., 2008. The Contagion Effects of Accounting Restatements. The Accounting Review 83, 83–110.

Goel, S., Shawky, H. A., 2009. Estimating the market impact of security breach announcements on firm values. Information & Management 46, 404–410.

Gofman, M., Jin, Z., 2024. Artificial Intelligence, Education, and Entrepreneurship. The Journal of Finance 79, 631–667.

Gordon, L. A., Loeb, M. P., Sohail, T., 2010. Market Value of Voluntary Disclosures Concerning Information Security. MIS Quarterly 34, 567–594.

Grennan, J., Michaely, R., 2020. Artificial Intelligence and High-Skilled Work: Evidence from Analysts .

Halawi, D., Zhang, F., Yueh-Han, C., Steinhardt, J., 2024. Approaching Human-Level Forecasting with Language Models.

Hansen, A. L., Kazinnik, S., 2024. Can ChatGPT Decipher Fedspeak? .

Hassan, T. A., Hollander, S., van Lent, L., Tahoun, A., 2019. Firm-Level Political Risk: Measurement and Effects*. The Quarterly Journal of Economics 134, 2135–2202.

Henry, E., Leone, A. J., 2016. Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. The Accounting Review 91, 153–178.

Hinz, O., Nofer, M., Schiereck, D., Trillig, J., 2015. The influence of data theft on the share prices and systematic risk of consumer electronics companies. Information & Management 52, 337–347.

Holt, T. J., , Lampke, E., 2010. Exploring stolen data markets online: products and market forces. Criminal Justice Studies 23, 33–50.

Hovav, A., , D'Arcy, J., 2004. The Impact of Virus Attack Announcements on the Market Value of Firms. Information Systems Security 13, 32–40.

Islam, M. S., Wang, T., Farah, N., Stafford, T., 2022. The spillover effect of focal firms' cybersecurity breaches on rivals and the role of the CIO: Evidence from stock trading volume. Journal of Accounting and Public Policy 41, 106916.

Jamilov, R., Rey, H., Tahoun, A., 2021. The Anatomy of Cyber Risk. Working Paper.

Jansen, M., Nguyen, H., Shams, A., 2023. Rise of the Machines: The Impact of Automated Underwriting.

Jeong, C. Y., Lee, S.-Y. T., Lim, J.-H., 2019. Information security breaches and IT security investments: Impacts on competitors. Information & Management 56, 681–695.

Jha, M., Qian, J., Weber, M., Yang, B., 2025. ChatGPT and Corporate Policies .

Jiang, J., Kelly, B., Xiu, D., 2023. (Re-)Imag(in)ing Price Trends. The Journal of Finance 78, 3193–3249.

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., Stulz, R. M., 2018. What is the Impact of Successful Cyberattacks on Target Firms?

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., Stulz, R. M., 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. Journal of Financial Economics 139, 719–749.

Kang, E., 2008. Director Interlocks and Spillover Effects of Reputational Penalties from Financial Reporting Fraud. The Academy of Management Journal 51, 537–555.

Kashmiri, S., Nicol, C. D., Hsu, L., 2017. Birds of a feather: intra-industry spillover of the Target customer data breach and the shielding role of IT, marketing, and CSR. Journal of the Academy of Marketing Science 45, 208–228.

Kelton, A. S., Pennington, R. R., 2020. Do Voluntary Disclosures Mitigate the Cybersecurity Breach Contagion Effect? Journal of Information Systems 34, 133–157.

Kim, A. G., Muhn, M., Nikolaev, V. V., 2024. From Transcripts to Insights: Uncovering Corporate Risks Using Generative AI .

Ko, M., Dorantes, C. A., 2006. The impact of information security breaches on financial performance of the breached firms: An empirical investigation. Journal of Information Technology Management 17, 13–22.

Konchitchki, Y., Xie, J., 2023. Undisclosed material inflation risk. Journal of Monetary Economics 140, S82–S100.

Kuypers, M. A., Maillart, T., Paté-Cornell, E., 2016. An Empirical Analysis of Cyber Security Incidents at a Large Organization.

Lattanzio, G., Ma, Y., 2023. Cybersecurity risk and corporate innovation. Journal of Corporate Finance 82, 102445.

Lawrence, A., Minutti-Meza, M., Vyas, D., 2018. Is Operational Control Risk Informative of Financial Reporting Deficiencies? AUDITING: A Journal of Practice & Theory 37, 139–165.

Li, F., 2010. The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach. Journal of Accounting Research 48, 1049–1102.

Li, K., Mai, F., Shen, R., Yan, X., 2021. Measuring Corporate Culture Using Machine Learning. The Review of Financial Studies 34, 3265–3315.

Lopez-Lira, A., Tang, Y., 2023. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models.

Loughran, T., Mcdonald, B., 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance 66, 35–65.

Loughran, T., Mcdonald, B., 2016. Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research 54, 1187–1230.

Loughran, T., McDonald, B., 2022. Master Loughran-McDonald Word Dictionary.

Lowry, M., Michaely, R., Volkova, E., 2020. Information Revealed through the Regulatory Process: Interactions between the SEC and Companies ahead of Their IPO. The Review of Financial Studies 33, 5510–5554.

Martin, K. D., Borah, A., Palmatier, R. W., 2017. Data Privacy: Effects on Customer and Firm Performance. Journal of Marketing 81, 36–58.

Matsumura, E. M., Prakash, R., Vera-Mu oz, S. C., 2014. Firm-Value Effects of Carbon Emissions and Carbon Disclosures. The Accounting Review 89, 695–724.

Modi, S. B., Wiles, M. A., Mishra, S., 2015. Shareholder value implications of service failures in triads: The case of customer information security breaches. Journal of Operations Management 35, 21–39.

Murray, S., Xia, Y., Xiao, H., 2024. Charting by machines. Journal of Financial Economics 153, 103791.

Pennebaker, J. W., Boyd, R. L., Jordan, K., Blackburn, K., 2015. LIWC2015 Manual: Operation Instructions. University of Texas at Austin.

Pirounias, S., Mermigas, D., Patsakis, C., 2014. The relation between information security events and firm market value, empirical evidence on recent disclosures: An extension of the GLZ study. Journal of Information Security and Applications 19, 257–271.

Pooser, D. M., Browne, M. J., Arkhangelska, O., 2018. Growth in the Perception of Cyber Risk: Evidence from U.S. P&C Insurers. The Geneva Papers on Risk and Insurance. Issues and Practice 43, 208–223.

Pyo, Y., Lustgarten, S., 1990. Differential intra-industry information transfer associated with management earnings forecasts. Journal of Accounting and Economics 13, 365–379.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving Language Understanding by Generative Pre-Training.

Ramnath, S., 2002. Investor and Analyst Reactions to Earnings Announcements of Related Firms: An Empirical Analysis. Journal of Accounting Research 40, 1351–1376.

Ray, P. P., 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 3, 121–154.

Romanosky, S., Hoffman, D., Acquisti, A., 2014. Empirical Analysis of Data Breach Litigation. Journal of Empirical Legal Studies 11, 74–104.

Schmidt, P. J., Wood, J. T., Grabski, S. V., 2016. Business in the Cloud: Research Questions on Governance, Audit, and Assurance. Journal of Information Systems 30, 173–189.

SEC, 2011. CF Disclosure Guidance: Topic No. 2 - Cybersecurity.

SEC, 2018. Commission Statement and Guidance on Public Company Cybersecurity Disclosures.

SEC, 2023. SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies.

Sen, R., , Borle, S., 2015. Estimating the Contextual Risk of Data Breach: An Empirical Approach. Journal of Management Information Systems 32, 314–341.

Smith, C. W., Stulz, R. M., 1985. The Determinants of Firms' Hedging Policies. The Journal of Financial and Quantitative Analysis 20, 391–405.

Spanos, G., Angelis, L., 2016. The impact of information security events to the stock market: A systematic literature review. Computers & Security 58, 216–229.

Tweneboah-Koduah, S., Atsu, F., Prasad, R., 2020. Reaction of stock volatility to data breach: An event study. Journal of Cyber Security and Mobility 9, 355–384.

Wang, T., , Jackie Rees, U., , Kannan, K., 2013a. The Textual Contents of Media Reports of Information Security Breaches and Profitable Short-Term Investment Opportunities. Journal of Organizational Computing and Electronic Commerce 23, 200–223.

Wang, T., Kannan, K. N., Ulmer, J. R., 2013b. The Association Between the Disclosure and the Realization of Information Security Risk Factors. Information Systems Research 24, 201–218.

Wang, Z., Xie, Q., Feng, Y., Ding, Z., Yang, Z., Xia, R., 2023. Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study. arXiv e-prints p. arXiv:2304.04339.

Wheatley, S., Maillart, T., Sornette, D., 2016. The extreme risk of personal data breaches and the erosion of privacy. The European Physical Journal B 89, 7.

Woodhouse, D., Charlesworth, A., 2023. Can ChatGPT Predict Future Interest Rate Decisions? .

Yayla, A. A., Hu, Q., 2011. The impact of information security events on the stock value of firms: the effect of contingency factors. Journal of Information Technology 26, 60–77.