Valuing Companies' Data Safety: An Empirical Analysis of Investors' and Customers' Responses to Cybersecurity Failures

Alice Ferrari

Dissertation written under the supervision of Professor José Garcia Revelo

Dissertation submitted in partial fulfillment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, January 2025.

Abstract

Cyber breaches are becoming more and more frequent due to the increasing connection allowed by technology. This implies a loss of sensitive data about companies or clients. Cyber breaches can have several causes and can happen in different forms. This paper aims to analyze how cyber data leaks affect the value of publicly traded firms by considering a sample of US companies that experienced breaches between 2009 and 2019. The research will provide some consensus about how customers react to cyber breach announcements. Both investors and customers are expected to react negatively to such events. Results show that the market reaction to such announcements is negative and significant, and cumulative abnormal returns are negatively and significantly impacted by the characteristics of the breach. However, customers' reactions are not as negative as investors' reactions, indicating that the two categories of stakeholders consider different aspects to value their relationship with companies.

Keywords: *cybersecurity, data breach, breach announcement, cumulative abnormal returns, net sales, investors, customers*

Title: Valuing Companies' Data Safety: An Empirical Analysis of Investors' and Customers' Responses to Cybersecurity Failures

Author: Alice Ferrari

Resumo

As violações cibernéticas estão a tornar-se cada vez mais frequentes, devido ao aumento da conectividade proporcionada pela tecnologia. Isto implica a perda de dados sensíveis de empresas ou clientes. As violações cibernéticas podem ter várias causas e ocorrer de diferentes formas. Este artigo tem como objetivo analisar como as fugas de dados cibernéticos afetam o valor das empresas cotadas em bolsa, considerando uma amostra de empresas norte-americanas que sofreram violações entre 2009 e 2019. A pesquisa fornecerá algum consenso sobre como os clientes reagem aos anúncios de violações cibernéticas. Tanto os investidores quanto os clientes são esperados a reagir negativamente a tais eventos. Os resultados mostram que a reação do mercado a esses anúncios é negativa e significativa, e os retornos anormais acumulados são negativamente e significativamente impactados pelas características da violação. No entanto, as reações dos clientes não são tão negativas quanto as dos investidores, indicando que as duas categorias de stakeholders consideram diferentes aspetos para avaliar a sua relação com as empresas.

Palavras-chave: cibersegurança, violação de dados, anúncio de violação, retornos anormais acumulados, vendas líquidas, investidores, clientes
Título: Avaliando a Segurança de Dados das Empresas: Uma Análise Empírica das Respostas de Investidores e Clientes a Falhas de Cibersegurança

Autor: Alice Ferrari

Acknowledgments

I wish to express my most sincere gratitude to my supervisor, José Garcia Revelo, for his guidance during the writing of this thesis. He generously shared his expertise with me, as well as important suggestions that helped me reach this objective. His support and advice were fundamental to completing this work.

I would also like to thank the inspiring professors of Católica Lisbon, especially those I crossed paths with, for the engagement they showed in teaching and the invaluable experiences they allowed me to gain. They have significantly increased my academic and personal knowledge.

Finally, I wish to express my gratefulness towards my classmates, those that I had the pleasure to work with during the courses and the preparation of the thesis. The collaboration with them has incredibly enhanced my background.

Executive summary

This paper can be of particular usefulness to US insurance companies that offer policies against cyber-attacks, US regulators, US policymakers, and companies that wish to be informed about cybersecurity matters.

A sample of US firms was analyzed to understand how their shareholders and customers reacted to cybersecurity breach announcements. The observations go from 2009 to 2019 and involve 260 observations in the analysis of cumulative abnormal returns. The subsequent regressions, that analyze the factors that impact CARs and the customers' reactions, involve 177 observations, as the extreme values were deleted from the sample to prevent them from creating noisy results.

Literature about investors' reactions to these events is very complete, and there is a clear consensus. All authors find a negative impact on CARs after the announcement of cyber breaches, and this research finds that they are negatively impacted by any kind of information stolen; they are also negatively and significantly impacted when they involve sectors such as the financial, medical, retail, and education ones. Also, a very negative and significant impact is due to hackers' action, but neither the number of people affected by the breach nor its duration appear to be of any significance to investors.

Literature about customers' reactions is very scarce. This paper will show how the relative change in net sales is impacted by the announcement of cyber events. The findings indicate that, unless the stolen information is highly sensitive, breaches could even generate indirect advertising for the breached companies, increasing the relative change in net sales from the quarter preceding the breach announcement to the quarter of the breach announcement. However, the change in net sales is negatively and significantly impacted when it happens in sectors that deal with strictly confidential information, such as the financial and medical sectors.

However, while reading this paper, the reader should keep in mind some limitations. First, usually, companies impacted by a cyber breach tend to delay the disclosure of such negative events. Second, this research is carried out over a limited period and only considers US companies, which is why some hints for future research are proposed in the conclusion.

Table of Contents

1. Introduction

In the last few decades, digital transformation has had as a direct consequence the creation of new threats in terms of cyber security. Cyber security has become one of the central concerns of companies, as data is one of the most important assets that a firm can own (Boroomand et al., 2022). In particular, some sectors are becoming more and more reliant on technology, implying big damage if they are hit by any cyber-attack. When data leaks occur, they impact the value of the involved companies, having markets react to such events, even though cyber costs are very difficult to quantify (Aldasoro et al., 2020), as they might have both monetary and/or reputational costs. This is why some targeted companies avoid reporting the incidents to financial authorities. Aldasoro et al. (2020) provide in their paper some statistics showing that, between 2002 and 2019, the average cost per cyber incident over the period was $2.6 million, but of course, tail events are not considered, so in extreme scenarios, the cost for a cyber incident could be much higher. The possible losses implied by a cyber-attack, however, are not limited to money or reputation, but they could also imply operational disruptions, such as the impossibility of relying (even if for a limited amount of time) on cloud computing (i.e., Google Cloud, etc.) or unavailability of payment systems. Not only the aforementioned services could be made impossible during a cyber breach, but they could be the spread vehicle of the cyber-attack itself.

In light of this, it becomes important to understand if and in what measure the shareholders' value of a company is impacted by the announcement of a data breach, as well as what factors influence the change in value, considering the characteristics of the data breach. This research will not only try to provide an answer to these matters, but it will also consider the reaction of customers to the announcement of the same event to provide some consensus on the topic. Results will show that cumulative abnormal returns are negative and significant, and they are mainly negatively and significantly impacted by the characteristics of the breach. Customers' reactions are different with respect to shareholders', they don't perceive the breach as badly as investors, so some characteristics of the breach have a positive and significant impact on net sales in the quarter of the breach; some others are negative and significant. However, a limitation that should be considered throughout the reading of this paper is that companies tend not to disclose these negative events, so the announcement could be due to third-party disclosures or to breached companies' disclosure, but later with respect to the discovery of the data leak.

Cyber-attacks are just one of the idiosyncratic risks due to technology. Cyber incidents can have multiple causes and actors, as well as consequences. Aldasoro et al. (2020) list in their paper some possible actors, their motives, the methods they use, and the possible consequences. Actors include criminal organizations, industrial spies, hacktivists, terrorists, and insiders. A figure that gained importance in the last years, with the historical events that happened, is the hacktivist. Hacktivists operate with political or social purposes and conduct cyber-attacks to support their ideals in a conflict. Some of the most famous attacks that have been recently performed by the cyber-activist group Anonymous are the ones against Russia and Putin since the war between Russia and Ukraine started. As documented by Tidy (2022), they aimed to weaken Russia and move some of the resources it was using for the war to another front. This, therefore, represents a geopolitical motive, which is one of those identified in the paper by Aldasoro et al. (2020), together with profit or discontent, however, some cyber incidents could also be unintentional. The consequences include "simple" fraud or data theft, but as mentioned previously, they could imply business disruptions or reputational costs, as well as (but more rarely) systemic crisis, depending on the target of the attack. The causes of cyber incidents could be many, even non-malicious, just like in the case of IT implementation errors or other types of insider mistakes; intentional attacks may include malware (software that can damage a network, computer, or server, as defined by Yasar et al., 2024), phishing (they exploit human mistakes to obtain the access to data, as defined by Gillis, 2024), password attack, denial-of-service (they forbid the access of users to data or services, defined by Ferguson, 2021), man-in-the-middle (the hacker positions himself in a conversation between a user and an application, defined by Yasar et al., 2022).

Cybersecurity is an important part of today's world, and this is shown by the estimated cost of cybercrime in 2022, which was $8 trillion worldwide, as reported by Hernandez (2023). In particular, as computed by Petrosyan (2024), the sectors that incur the highest costs when involved in a data breach are the healthcare sector (9.77 million U.S. dollars on average) and the financial sector (6.08 million U.S. dollars on average), considering a worldwide sample. However, since this research focuses on data breaches that occurred in the United States, it would be important to consider the monetary cost of data breaches when they happen in that region. According to Petrosyan (2024), as of 2024, it amounts to 9.36 million U.S. dollars for U.S. firms, much higher with respect to the world average, which amounts to 4.88 million U.S. dollars. In light of this, in 2018, the urge for some regulation to guide companies toward the correct management of data led to the implementation of the European GDPR (General Data

Protection Regulation), which impacted all the companies doing business with European firms (Boroomand et al., 2022).

To help the readers best navigate this paper, a short recap of its structure is presented as follows: this thesis will be divided into five sections; after this first introductory section, in section 2, some previous literature will be deeply reviewed in order to provide a proper background for the topics that will be later discussed. Section 3 will explain how data were retrieved, the hypotheses and models, as well as the methodology used to perform the analysis of data. Section 4 will present the results obtained through the data analysis, comparing them to the previous literature. In Section 5, the conclusion to the paper will be presented, together with some hints as a starting point for future research.

2. Literature Review

With the evolution of technology, firms have been increasingly facing the risk of cyber-attacks, which, as well as other types of risks, impact their way of conducting everyday business. Weaknesses in the online activities of companies could represent a threat to their well-functioning and, therefore, to their value. This is why the topic of cyber security has now been studied by many authors, and new ways of preventing cyber risk are being investigated. Previous literature analyses the subject from different perspectives, which can be very useful because it creates an exhaustive context to place this research. However, since cybersecurity is a relatively new type of risk, it would be interesting to analyze how similar types of firm-specific or idiosyncratic risk (namely, the risk is unique to a specific firm, and it is independent of market movements, as defined in Fu, 2009) impacts companies, considering some variations of it. For example, the supply chain risk and the operational risk could be taken into consideration because they follow shocks that usually disrupt the normal operations of companies, just like a cyber-attack. Hendricks et al. (2019) explored how the supply chain disruptions following the 2011 Great East Japan earthquake impacted the stock prices of companies. Supply chain risk, as defined by March and Shapira (1987), is the change in the supply chain, the probability of its disruption, and the consequent monetary or reputational losses. Results by Hendriks et al. (2019) showed, through the event study methodology, that, on average, companies lost a statistically significant 5,21% of their shareholder value. Cummins et al. (2006) showed that the value of banks and insurance companies is significantly and negatively impacted by operational losses, which are defined as the risk of a loss caused by the firm's operating technology or the risk of a loss caused by agency costs.

Now, diving into cyber risk, some papers provide important insights that are interesting to analyze before conducting this research.

2.1 Investors' reaction

Hogan et al. (2023) showed that the industries that have more direct interactions with their clients are more subject to cyber-attacks. Using the event study methodology, the authors found that cybercriminals prefer to steal personal identity, financial, and health information rather than information related to the company's business; also, it is more frequent that cyber events originate from actors that are external to the firm. As for how data leaks impact the value of companies, the authors found that the companies collect significant negative returns around the announcement date of the cyber-attack, even though their magnitude is smaller with respect to what prior literature had found. They guess that this different result is due to the sample size, which in their research is much wider than those used previously. Moreover, analyzing each industry separately, the results showed that the service industry suffers large negative returns, while by the 5th day after the announcement date, the finance industry experiences a market reversal. This is the case because, based on the findings, investors react worse to leaks of personal information; loss of identity and health information still cause cumulative negative returns, but they are smaller. Arcuri et al. (2018) also found an overall negative stock market reaction, but, as opposed to the previous paper, they found that the most negative return among the analyzed sectors happen in the financial industry, even more than in the other economic sectors. The authors speculate that this is due to the fact that cybercriminals commit insider trading. The results, obtained from running an event study, support the idea that companies should invest in cybersecurity measures, including insurance. The stronger impact in the financial sector on other industries was confirmed by Buchanan et al. (2018), who also showed that technology firms don't have a big reaction to cyberattack announcements; the authors suppose it is because that type of firms have more tools to act quickly in response to those events.

Juma'h et al. (2019) found that breached companies, on average, performed worse in the quarter of the breach in terms of ROA and ROE. The findings show that announcements of data breaches are seen by investors as signals of breached companies' internal weakness, which implies that those companies are considered liable to their customers, employees, and investors. The signaling action of cyber-attacks is also highlighted by Kamiya et al. (2021), who argued that if the attack gives a clear signal to the firm and its stakeholders about the cost and the likelihood of future attacks, that information is likely to influence the breached company's

value. A part of the costs includes reputation costs, but apparently, shareholders value more personal financial information. The cyber-attacks that imply a loss of personal financial information cause a more significant loss in shareholders' wealth. According to the authors, the decline in companies' value is mostly due to factors that include the likelihood of future cyber-attacks and the cost attributable to past attacks. The higher those estimates, the higher the company is perceived as risky. This is why attacked companies tend to invest more in risk management and decrease management risk-taking incentives to compensate for cyber risk. However, the paper also proved how not only the breached company is affected by the attack, but its competitors as well; the cyber breach reveals the expected cost of attacks for competitor firms as well.

As Tosun (2021) showed, in the short term, after the announcement of the cyber-attack, the attention to the breached company increases, and investors react by selling shares, decreasing the value of the firm. Interestingly, the author discovered that, in the short term, target firms rely on the existing management to overcome the shock, while, in the longer term, they are more likely to invest more in their CEO. Leroy (2022) defined instead a few strategies that, according to the results, could be useful after a cyber-attack to help the company recover in terms of time and money. The author named seven strategies to manage companies' reputation, including transparency, reporting, and security, showing that the more strategies are used, the less time will be needed by companies to recover from a cyberattack in terms of share price. Bose and Leung (2013) showed that, when companies adopt tools and measures against data thefts, their stock prices increase. The same happens if companies create the position of a CIO (Chief Information Officer), according to Chatterjee et al. (2001).

2.2 Customers' reaction

In terms of customers' reaction to the announcement of data breaches, there is little consensus. Hogan et al. (2023) showed that customers' reaction, proxied by net sales change, is worse when breaches happen in the financial sector rather than in others, but the impact is close to zero when breaches involve personal financial or health information. Close to zero is also the impact of variables that describe firm characteristics, such as size and liquidity.

Overall, the literature studied data breaches and the way they influence companies from several perspectives, but what clearly emerges is that a decrease in market prices of companies is expected after the announcement of a cyber-attack. There is not a clear consensus about what sector is the most affected in these terms, but all the authors propose several ways to prevent

and react to these events to try to limit the damages to the targeted companies. Also, customers' reaction appears to be much lighter with respect to investors' ones, and it also depends on different factors, but especially this last point needs to be more thoroughly investigated as research about it is very scarce.

3. Hypotheses, Data, and Methodology

3.1 Hypotheses

As mentioned in the literature section, cyber events belong to the idiosyncratic type of risk, so they impact companies independently from market movements. It is logical to guess that cyber-attacks are considered negative news by companies' shareholders. Therefore, it is reasonable to state that, in the short run, companies have negative returns around the announcement date of cyber-attacks.

H1: companies' short-term returns are negative around the announcement date of cyber events.

To explain CARs, the characteristics of the breaches will be used. The analysis will consider the industry the breached company belongs to, the type of stolen information, the incident type, the length of the attack as a measure of intensity, and the natural logarithm of the number of people impacted by the attack, as in Hogan et al. (2023). So,

H2: companies' CARs following announcements of data breaches are determined by the organization type, the type of stolen information, the breach type, the intensity of the attack, and the natural logarithm of people impacted by the attack.

CARs computation explains the market perception of the event. To analyze customer perception of cyber-attacks and provide some additional consensus to the literature, an analysis of the relative change of net sales between the quarter preceding the breach announcement and the quarter of the breach announcement will be carried out. The relative change in net sales will be partly explained using the information of companies' financial statements data from the quarter preceding the breach announcement. Additionally, information about breaches will be included (i.e., the type of attack, the type of information stolen, the sector the breached company belongs to, the number of people impacted by the breach, and the breach intensity). The quarterly income statement data includes Cash, Assets, and BV/MV, all from the quarter preceding the breach announcement, as the methodology used in Hogan et al. (2023). However, I am using

accounting data just one quarter before the breach announcement, which is most likely very similar to those in the breach announcement quarter, unlike in Hogan et al. (2023), which considers the accounting data one year apart from the breach announcement. To avoid endogeneity issues, I have therefore computed the relative change of net sales between $t = -1$ and $t = 0$. The relative change of net sales was computed by subtracting the natural logarithm of net sales in the quarter preceding the breach announcement from the natural logarithm of net sales in the quarter of the breach announcement, to avoid inserting as an explanatory variable Net Sales at $t = -1$, as in Hogan et al. (2023). To these variables, I have added the firm size proxied by Net Assets Value, while R&D has been excluded as none of the companies in this sample shows R&D expenses. Therefore, four different models for relative change in net sales will be estimated so that each single model will add different combinations of variables to allow the study of the impact of each variable without and with the impact of others (i.e., a model with numerical variables, a model with dummy variables, a model with the previous two combined, a model with the first two combined and the addition of another dummy "the Quarter Dummy").

H3: the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity and data from the quarter preceding the breach, including cash, assets, BV/MV, and the number of people impacted by the breach.

H4: the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the type of stolen information, type of breach, company type, the number of people impacted by the breach, and breach intensity.

H5: the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity, and data from the quarter preceding the breach, including cash, assets, BV/MV, and type of stolen information, type of breach, and company type, and the number of people impacted by the breach.

H6: the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity, the quartile dummy, and data from the quarter preceding the breach, including cash, assets,

BV/MV, and type of stolen information, type of breach, and company type, and the number of people impacted by the breach.

The last hypothesis considered in this paper (*H4*) has as a starting point *H3*. It then adds the effect of the Quarter Dummy, which is a dummy that equals 1 if the value of the variable $TIME_i$ is in its first quartile, 0 otherwise. The variable $TIME_i$ reports the number of days that passed from the end of the breach to its announcement. The dummy aims to show that the more time passes from the breach end to its announcement, the less it will impact the stock price of companies. This could be because the more time passes from the attack, the less it will be considered dangerous, as the stolen information will become obsolete.

The section that follows describes the data and models that will be used to investigate the hypotheses mentioned.

3.2 Sample and Data Description

The sample of public companies that faced a data leak was extracted from the PRC database[1], which contains data breaches that happened between 2005 and 2019. It includes data about the targeted firms, including their name, the sector they operate in, the announcement date of the data breach (which is used as the event date for the purpose of this research), as well as the start and end date of the breach, a description of the company and the event (if available). I selected only companies based in the US, as there were too few from other countries to make a reliable comparison, and eliminated all observations that included missing variables. Then, since both public and private companies were listed in the sample, I manually obtained the public companies' tickers, which reduced the sample to 260 observations. The observations that were left were made between 2009 and 2019.

An additional variable was added to the ones already provided by the database, called "*Intensity*" and computed as the number of days the data breach lasted, as a proxy for the intensity of the attack indeed.

The tables below indicate how data are distributed according to some variables (namely, sector, breach types, stolen information, year, number of people impacted, intensity, and time); Table 1 shows how the data breaches are distributed across industries.

[1] PRC (Privacy Rights Clearinghouse) is a non-profit foundation, whose aim is to support consumers' awareness about privacy and data protection, as explained by Tosun (2021).

Table 1

Frequency distribution by Industry of 260 Data Breaches, announced by US firms

Industry	Frequency	%
Financial	63	24,2
Retail	27	10,4
Education	1	0,4
Medical	24	9,2
Other	145	55,8
Total	260	100

A clarification about the types of sectors presented in the table above is needed in order to properly interpret it. Financial stands for the financial sector (including investment funds, banks, insurance companies, etc.); Retail refers to the retail and merchant businesses; Education indicates educational institutions; Medical stands for healthcare entities; Other encompasses all the businesses that do not belong to the previous categories. The precise definition for each variable was found in the "Overview of Data Breach Chronology Database" provided by PRC.

Table 2 shows how they are distributed depending on the breach type:

Table 2

Frequency distribution by Breach Types of 260 Data Breaches, announced by US firms

Breach Type	Frequency	%
Card	2	0,8
Malware	174	66,9
Insider	25	9,6
Physical	2	0,8
Portable	6	2,3
Stationary	2	0,8
Unintended	49	18,8
Total	260	100

where Card indicates credit/debit card frauds (not related to hackings); Malware indicates hacking/malware attack by external criminals; Insider includes insiders' breaches; Physical indicates physical documents loss or theft; Portable stands for the theft of portable devices; Stationary is for the loss or theft of stationary computers; Unintended indicates non-intentional disclosures, that do not involve physical loss or hackings (from the manual "Overview of Data

Breach Chronology Database"). Data therefore suggests that breaches are mainly due to actors that are external to the firms.

Table 3 displays how data breaches are distributed by type of stolen data:

Table 3

Frequency distribution by Stolen Information of 260 Data Breaches, announced by US firms

Information	Frequency	%
Identifier	243	93,5
Login	11	4,2
Government	4	1,5
Health	1	0,4
Commercial	1	0,4
Total	260	100

More in detail, Identifier includes a wide range of personal information, such as real names, postal or email addresses; Login indicates sensitive log-in information, credit/debit card numbers and their security codes, etc.; Government stands for sensitive government identifiers, for example passport numbers or driver's license numbers; Health indicates personal health information; Commercial encompasses information of personal property, including products or services purchased (from the manual "Overview of Data Breach Chronology Database"). What is remarkable, after some additional statistical computations, is that out of 211 non-unintended breaches, 195 targeted personal information. This supports the findings of Hogan et al. (2023) in terms of cybercriminals' preference for data to steal.

Table 4 reports how they are distributed across years. Due to the elimination of observations with missing variables, the sample only includes events starting from 2009. There has been a clear increase in the number of breaches over the years, with an average increase of 117% each year. This is most likely linked to the improvement in technology and an augmented connection in networks. The table is reported below:

Table 4

Frequency distribution by Year of 260 Data Breaches, announced by US firms

Year	Frequency	%
2009	1	0,4
2010	3	1,2
2011	2	0,8
2012	1	0,4
2013	8	3,1
2014	8	3,1
2015	15	5,8
2016	50	19,2
2017	50	19,2
2018	56	21,5
2019	66	25,4
Total	260	100

In the regressions, three numerical variables describe some characteristics of the attack. They are: *People Impacted,* which reports how many people were impacted by the attack; *Intensity*, which is the number of days that the breach lasted; *Time*, which measures the number of days between the end of the breach and its announcement and will be used to compute the Quartile Dummy. Each of these variables had very big variances, so to eliminate the noise from the dataset, I got rid of extreme values[2]. The number of observations left is now 177, which will be used to run the regressions about investors' and customers' reactions. Table 5 shows some statistics about two numerical values (e.g., the number of people impacted by the attack, the breach intensity, and time). The following table contains Min, Max, Median, Average, 25th, and 75th quartiles.

[2] Information about the computation to identify the outliers is in Appendix A.

Table 5

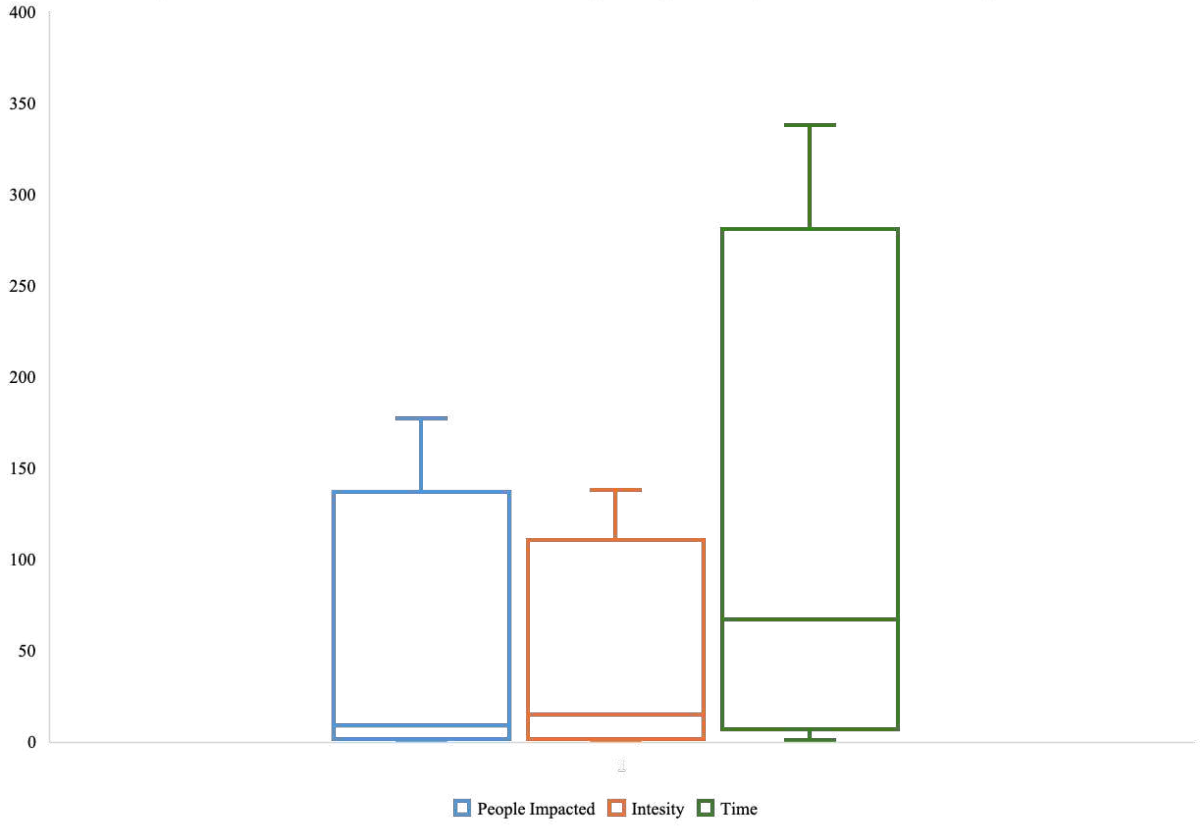
Descriptive statistics about the number of people impacted by the breach and its intensity

	People Impacted	Intesity	Time
Min	1	1	1
25th percentile	2	2	24
Mean	19	20	78
Median	4	8	40
75th percentile	16	28	110
Max	177	138	338

To make the descriptive statistics about the three variables more intuitive, a football field graph is presented.

Graph 1

Descriptive statistics about the number of People Impacted by the breach, Intensity, and Time



This graph was created by the author using her own calculations.

This research will also investigate the customers' reaction to such announcements. To do so, data was partly extracted from the data obtained from the PRC database (i.e., the sector, the type of stolen information, the type of breach, and the time), while data about companies'

quarterly financial statements was obtained from Fundamentals Quarterly on Compustat – Capital IQ, and they are expressed in million dollars. Below are the descriptive statistics about Net Sales both in the quarter of breach announcement and in the preceding quarter: Cash; Assets; BV/MV, which was computed using as a proxy for BV the total stockholders' equity.

Table 6

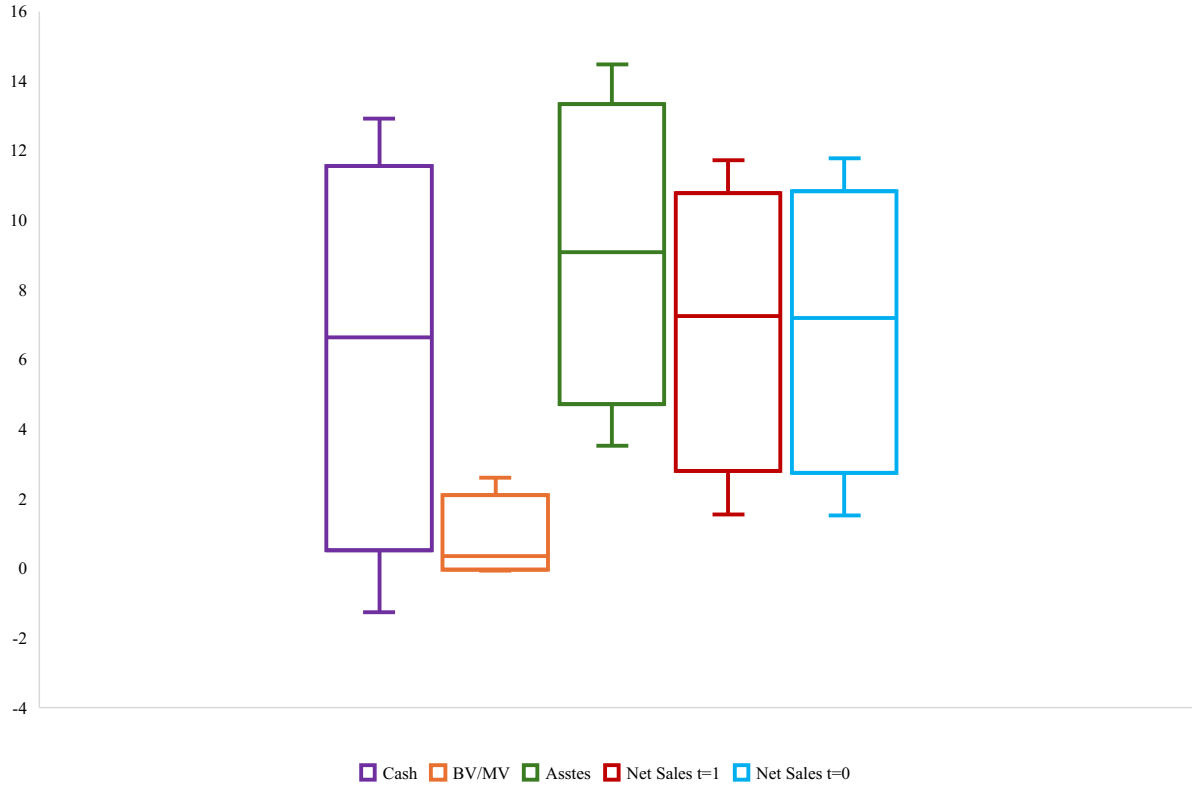
Descriptive statistics about the accounting data, namely Cash, BV/MV, Assets, Net Sales in the quarter of breach announcement, and Net Sales in the quarter preceding the breach announcement (in M$)

	Cash	BV/MV	Asstes	Net Sales t=1	Net Sales t=0
Min	0	-0,08	33	5	5
25th percentile	334	0,04	3.833	663	623
Mean	6.793	0,40	63.077	3.690	3.611
Median	777	0,25	4.940	1.016	842
75th percentile	1.697	0,63	19.433	2.884	2.812
Max	399.687	2,59	1.899.511	120.319	129.667

To make the descriptive statistics about the variables more intuitive, a football field graph is presented, containing the variables expressed with their natural logarithms (except for BV/MV) in order to decrease the range of values and make them representable in a graph.

Graph 2

Descriptive statistics about time and accounting data, namely Cash, BV/MV, Assets, Net Sales at t=1, and Net Sales at t=0



This graph was created by the author using her own calculations.

In the regressions, these variables (except BV/MV) will be included using their natural logarithm to make them comparable to other variables that have smaller values.

The next paragraph explains how CARs will be explained using the information contained in the PRC database regarding the breach.

3.3 Estimation of Abnormal Returns

The methodology used in this paper to assess whether the shareholders of the breached companies suffered a negative impact on their short-term wealth is the Event Study Methodology. To conduct an Event Study, it is necessary to make some fundamental assumptions, as defined in Sorescu et al. (2017): i) markets are efficient, which implies that stock prices reflect all information that is publicly available and that they immediately change to embed new information as soon as it becomes publicly available; ii) it can be derived that the event needs to be unanticipated, otherwise, it would be reflected in the stock price before its announcement; iii) during the event window, there must be no confounding event, unless it is accounted for. In this paper, the assumption of no confounding events has been made.

The Event Study conducted in this research aims to investigate the cumulative abnormal returns (CARs) that the breached companies collected after data leak announcements to check whether their shareholders' wealth changed. To do so, the Market Model (or CAPM – Capital Asset Pricing Model– by Markowitz) was used to compute the expected companies' returns, as in Bello (2008):

$$E(R_{it}) = r_{it} - r_{ft} = \alpha_i + \beta_i\left(r_{mt} - r_{ft}\right) + \varepsilon_{it}, \text{ where:}$$

$E(R_{it}) = r_{it} - r_{ft}$ is the expected return on company i's stock at time t;

α_i is the return on company i at time t if the dependent variables are equal to zero;

r_{ft} is the risk-free return at time t;

$r_{mt} - r_{ft}$ is the market risk premium, computed as the difference between the average rate of return of all stocks trading in the market at time t (r_{mt}) and the risk-free return in period t;

ε_{it} is the residual excess return for the portfolio i's stock in period t;

β_i is the factor-beta of the excess return on the market[3];

[3] Usually estimated with a regression over the 100 or more days prior to the event date, Sorescu et al. (2017), Event Study in the Marketing Literature: an Overview

i indicates the breached companies.

The cumulative abnormal returns were obtained through the U.S. Daily Event Study on the Center for Research in Security Prices (CRSP). More than one event window will be considered and $t = 0$ is the event date. The period selected to compute expected returns (e.g., estimation period) goes from $t = -141$ to $t = -21$, with 20 days of gap between the estimation window and the event window, as in Arcuri et al. (2018).

CARs are defined as the sum of abnormal returns observed during the event window, i.e.,

$$CAR = \sum AR_{it} ,$$

where, AR_{it} stands for the abnormal returns of company i at time t.

CRSP defines abnormal returns as

$$AR_{it} = R_{it} - [R_{ft} + \alpha_i + \beta_i(r_{mt} - r_{ft})],$$

It is, therefore, possible to derive that

$$AR_{it} = R_{it} - E(R_{it}).$$

Assuming no cross-sectional dependence in (cumulative) abnormal returns, to assess the significance of the results, the cross-sectional t-test for both abnormal returns and cumulative abnormal returns was obtained from CRSP as well. The cross-sectional t-test on CRSP is defined as the non-statistically significant difference from zero of mean (cumulative) abnormal returns, and according to the manual for overview of significance tests on WRDS, its formula is:

$$t_{it} = \frac{\frac{1}{M}\sum_{i=1}^{M} RetVar_{it}}{\sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}[RetVar_i - \frac{1}{M}\sum_{i=1}^{M} RetVar_i]^2}}$$

where,

$$RetVar_{it} = AR_i, CAR_i$$

To explain CARs, a regression will be run to assess whether *H2* should be rejected or not. The regression (Equation I.1) is the following:

$$CARs_i = \alpha_i + \beta_{i1}\, INFO\_ID_i + \beta_{i2} INFO\_LOGIN_i + \beta_{i3}\, INFO\_HEALTH_i +$$
$$\beta_{i4} INFO\_COMMERCIAL_i + \beta_{i5} SECTOR\_FIN_i + \beta_{i6}\, SECTOR\_RETAIL_i + \beta_{i7}\, SECTOR\_EDU_i +$$

$$\beta_{i8}\, SECTOR\_MED_{\,i} + \beta_{i9}\, BREACH\_CARD_{\,i} + \beta_{i10}\, BREACH\_HACK_{\,i} +$$
$$\beta_{i10}\, BREACH\_INSIDER_{\,i} + \beta_{i12}\, BREACH\_PHYSICAL_{\,i} + \beta_{i13}\, BREACH\_PORTABLE_{\,i} +$$
$$\beta_{i14}\, BREACH\_STATIONARY_{i} + \beta_{i15}\, INTENSITY_{i} + \beta_{i16} \ln (PEOPLE\ IMPACTED)_{\,i} + \varepsilon_{i}.$$

The variables in the model are the following:

- $INFO\_ID_{\,i}$, which equals 1 if the stolen information is identity, 0 otherwise information, for company i.
- $INFO\_LOGIN_{i}$, which equals 1 if the stolen information is sensitive login information, 0 otherwise, for company i.
- $INFO\_HEALTH_{\,i}$, which equals 1 if the stolen information is health information, 0 otherwise, for company i.
- $INFO\_COMMERCIAL_{\,i}$, which equals 1 if the stolen information is commercial information, 0 otherwise, for company i.
- $SECTOR\_FIN_{i}$, which equals 1 if the sector of the breached company is the financial sector, 0 otherwise, for company i.
- $SECTOR\_RETAIL_{i}$, which equals 1 if the sector of the breached company is the retail sector, 0 otherwise, for company i.
- $SECTOR\_EDU_{\,i}$, which equals 1 if the sector of the breached company is the education sector, 0 otherwise, for company i.
- $SECTOR\_MED_{\,i}$, which equals 1 if the sector of the breached company is the medical sector, 0 otherwise, for company i.
- $BREACH\_CARD_{\,i}$, which equals 1 if the sector of the type of breach involves the theft of credit or debit card information, 0 otherwise, for company i.
- $BREACH\_HACK_{\,i}$, which equals 1 if the sector of the type of breach involves the theft of credit or debit card information, 0 otherwise, for company i.
- $BREACH\_INSIDER_{\,i}$, which equals 1 if the sector of the type of breach involves an insider's action, 0 otherwise, for company i.
- $BREACH\_PHYSICAL_{\,i}$, which equals 1 if the sector of the type of breach involves the theft of physical documents containing sensitive data, 0 otherwise, for company i.
- $BREACH\_PORTABLE_{\,i}$, which equals 1 if the sector of the type of breach involves the theft of portable devices containing sensitive data, 0 otherwise, for company i.
- $BREACH\_STATIONARY_{i}$, which equals 1 if the sector of the type of breach involves the theft of stationary devices containing sensitive data, 0 otherwise, for company i.
- $INTENSITY_{i}$ is equal to the number of days that passed from the start to the end of the breach, for company i.

- $\ln(PEOPLE\ IMPACTED)_i$ is the natural logarithm of the number of people impacted by the breach, for company i.

What is expected from this regression is an overall negative impact for all (or at least most) variables, as any breach characteristic is likely to produce a negative reaction in markets. The same is valid for $INTENSITY_i$ and $\ln(PEOPLE\ IMPACTED)_i$, as the more they grow, the worse the reaction is expected from the shareholders. The same regression was run, introducing the *Year Fixed Effect* to see how results change.

As anticipated in the previous section, the final sample used to run the regressions studying investors' and customers' reactions does not consider the outliers and is made of 177 observations. Observations start in 2009 and end in 2019, with no useful observations for 2012. The data description for the new sample can be found in Appendix C.

The next paragraph contains an explanation of the models used to test how much the relative change in net sales is impacted by cyber breaches.

3.4 Estimation of Net Sales in the quarter of the breach

Hypotheses *3* to *6* will have the following forms in their respective models, and the variables they include are explained below.

Model 1 (referring to Hypothesis 3, namely *the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity, and data from the quarter preceding the breach, including cash, assets, BV/MV, and the number of people impacted by the breach*):

$$REL\_CHANGE\_NET\_SALES_i = \alpha_i + \beta_{i1} \ln(CASH_{-1,i}) + \beta_{i2}\ (BV/MV_{-1,i}) + \beta_{i3} \ln(ASSETS_{-1,i}) + \beta_{i4}\ INTENSITY_i + \beta_{i5} \ln(PEOPLE\ IMPACTED)_i + \varepsilon_i.$$

In this model (Equation C.1), the information about the breach is considered to be embedded in the $INTENSITY_i$ and $PEOPLE\ IMPACTED_i$ variables. This is to check how the duration of a breach and the number of people affected by it impact the net sales of a company, considering also the accounting results it obtained in the quarter preceding the breach announcement. The variables that compose this model are:

- $INTENSITY_i$ is equal to the number of days that passed from the start to the end of the breach, for company i.
- $BV/MV_{-1,i}$ is the book-to-market value in the quarter preceding the breach announcement, where BV is proxied by the total value of stockholder's equity of the

companies, which is common and preferred equity and nonredeemable noncontrolling interests, for company i.

- ln ($CASH_{-1,i}$) is the natural logarithm of cash that companies owned in the quarter preceding the breach announcement, for company i.
- ln ($ASSETS_{-1,i}$) is the natural logarithm of the total amount of assets that the company had in the quarter preceding the breach announcement, which is the proxy for firm size, for company i.
- $\ln(PEOPLE\ IMPACTED)_i$ is the natural logarithm of the number of people impacted by the breach, for company i.

What is expected to find from this regression is an overall positive impact for all accounting variables. $INTENSITY_i$ will probably be slightly positive as well, as the more time a breach lasts, the more information will be leaked, but it will become obsolete as well.

Model 2 (referring to Hypothesis 4, namely *the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the type of stolen information, type of breach, company type, the number of people impacted by the breach, and the breach intensity*):

$$REL\_CHANGE\_NET\_SALES_i = \alpha_i + \beta_{i1} INTENSITY_i + \beta_{i2} \ln(PEOPLE\ IMPACTED)_i + \beta_{i3}\ INFO\_ID_i + \beta_{i4}\ INFO\_LOGIN_i + \beta_{i5}\ INFO\_HEALTH_i + \beta_{i6}\ INFO\_COMMERCIAL_i + \beta_{i7}\ SECTOR\_FIN_i + \beta_{i8}\ SECTOR\_RETAIL_i + \beta_{i9}\ SECTOR\_EDU_i + \beta_{i10}\ SECTOR\_MED_i + \beta_{i11}\ BREACH\_CARD_i + \beta_{i12}\ BREACH\_HACK_i + \beta_{i13}\ BREACH\_INSIDER_i + \beta_{i14}\ BREACH\_PHYSICAL_i + \beta_{i15}\ BREACH\_PORTABLE_i + \beta_{i16}\ BREACH\_STATIONARY_i + \varepsilon_i.$$

The model (Equation C.2) aims to explain the relative change in net sales by considering the characteristics of the breach to check for those characteristics that clients deem important when a cyber breach is announced. The variables for this model that were not previously explained are:

- $INFO\_ID_i$, which equals 1 if the stolen information is identity information, 0 otherwise, for company i.
- $INFO\_LOGIN_i$, which equals 1 if the stolen information is sensitive login information, 0 otherwise, for company i.
- $INFO\_HEALTH_i$, which equals 1 if the stolen information is health information, 0 otherwise, for company i.
- $INFO\_COMMERCIAL_i$, which equals 1 if the stolen information is commercial information, 0 otherwise, for company i.

- $SECTOR\_FIN_i$, which equals 1 if the sector of the breached company is the financial sector, 0 otherwise for company i.
- $SECTOR\_RETAIL_i$, which equals 1 if the sector of the breached company is the retail sector, 0 otherwise for company i.
- $SECTOR\_EDU_i$, which equals 1 if the sector of the breached company is the education sector, 0 otherwise for company i.
- $SECTOR\_MED_i$, which equals 1 if the sector of the breached company is the medical sector, 0 otherwise for company i.
- $BREACH\_CARD_i$, which equals 1 if the sector of the type of breach involves the theft of credit or debit card information, 0 otherwise for company i.
- $BREACH\_HACK_i$, which equals 1 if the sector of the type of breach involves the theft of credit or debit card information, 0 otherwise for company i.
- $BREACH\_INSIDER_i$, which equals 1 if the sector of the type of breach involves an insider's action, 0 otherwise for company i.
- $BREACH\_PHYSICAL_i$, which equals 1 if the sector of the type of breach involves the theft of physical documents containing sensitive data, 0 otherwise, for company i.
- $BREACH\_PORTABLE_i$, which equals 1 if the sector of the type of breach involves the theft of portable devices containing sensitive data, 0 otherwise for company i.
- $BREACH\_STATIONARY_i$, which equals 1 if the sector of the type of breach involves the theft of stationary devices containing sensitive data, 0 otherwise for company i.

What is expected from this regression is an overall negative impact for all (or at least most) variables, as any breach characteristic is likely to produce a negative reaction from customers.

Model 3 (referring to Hypothesis 5, namely *the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity, and data from the quarter preceding the breach, including cash, assets, BV/MV, and type of stolen information, type of breach, and company type, and the number of people impacted by the breach*):

$$REL\_CHANGE\_NET\_SALES_i = \alpha_i + \beta_{i1} \ln(CASH_{-1,i}) + \beta_{i2} (BV/MV_{-1,i}) + \beta_{i3} \ln(ASSETS_{-1,i}) + \beta_{i4} INTENSITY_i + \beta_{i5} \ln(PEOPLE\ IMPACTED)_i + \beta_{i6} INFO\_ID_i + \beta_{i7} INFO\_LOGIN_i + \beta_{i8} INFO\_HEALTH_i + \beta_{i9} INFO\_COMMERCIAL_i + \beta_{i10} SECTOR\_FIN_i + \beta_{i11} SECTOR\_RETAIL_i + \beta_{i12} SECTOR\_EDU_i + \beta_{i13} SECTOR\_MED_i + \beta_{i14} BREACH\_CARD_i + \beta_{i15} BREACH\_HACK_i + \beta_{i16} BREACH\_INSIDER_i + \beta_{i17} BREACH\_PHYSICAL_i + \beta_{i18} BREACH\_PORTABLE_i + \beta_{i19} BREACH\_STATIONARY_i + \varepsilon_i.$$

This model (Equation C.3) tries to explain the relative change in net sales using accounting data from the quarter preceding the breach announcement, the time that passed from a breach to its announcement, and the characteristics of the breach, therefore combining Model 1 and Model 2. The variables for Model 3 are those from Model 1 and Model 2.

Model 4 (referring to Hypothesis 6, namely *the relative change in net sales between the quarter preceding the breach announcement and the quarter of the breach announcement is determined by the breach intensity, the quartile dummy, and data from the quarter preceding the breach, including cash, assets, BV/MV, and type of stolen information, type of breach, and company type, and the number of people impacted by the breach*):

$$REL\_CHANGE\_NET\_SALES_i = \alpha_i + \beta_{i1} \ln(CASH_{-1,i}) + \beta_{i2} \, (BV/MV_{-1,i}) + \beta_{i3} \ln(ASSETS_{-1,i}) + \beta_{i4} INTENSITY_i + \beta_{i5} QUARTILE\_DUMMY_i + \beta_{i6} \ln(PEOPLE\ IMPACTED)_i + \beta_{i7}\, INFO\_ID_i + \beta_{i8}\, INFO\_LOGIN_i + \beta_{i9}\, INFO\_HEALTH_i + \beta_{i10} INFO\_COMMERCIAL_i + \beta_{i11} SECTOR\_FIN_i + \beta_{i12}\, SECTOR\_RETAIL_i + \beta_{i13}\, SECTOR\_EDU_i + \beta_{i14}\, SECTOR\_MED_i + \beta_{i15}\, BREACH\_CARD_i + \beta_{i16}\, BREACH\_HACK_i + \beta_{i17}\, BREACH\_INSIDER_i + \beta_{i18}\, BREACH\_PHYSICAL_i + \beta_{i19}\, BREACH\_PORTABLE_i + \beta_{i20}\, BREACH\_STATIONARY_i + \varepsilon_i.$$

This model (Equation C.4) tries to explain the relative change in net sales using accounting data from the quarter preceding the breach announcement, the time that passed from the breach start to its end, the characteristics of the breach, and a dummy variable that indicates if the time passed between the end of a breach and its announcement date falls in the first quartile, considering all the companies in the sample. The variables for Model 4 are the same as for Model 3, and the Quartile Dummy:

- $QUARTILE\_DUMMY_i$ equals 1 if the time passed between the end of the breach and the announcement date falls in the first quartile (considering all the companies in the sample) and 0 otherwise. This variable was computed starting from the variable $TIME_i$, defined as the number of days that separate the end of the breach and its announcement for company i.

What is expected to find from this regression is an overall positive impact for all accounting variables. $QUARTILE\_DUMMY_i$ will probably be negative, as the more time passes from the end of the breach to its announcement, the more the lost data will be useless as it becomes obsolete. Breach characteristics variables are likely to produce a contrasting reaction in customers, with different significance levels.

All four models were then run, including the *Year Fixed Effect*, to check how results change when controlling for unobserved factors that impact all companies in a given year.

4. Results

In this section, the results of the research are presented. In the first sub-section, the results of the investors' reaction will be shown, while in the second sub-section, the customers' reaction is described.

4.1 Investors' reaction

It would be logical to believe that the announcement of a cyber-attack provokes a negative market reaction, implying a decrease in the stock prices of the breached companies. Also, if the information about the breach is spread in the market before the announcement, the negative impact could start before the announcement.

With a focus on the entire data sample, including 260 observations, the results obtained show that CARs are always negative after a cyber-attack but not always significant. Most of the following event windows are used in some past papers, for example (–10; +10), (–5; +5), (–3; +3), (–5; –1), (–3; –1) were used in Arcuri et al. (2018), (–1; +1), (–1; +3), (–1; +5) were used in Hogan et al. (2023), (–3; +1) was used in Tosun (2021). I added the event windows (–10; +1), (–5; +1), (–1; +10) to allow considerations about the possibility of information leaks in the market prior to the announcement of cyber-attacks, for those event windows that were not included in the previous literature.

Table 7a
CARs and their test-statistics, computed using the Market Model

Event Window	Average CAR (%)	CAR t-statistics
(–5; –1)	-0.448	-2.298**
(–3; –1)	-0.277	-1.667**
(–10; +10)	-0.217	-0.483
(–5; +5)	-0.740	-2.141**
(–3; +3)	-0.640	-2.281**
(–1; +1)	-0.156	-0.914
(–10; +1)	-0.088	-0.284
(–5; +1)	-0.573	-2.346***
(–3; +1)	-0.402	-1.779**
(–1; +3)	-0.397	-1.692**
(–1; +5)	-0.330	-1.086
(–1; +10)	-0.286	-0.733
*p < 0.10, **p < 0.05, ***p < 0.01 (one-tailed test)		

As shown in the table above, the strongest significance is found on the event window (–5; +1), implying that maybe, towards the announcement of cyber-attacks, there could be a leakage of information about the breach in the market, impacting companies' stock prices starting before the announcement. In that event windows cumulative abnormal returns decrease on average by

0.573% for the companies that announce a cyber breach at $t = 0$. Therefore, the results are in line with the hypothesis. However, comparing the results obtained in this research with previous literature, some differences in significance and CARs averages are found. This is most likely due to possible differences in the samples used.

The same event windows were considered using the Fama and French Three Factor Model (i.e., $E(R_{it}) = r_{it} - r_{ft} = \alpha_i + \beta_{1i}(r_{mt} - r_{ft}) + \beta_{2i}(SMB_i) + \beta_{3i}(HML_i) + \varepsilon_{it}$, as mentioned in Bello, 2008) and the Carhart Model (i.e., $E(R_{it}) = r_{it} - r_{ft} = \alpha_i + \beta_{1i}(r_{mt} - r_{ft}) + \beta_{2i}(SMB_i) + \beta_{3i}(HML_i) + \beta_{4i}(MOM_i) + \varepsilon_{it}$, as mentioned in Bello, 2008) to check the robustness of CARs. The results appear to be quite robust, even if some event windows have different significance levels. The CARs computed with these models were downloaded from CRSP as well. The table with the two alternative models is presented below:

Table 7b

CARs and their test-statistics, computed with the Fama-French Three Factor Model and the Carhart Model, respectively

Fama-French Three Factor Model			**Carhart Model**		
Event Window	**Average CAR (%)**	**CAR t-statistics**	**Event Window**	**Average CAR (%)**	**CAR t-statistics**
(–5; –1)	-0.312	-1.576*	(–5; –1)	-0.257	-1.300*
(–3; –1)	-0.289	-1.734**	(–3; –1)	-0.252	-1.514**
(–10; +10)	0.011	0.026	(–10; +10)	-0.037	-0.087
(–5; +5)	-0.614	-1.833**	(–5; +5)	-0.618	-1.869**
(–3; +3)	-0.637	-2.441***	(–3; +3)	-0.607	-2.367***
(–1; +1)	-0.107	-0.651	(–1; +1)	-0.057	-0.346
(–10; +1)	0.093	0.312	(–10; +1)	0.146	0.474
(–5; +1)	-0.369	-1.530*	(–5; +1)	-0.293	-1.226
(–3; +1)	-0.347	-1.574**	(–3; +1)	-0.290	-1.323*
(–1; +3)	-0.403	-1.913**	(–1; +3)	-0.380	-1.817**
(–1; +5)	-0.370	-1.298*	(–1; +5)	-0.406	-1.426*
(–1; +10)	-0.220	-0.590	(–1; +10)	-0.275	-0.742
*p < 0.10, **p < 0.05, ***p < 0.01 (one-tailed test)					

Therefore, the market perceives data breaches as a negative event for the impacted companies, reflecting this information in a decrease in stock prices. This is a robust finding across all models used in terms of significance levels, while CARs coefficients slightly change.

Cumulative abnormal returns are supposed to be influenced by several factors. As anticipated in the previous section, the regression described by Equation I.1 tries to explain the cumulative abnormal returns through information about cyber-attacks, using 177 observations that exclude outliers. It is expected that some kinds of sectors, as well as some kinds of stolen information or breaches, will have more negative impacts with respect to others. Looking at the literature, it is more likely that the sectors with the worst impacts on their stock prices are retail (Hogan et al., 2023) and financial (Arcuri et al., 2018) sectors.

The results show an overall negative impact of the variables on CARs, except for the theft of portable devices, which is, however, very close to zero. Cumulative abnormal returns of breached companies are not influenced by the length of the attack or by the number of people impacted, which could also be due to the non-disclosure of these details. Therefore, CARs of breached companies appear to be impacted by the breach itself and not by its magnitude. The analysis reveals significance in the other variables (i.e., type of attack, type of stolen information, companies' sector) instead. For each characteristic, a variable was dropped to avoid multicollinearity; for the type of attack, the "Unintended Disclosure" was dropped; for the type of stolen information, the "Government" variable was dropped; for the sectors, the "Other Sector" was dropped. The interpretation of the characteristic variables should be made with respect to the variable dropped for each category. Sensitive login information results significantly at a 1% significance level, and it decreases CARs by 5.4% with respect to the sector "Government" information, considering a one-tailed test and no fixed effect. When year-fixed effects are added, the same variable remains significant at a 1% level, with a negative coefficient of 6.1%. Commercial information is negative and significant as well, at a 10% level, both with and without year-fixed effects. As regards the sector, financial, education, and retail sectors are significant at a 10% level, while the medical one is significant at a 1% level, but when the year fixed effects are added, the medical and education sectors lose their significance. All the sectors reduce CARs between 1% and 6% with respect to the "Other" sector. The types of attacks that are significant are very few. The attacks that involve hackers significantly reduce CARs, with respect to unintended disclosures, with a 5% and 1% significance and coefficients of -1.6% or -2% for models without and with year-fixed effects, respectively. The theft of stationary devices is significant at a 10% significance level, with a significance of -5%. What can be derived from the results is that the market does not negatively perceive the length or the number of people impacted by the attack, but it reacts particularly negatively when personal or financial information is stolen (e.g., sensitive login information; financial, educational, and medical sector), and when it is due to external criminals (e.g., hacking; physical documents or stationary devices theft). Therefore, the results support what was found by Hogan et al. (2023) in terms of the type of information stolen and type of attack, and they are also coherent with the findings by Arcuri et al. (2018) in terms of sector. After getting rid of the variables with few observations (namely, $INFO\_HEALTH_i$, $INFO\_COMMERCIAL_i$, $SECTOR\_EDU_i$, $BREACH\_CARD_i$, $BREACH\_PHYSICAL_i$, $BREACH\_STATIONARY_i$), the results show similar findings, both in terms of coefficients and significance levels, which shows the robustness of the results. For both cases, the R^2 increases when year-fixed effects are added to the models,

showing that, controlling for shocks that all companies face enhances the ability of the model to explain the variance in CARs.

Below detailed Table 8 with all variables and their significance is presented:

Table 8
Impact of Intensity, People Impacted, Seector, type of stolen information, breach type on CARs

	All variables are included		Variables with few observations were dropped	
Variables	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
$INTENSITY_i$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$ln(PEOPLE\ IMPACTED)_i$	0.000 (0.002)	0.001 (0.002)	0.000 (0.002)	0.000 (0.002)
$INFO\_ID_i$	-0.018 (0.019)	-0.025 (0.019)	-0.018 (0.019)	-0.010 (0.020)
$INFO\_LOGIN_i$	-0.054 (0.023)***	-0.061 (0.024)***	-0.054 (0.023)***	-0.046 (0.025)***
$INFO\_HEALTH_i$	-0.031 (0.043)	-0.044 (0.043)	–	–
$INFO\_COMMERCIAL_i$	-0.060 (0.041)*	-0.059 (0.042)*	–	–
$SECTOR\_FIN_i$	-0.013 (0.008)*	-0.014 (0.009)*	-0.013 (0.008)**	-0.013 (0.008)*
$SECTOR\_RETAIL_i$	-0.016 (0.011)*	-0.016 (0.012)*	-0.016 (0.011)*	-0.014 (0.011)*
$SECTOR\_EDU_i$	-0.057 (0.038)*	-0.013 (0.052)	–	–
$SECTOR\_MED_i$	0.005 (0.009)***	0.005 (0.013)	0.005 (0.012)	0.006 (0.013)
$BREACH\_CARD_i$	-0.025 (0.039)	-0.022 (0.039)	–	–
$BREACH\_HACK_i$	-0.016 (0.008)**	-0.020 (0.008)***	-0.016 (0.008)**	-0.016 (0.008)**
$BREACH\_INSIDER_i$	-0.008 (0.012)	-0.007 (0.012)	-0.008 (0.012)	-0.008 (0.012)
$BREACH\_PHYSICAL_i$	-0.015 (0.037)	-0.021 (0.037)	–	–
$BREACH\_PORTABLE_i$	0.019 (0.020)	0.012 (0.020)	0.019 (0.020)	0.015 (0.020)
$BREACH\_STATIONARY_i$	-0.047 (0.037)	-0.050 (0.038)*	–	–
Constant	0.029 (0.020)*	-0.009 (0.042)	0.029 (0.020)*	0.023 (0.043)*
No. of observations	177	177	171	171
Year Fixed Effect	No	Yes	No	Yes
R-squared	0.119	0.182	0.101	0.146

*p < 0.10, **p < 0.05, ***p < 0.01 (one-tailed test)
Standard errors are reported in parenthesis

Check the appendix for the variables' abbreviations explanation.

4.2 Customers' reaction

As regards customers' perception of the cyber-attack, four other regressions were run to analyze how the percentage change in net sales between the quarter preceding the attack announcement and the quarter of the attack announcement are influenced by a number of variables, both regarding companies' books data and data about the cyber breach. More specifically, the regression takes into consideration the intensity of the breach, measured as the number of days the breach lasted, a time dummy indicating if the firm announced the breach in a number of days that fall in the first quartile (the Quartile Dummy variable equals 1 if the time between the end of the breach and the announcement date falls in the first quartile, 0 otherwise. The first quartile value is 23, so firms that fall in the first quartile announced the breach within 23 days from its end), and some data from the quarter before the breach announcement (i.e., Cash, Assets, BV/MV), as in Hogan et al. (2023). As anticipated in the methodology section, I will consider four models:

1. The model described by Equation C.1
2. The model described by Equation C.2
3. The model described by Equation C.3
4. The model described by Equation C.4

Model 1 explains the relative change in net sales between the quarter preceding the attack announcement and the quarter of the attack announcement using Cash, BV/MV, Assets, the intensity of the breach, and the number of people affected by the breach. The results show no significance when there is no year-fixed effect; however, when they are added, the variable BV/MV becomes negative and significant at a 10% level. However, this is most likely unrelated to the breach announcement unless there was a leakage of information in the market about the event. In this case, the negative coefficient could be connected to a loss of confidence on the investors' side, which is therefore supported by the findings about CARs. Results for Model 1 are the same even after eliminating the variables with few observations (i.e., $INFO\_HEALTH_i$, $INFO\_COMMERCIAL_i$, $SECTOR\_EDU_i$, $BREACH\_CARD_i$, $BREACH\_PHYSICAL_i$, $BREACH\_STATIONARY_i$), as it is not impacted by this change. Model 2 explains the relative change in net sales using data about the breach itself, including the type of stolen data, the sector the breached company belongs to, and the type of breach, as well as the intensity, the number of people impacted, and the intensity of the breach. While intensity is always very close to zero, the number of people impacted is significant at a 10% level, considering year-fixed effects. However, this result does not appear very robust, as it is the only significance this variable has

in all models. For the other variables, the results appear to be more significant when year-fixed effects are added, showing positive and significant impacts of the type of stolen information for identity, login, and medical information, respectively, at 1%, 1%, and 10% levels. Their coefficients are respectively 86.2%, 75.1%, and 79.3%, compared to "Government information". In the first two cases, this could be due to the fact that, after a breach announcement, a company gets more visibility, which could paradoxically increase its sales. In the case of medical information, the significance could be little reliable, as there is only one observation for that type of information, which is indeed eliminated in Table 9b. For all other variables, robustness is shown after getting rid of the variables with few observations. After controlling for year-fixed effects, breaches in the medical and financial sectors provoke a negative impact on the relative change in net sales, with coefficients of 20.7% and 24.4%, compared to the "Other" sector, probably because clients fear that the stolen data involves highly sensitive personal information. The robustness of the results is confirmed in Table 9b. Model 3 combines the explanatory variables of both Model 1 and Model 2, showing significance for the variables that were significant in Model 2 at the same levels, except for the financial information, which is now significant at a 10% level. Therefore, compared to the "Government information", a breach that happens in the financial sector reduces the net sales relative change by 20%, with a significance of 10%. Model 4 is the same as Model 3, but it also adds the quartile dummy. The significant variables are the same as for Model 3, plus the quartile dummy and retail sector. These two variables are respectively significant at 1% and 10% levels in the model considering the year fixed effects, and their coefficients are -21.3% and -19.1%. Therefore, if a company announces the breach within 23 days from its end, then its relative change in net sales will decrease by 21.3%, indicating that the sooner the company announces the breach, the worse the clients' reaction; this could indicate that as time goes by, the clients perceive the stolen information as too obsolete to be used against the company or them. For a breach that involves a company in the retail sector, the relative change in net sales for that company will decrease by 19.1%, compared to a company in the "Other" (non-specified) sector, probably because customers fear the risk of strictly personal information theft. The results for Model 4 are also robust when the variables with few observations are eliminated.

Overall, the results indicate that when the stolen information does not involve highly sensitive data, the relative change in net sales is likely to be increased thanks to the involuntary public attention towards the breached company. Customers appear to be more worried when the breached companies operate in sectors that are likely to deal with strictly personal information.

They do not mind whether the data were involuntarily leaked, or external criminals stole them, probably because they could possibly become public anyway. A good accounting situation appears to be of little to no help in these cases for customers' interests. Also, as anticipated, customers do not seem to care too much about the number of people impacted (since, despite being significant at a 10% level in Model 2, no robustness about this result is shown) or the breach intensity; this could be due either to the non-disclosure of these details or to the fact that customers simply care about the fact that data was lost anyway, no matter for how many days or how many people were impacted. These findings appear to be quite in line with those by Hogan et al. (2023), supporting the fact that the accounting characteristics of a firm do not significantly improve the relative change in net sales, which is worsened by some breach characteristics.

The results obtained from the regressions are presented in Table 9a (all variables are included) and 9b (variables with few observations were dropped) below:

Table 9a

Impact of ln (Cash), BV/MV, ln (Asstets), ln (Net Sales), Time, Quartile Dummy, Seector, type of stolen information, breach type on ln (Net Sales) in the breach quarter – all variables are included

Variables	Model 1 Coefficients	Model 1 Coefficients	Model 2 Coefficients	Model 2 Coefficients	Model 3 Coefficients	Model 3 Coefficients	Model 4 Coefficients	Model 4 Coefficients
$ln(CASH_{-1,i})$	-0.011 (0.030)	0.013 (0.029)			-0.011 (0.034)	0.031 (0.033)	-0.007 (0.034)	0.038 (0.032)
$BV/MV_{-1,i}$	-0.110 (0.094)	-0.140 (0.093)*			-0.122 (0.114)	-0.062 (0.112)	-0.125 (0.113)	-0.063 (0.110)
$ln(ASSETS_{-1,i})$	0.003 (0.035)	-0.018 (0.033)			-0.002 (0.040)	-0.036 (0.037)	-0.006 (0.040)	-0.039 (0.037)
$INTENSITY_i$	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.000 (0.001)
$QUARTILE\_DUMMY_i$							-0.157 (0.093)**	-0.213 (0.087)***
$ln(PEOPLE\ IMPACTED)_i$	0.026 (0.024)	0.029 (0.023)	0.030 (0.025)	0.030 (0.023)*	0.028 (0.026)	0.027 (0.024)	0.024 (0.026)	0.020 (0.023)
$INFO\_ID_i$			0.708 (0.240)***	0.862 (0.230)***	0.723 (0.242)***	0.875 (0.231)***	0.708 (0.240)***	0.851 (0.228)***
$INFO\_LOGIN_i$			0.645 (0.302)***	0.751 (0.287)***	0.655 (0.304)**	0.782 (0.289)***	0.672 (0.302)**	0.790 (0.284)***
$INFO\_HEALTH_i$			0.731 (0.550)*	0.793 (0.507)*	0.756 (0.556)*	0.796 (0.512)*	0.878 (0.558)*	0.966 (0.508)**
$INFO\_COMMERCIAL_i$			0.492 (0.526)	0.587 (0.492)	0.509 (0.528)	0.566 (0.494)	0.440 (0.526)	0.487 (0.487)
$SECTOR\_FIN_i$			-0.119 (0.108)	-0.244 (0.104)***	-0.027 (0.130)	-0.200 (0.128)*	-0.040 (0.129)	-0.255 (0.126)**
$SECTOR\_RETAIL_i$			-0.040 (0.136)	-0.159 (0.137)	-0.033 (0.138)	-0.166 (0.139)	-0.064 (0.138)	-0.191 (0.137)*
$SECTOR\_EDU_i$			0.014 (0.484)	-0.126 (0.616)	0.088 (0.489)	-0.009 (0.629)	-0.008 (0.489)	-0.253 (0.627)
$SECTOR\_MED_i$			-0.089 (0.159)	-0.207 (0.148)*	-0.043 (0.169)	-0.233 (0.159)*	-0.100 (0.172)	-0.302 (0.159)**
$BREACH\_CARD_i$			-0.055 (0.496)	-0.017 (0.456)	-0.126 (0.508)	0.025 (0.471)	0.008 (0.510)	0.189 (0.468)
$BREACH\_HACK_i$			0.007 (0.108)	-0.015 (0.100)	0.003 (0.114)	-0.042 (0.106)	-0.036 (0.115)	-0.093 (0.107)
$BREACH\_INSIDER_i$			0.023 (0.151)	-0.028 (0.141)	0.069 (0.157)	0.007 (0.147)	0.018 (0.159)	-0.068 (0.148)
$BREACH\_PHYSICAL_i$			-0.223 (0.482)	-0.257 (0.443)	-0.247 (0.489)	-0.237 (0.448)	-0.335 (0.489)	-0.366 (0.444)
$BREACH\_PORTABLE_i$			-0.034 (0.255)	-0.087 (0.235)	-0.010 (0.257)	-0.070 (0.236)	-0.038 (0.256)	-0.097 (0.233)
$BREACH\_STATIONARY_i$			-0.031 (0.482)	-0.089 (0.444)	-0.036 (0.495)	-0.053 (0.459)	-0.049 (0.494)	-0.187 (0.455)
Constant	0.090 (0.0186)	-0.067 (0.466)	-0.656 (0.252)***	-0.670 (0.492)*	-0.549 (0.333)**	-0.597 (0.520)	-0.427 (0.338)	-0.333 (0.523)
No. of observations	177	177	177	177	177	177	177	177
Year Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.027	0.193	0.073	0.269	0.084	0.279	0.100	0.307

*p < 0.10, **p < 0.05, ***p < 0.01 (one-tailed test)
Standard errors are reported in parenthesis

Check the appendix for the variables' abbreviations explanation.

Table 9b

Impact of ln (Cash), BV/MV, ln (Asstets), ln (Net Sales), Time, Quartile Dummy, Seector, type of stolen information, breach type on ln (Net Sales) in the breach quarter – variables with few observations were dropped

Variables	Model 1 Coefficients	Model 1 Coefficients	Model 2 Coefficients	Model 2 Coefficients	Model 3 Coefficients	Model 3 Coefficients	Model 4 Coefficients	Model 4 Coefficients
$ln(CASH_{-1,i})$	-0.013 (0.032)	0.013 (0.031)			-0.012 (0.034)	0.031 (0.033)	-0.007 (0.034)	0.038 (0.032)
$BV/MV_{-1,i}$	-0.116 (0.097)	-0.149 (0.095)*			-0.122 (0.114)	-0.062 (0.122)	-0.125 (0.113)	-0.063 (0.110)
$ln(ASSETS_{-1,i})$	0.006 (0.036)	-0.017 (0.034)			-0.002 (0.040)	-0.036 (0.037)	-0.006 (0.040)	-0.039 (0.037)
$INTENSITY_i$	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.000 (0.001)
$QUARTILE\_DUMMY_i$							-0.157 (0.093)**	-0.213 (0.087)***
$ln(PEOPLE\ IMPACTED)_i$	0.028 (0.025)	0.029 (0.023)	0.030 (0.025)	0.030 (0.023)*	0.028 (0.001)	0.027 (0.024)	0.024 (0.026)	0.020 (0.023)
$INFO\_ID_i$			0.708 (0.240)***	0.862 (0.230)***	0.723 (0.242)***	0.875 (0.231)***	0.708 (0.240)***	0.851 (0.228)***
$INFO\_LOGIN_i$			0.645 (0.302)***	0.750 (0.287)***	0.656 (0.304)**	0.782 (0.289)***	0.672 (0.302)**	0.790 (0.284)***
$INFO\_HEALTH_i$			–	–	–	–	–	–
$INFO\_COMMERCIAL_i$			–	–	–	–	–	–
$SECTOR\_FIN_i$			-0.119 (0.108)	-0.244 (0.104)***	-0.027 (0.130)	-0.200 (0.128)*	-0.040 (0.129)	-0.255 (0.126)**
$SECTOR\_RETAIL_i$			-0.040 (0.136)	-0.159 (0.137)	-0.033 (0.138)	-0.166 (0.139)	-0.064 (0.138)	-0.191 (0.137)*
$SECTOR\_EDU_i$			–	–	–	–	–	–
$SECTOR\_MED_i$			-0.089 (0.159)	-0.207 (0.148)*	-0.043 (0.169)	-0.233 (0.159)*	-0.100 (0.172)	-0.302 (0.159)**
$BREACH\_CARD_i$			–	–	–	–	–	–
$BREACH\_HACK_i$			0.007 (0.108)	-0.015 (0.100)	0.003 (0.114)	-0.042 (0.106)	-0.036 (0.115)	-0.093 (0.107)
$BREACH\_INSIDER_i$			0.023 (0.151)	-0.028 (0.141)	0.069 (0.157)	0.007 (0.147)	0.018 (0.159)	-0.068 (0.148)
$BREACH\_PHYSICAL_i$			–	–	–	–	–	–
$BREACH\_PORTABLE_i$			-0.034 (0.255)	-0.087 (0.235)	-0.010 (0.257)	-0.070 (0.236)	-0.038 (0.256)	-0.097 (0.233)
$BREACH\_STATIONARY_i$			–	–	–	–	–	–
Constant	0.090 (0.191)	0.063 (0.473)	-0.656 (0.252)***	-0.669 (0.492)*	-0.549 (0.333)**	-0.597 (0.520)	-0.427 (0.338)	-0.333 (0.523)
No. of observations	171	171	171	171	171	171	171	171
Year Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.028	0.195	0.072	0.268	0.082	0.278	0.098	0.306

*p < 0.10, **p < 0.05, ***p < 0.01 (one-tailed test)
Standard errors are reported in parenthesis

Check the appendix for the variables' abbreviations explanation.

5. Conclusion

This research provides important insights into the cybersecurity topic. The first finding regards cybercriminals' preferences in terms of data. Cybercriminals prefer to steal data regarding personal information, and most of the breaches are due to actors that are external to the breached firms. Both of these findings are in line with the findings by Hogan et al. (2023).

The paragraphs that follow wrap up the more complex findings about investors' and customers' reactions.

5.1 Investors' reaction

The results section highlights that what was hypothesized in section 3 is confirmed by the empirical findings of this research. Compliant with the previous literature, results confirm that cumulative abnormal returns that follow the announcement of a data breach are negative and mainly significant; the difference in significance levels and significance of event windows might be explained by the differences in samples and general availability of data (i.e., data availability and origin about stock prices of companies listed in the PRC database).

The explanation of cumulative abnormal returns through the characteristics of the attack included some original clues, while some variables were taken from previous literature (i.e., the natural logarithm of impacted people, as in Hogan et al., 2023). Results show that almost all variables have a negative impact on CARs, and some of them are significant at different levels. Some types of stolen information (e.g., login and commercial information) significantly impact CARs, implying that that kind of information is probably perceived as more damaging to lose than some others by the market, just like for some types of data breaches (e.g., hacking), or some sectors (e.g., financial and retail sectors).

5.2 Customers' reaction

As regards customers' reactions to data breaches, the significance variables depend on their category. For most of the models, accounting data is not significant, while some breach characteristics, such as type of stolen information and sector, provide insightful results. Apparently, when the type of information that is stolen does not involve highly sensitive data, it can increase the relative change in net sales, thanks to public exposure of the company after the announcement of the breach. However, in the case of specific sectors involved, such as financial, medical, and retail, the relative change in net sales is negatively impacted because customers perceive an increased risk of highly confidential information loss. Furthermore, if

the breach is announced within 23 days from its end, then the relative change in net sales is negatively impacted, as the information that was stolen can still be considered harmful, as it is not obsolete. As anticipated in the previous section, with respect to the results obtained by Hogan et al. (2023), the results obtained in this paper are quite similar despite different levels of significance, which is most likely due to differences in the datasets used.

To conclude this research, it should be underlined that data breaches represent harm for companies that are hit in monetary and reputational terms; this implies a negative effect on companies' value. It becomes, therefore, fundamental in this scenario that companies are able to prevent and/or properly react to cyber incidents. Interesting solutions for this are proposed in the paper by Aldasoro et al. (2020), who found that to face cyber-attacks, companies usually work on three main aspects. First, they all try to improve the time it takes them to recover from a breach from an operational point of view. They do so by investing more in cyber security and aligning their business goals to the implementation of measures that help protect their data. Also, after a breach, a company takes on a mindset that considers breaches as possible, which leads to keeping the attention high, at least in terms of internal workers. In some sectors, guidelines are being released to give companies guidance to follow in case they are targeted by cybercriminals. Another aspect that Alsadoro et al. (2020) suggest is for companies to consider cyber insurance; they can help cover the losses incurred by companies after a cyber-attack, excluding the losses incurred in tail events. There is also a new possibility of initiatives organized by the private sector, intending to support companies in the prevention of cyber-attacks and support in the recovery from them. They offer alternatives to the reliance on "third party dependencies" (i.e., cloud services) that could be interrupted as a consequence of data breaches.

The results obtained through this research could be of particular interest to US insurance companies that offer policies against cyber-attacks, US regulators, and US policymakers, as well as companies that are concerned about their cybersecurity safety. However, this research has some clear limitations. First of all, the sample considered is quite small due to a lack of data availability; furthermore, as anticipated in the introduction, in 2018, the GDPR was implemented in the European Economic Area, having impacts on all the companies that trade with the European ones. The latter topic mentioned was not considered in this research, though. Therefore, a potential starting point for future research is to expand the data breach sample, including companies beyond the US, and conduct similar research to compare the findings. Additionally, examining the reactions of other stakeholders, such as suppliers and employees,

would yield valuable insights for companies and interested parties. It would also be beneficial to divide the sample into two segments, before and after the implementation of GDPR, to assess any potential impact this regulation may have on shareholders' reactions to data breach announcements. Taking into account European companies would likely provide further insights into the effects of GDPR regulation.

Bibliography

Arcuri M. C., Brogi M., Gandolfi G. (2018). The Effect of Cyber-Attacks on Stock Returns. *Corporate Ownership & Control, Volume 15, Issue 2.*

Bello Z. Y. (2008). A Statistical Comparison of The CAPM to the Fama-French Three Factor Model and the Carhart's Model. *Global Journal of Finance and Banking Issues, Vol. 2, No. 2.*

Cummins J. D., Lewis C. M., Wei R. (2006). The Market Value Impact of Operational Loss Events for US Banks and Insurers. *Journal of Banking & Finance, Vol. 30, pp. 2605–2634.*

Ferguson K. (2021, April). *Definition. Denial of service attack.* TechTarget. Retrieved January 4, 2024, from https://www.techtarget.com

Fu F. (2009). Idiosyncratic Risk and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics, Vol. 91, No. 1, pp. 24-37.*

Gillis A. S. (2024, February). *Definition. Phishing.* TechTarget. Retrieved January 4, 2024, from https://www.techtarget.com

Hendricks K. B., Jacobs B. W., Singhal V. R. (2019). Stock Market Reaction to Supply Chain Disruptions From the 2011 Great East Japan Earthquake. *Manufacturing and Service Operations Management, 22(4).*

Hernandez J. R. (2023). What is the Actual Cost of Cybercrime? Evolve Security. https://www.evolvesecurity.com

Hogan K. M., Olson G. T., Mills J. D., Zaleski P. A. (2023). An Analysis of Cyber Breaches and Effects on Shareholders Wealth. *International Journal of the Economics of Business, 30:1, 51-78.*

Jarrow R. A. (2008). Operational Risk. *Journal of Banking & Finance, Vol. 32, Issue 5, pp. 870–879.*

Juma'h A. H., Alnsour Y. (2019). The Effect of Data Breaches on Company Performance. *International Journal of Accounting and Information Management, Vol. 28, No. 2, 2020, pp.* 275–301

Kamiya S., Kang J., Kim J., Milidonis A. (2021). Risk Management, Firm Reputation, and the Impact of Successful Cyberattacks on Target Firms. *Journal of Financial Economics 139, pp. 719–749.*

Leroy I. (2022). The Relationship Between Cyber-Attacks and Dynamics of Company Stock: the Role of Reputation Management. *International Journal of Electronic Security and Digital Forensics, Vol. 14, No. 4.*

March J. G. and Shapira Z. (1987). Managerial Perspectives on Risk and Risk Taking. *Management Science, Vol. 33, No. 11.*

Petrosyan A. (2024, September 11). *Average Cost of a Data Breach Worldwide from May 2020 to February 2024, by Industry (in million US Dollars)*. Statista. https://www.statista.com

Petrosyan A. (2024, October 10). *Average Cost of a Data Breach in the United States from 2006 to 2024 (in million US Dollars)*. Statista. https://www.statista.com

Privacy Rights Clearinghouse, PRC (n.d.). *Overview of Data Breach Chronology Database.* Retrieved December 27, 2024, from https://public.tableau.com/

Sorescu A., Warren N. L., Ertekin L. (2017). Event Study in the Marketing Literature: an Overview. *Journal of the Academy of Marketing Science, 45:186–207.*

Tidy J. (2022, March 20). *Anonymous: How Hackers are Trying to Undermine Putin.* BBC. https://www.bbc.com/news/technology-60784526

Tosun O. K., (2021). Cyber-attacks and Stock Market Activity. *International Review of Financial Analysis, Vol. 76.*

Yasar K. and Cobb. M. (2022, April). *Definition. Man-in-the-middle attack.* TechTarget. Retrieved January 4, 2024, from https://www.techtarget.com

Yasar K. and Lutkevich B. (2024, July). *Definition. What is malware? Prevention, detection and how attacks work.* TechTarget. Retrieved January 4, 2024, from https://www.techtarget.com

Appendices

Appendix A: identification of outliers for breaches' numerical characteristics

The outliers for breaches' numerical characteristics (e.g., People Impacted, Time, and Intensity) were identified by calculating the first and third quartile for each variable. Then, the difference between the two (i.e., the Interquartile Range, "IQR") was computed for all variables. The result obtained after subtracting the first quartile from the third was then used to compute the so-called "upper limit", which is the maximum acceptable value for the considered variable. A "lower limit" was not computed, as most of the observations included low numbers, therefore, imposing a lower limit would reduce the sample too much; for the same reason, the upper limit was computed using the third quartile. The formula used to compute the upper limit for each variable is

$$Upper\ Limit = Third\ Quartile + (1.5 \cdot IQR).$$

The observations that included values higher than the upper limit were deleted from the sample, as they were considered outliers.

Appendix B: data description for the 177 observations used to run the regressions

Table 1

Frequency distribution by Industry of 177 Data Breaches, announced by US firms

Industry	Frequency	%
Financial	41	23,2
Retail	15	8,5
Education	1	0,6
Medical	11	6,2
Other	109	61,6
Total	177	100

Table 2

Frequency distribution by Breach Types of 177 Data Breaches, announced by US firms

Breach Type	Frequency	%
Card	1	0,6
Malware	122	68,9
Insider	15	8,5
Physical	1	0,6
Portable	4	2,3
Stationary	1	0,6
Unintended	33	18,6
Total	177	100

Table 3

Frequency distribution by Stolen Information of 177 Data Breaches, announced by US firms

Information	Frequency	%
Identifier	164	92,7
Login	7	4,0
Government	4	2,3
Health	1	0,6
Commercial	1	0,6
Total	177	100

Table 4

Frequency distribution by Year of 177 Data Breaches, announced by US firms

Year	Frequency	%
2009	1	0,6
2010	1	0,6
2011	2	1,1
2012	0	0,0
2013	3	1,7
2014	2	1,1
2015	7	4,0
2016	47	26,6
2017	38	21,5
2018	39	22,0
2019	37	20,9
Total	177	100

Table A.1
Variables Short Names, Categories, and Descriptions

Variable Short Name	Variable Category	Variable Descritption
$INFO\_ID_i$	Type of Stolen Information	Theft of Identity Information for company i
$INFO\_LOGIN_i$	Type of Stolen Information	Theft of Sensitive Login Information for company i
$INFO\_HEALTH_i$	Type of Stolen Information	Theft of Health Information for company i
$INFO\_COMMERCIAL_i$	Type of Stolen Information	Theft of Commercial Information for company i
$SECTOR\_FIN_i$	Breached Companies' Sector	Breached Company is in the Financial Sector for company i
$SECTOR\_RETAIL_i$	Breached Companies' Sector	Breached Company is in the Retail Sector for company i
$SECTOR\_EDU_i$	Breached Companies' Sector	Breached Company is in the Education Sector for company i
$SECTOR\_MED_i$	Breached Companies' Sector	Breached Company is in the Medical Sector for company i
$BREACH\_CARD_i$	Type of Breach	Breach involves Credit/Debit Cards Frauds for company i
$BREACH\_HACK_i$	Type of Breach	Breach involves Hackers/Malwares Attacks for company i
$BREACH\_INSIDER_i$	Type of Breach	Breach involves Insiders' Actions for company i
$BREACH\_PHYSICAL_i$	Type of Breach	Breach involves Physical Documents Theft for company i
$BREACH\_PORTABLE_i$	Type of Breach	Breach involves Portable Devices Theft for company i
$BREACH\_STATIONARY_i$	Type of Breach	Breach involves Stationary Devices Theft for company i
$INTENSITY_i$	Numerical Variable	Number of Days between the Start and the End of the Breach for company i
$ln(PEOPLE\ IMPACTED)_i$	Numerical Variable	Natural Logarithm of People Impacted by the Breach for company i
$ACT\_DIFF\_NET\_SALES_i$	Accounting Information	Percentage Change in Net Sales between the Quarter preceeding the Breach Announcement and the Quarter of the Breach Announcement for company i
$ln(CASH_{-1,i})$	Accounting Information	Natural Logarithm of Breached Companies' Cash in the Quarter preceding the Breach Announcement for company i
$ln(ASSETS_{-1,i})$	Accounting Information	Natural Logarithm of Breached Companies' Assets in the Quarter preceding the Breach Announcement for company i
$BV/MV_{-1,i}$	Accounting Information	Breached Companies' Book-to-Market Value in the Quarter preceding the Breach Announcement for company i
$QUARTILE\_DUMMY_i$	Dummy Variable	Belonging of company i to the First Quartile of Companies that Announced the Breach Announcement after it ended