Norwegian School of Economics
Bergen, Fall 2020

The Stock Market Effect of Cybercriminals

An empirical study of the price effects on US listed companies targeted by a data breach

Håkon Høviskeland Berg and Simen Eide Hansen

Supervisors: Associate Professores Steffen Juranek and Carsten Bienz

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

The Stock Market Effect of Cybercriminals

An empirical study of the price effects on US listed companies targeted by a data breach

Håkon Høviskeland Berg and Simen Eide Hansen

Bergen, Fall 2020

Abstract

This study investigates the effect of a data breach with more than 30 000 records stolen on publicly US listed companies' share price. Utilizing the market model, we examine abnormal returns after an announcement of a data breach in the period between 2010 and 2019. Further, the study focuses exclusively on data breaches that either was officially confirmed by the targeted company through a press release, statements to the media or confirmed through independent media reports. We find a negative and statistically significant average reduction in the share price on the day of and following the announcement of a data breach. The cumulative effect of a data breach on the share price stabilizes at day six in the event window after the announcement of the data breach. Our findings are consistent over the analyzed event windows, indicating a negative abnormal return following a data breach. Furthermore, we find a considerable variance in the reduction in share price within the sample. Hence, we are looking closer into the heterogeneity of the data breaches. First, we investigate the differences between industries and find that the finance industry experiences the most severe decline. Secondly, in line with the increased media attention we explore the average market reaction of a data breach in the two periods 2010 to 2014 and 2015 to 2019. Our findings indicate a greater reduction in the market value in the period 2015 to 2019. Furthermore, we run a regression that accounts for firm-specific traits and variables that attempt to capture the individual data breaches' characteristics. The regression finds that the data sensitivity, number of records stolen, customer segment and firm size influence the market reaction. Lastly, a data breach can have large consequences for the management team as job losses are relatively common.

Preface

This thesis is written as a part of our Master of Science in Economics and Business Administration at the Norwegian School of Economics, NHH, and it marks the end of five challenging and educational years.

First and foremost, we would like to sincerely thank our supervisors, Steffen Juranek and Carsten Bienz, for their support, expertise on the topic, and guidance throughout the entire writing process. We would also like to thank our supervisors for the idea of conducting an event study to examine the stock market effect of cybercriminals. Last but not least, we would like to thank our family and friends for providing helpful comments and support throughout the writing of our thesis.

Bergen, December 2020

Håkon H. Berg

Håkon Høviskeland Berg

Simen E. Hansen

Simen Eide Hansen

Contents

List of figures

List of tables

1. Introduction

In the last 20 years, there has been a development in the general economy where businesses of all sizes and origins have become increasingly reliant on digital data, technology, and having a high workforce mobility. It has led to radical changes in the way companies conduct business and the business landscape, consequently making companies more dependent on data- and IT security to keep their customer and internal information safe from unauthorized individuals. At the same time, the threat of cybercrime has grown exponentially with the growth in technology as the adoption of new technology has given cybercriminals more attack surfaces such as smartphones and tablets. Due to this, nearly every company has a relationship with cybersecurity, which is defined as the company's measures to protect digital data from third-party attacks through the internet. Companies need to emphasize IT- and data security as they otherwise can be the target of a data breach, which can have a substantial influence on the firm, its customers, and regulators. Also, it can potentially result in a diminished reputation and customer trust, consequently increasing the abnormal customer turnover and reducing the overall firm value.

Although data breaches and cybersecurity gradually have become a significant concern to companies and regulators, the amount of literature is relatively scarce compared to other topics. Intrigued by the increasing relevance of IT security and data breaches, especially in a period influenced by the vast ripple effects from the Covid-19 pandemic, we want to contribute to the existing literature by examining a possible stock market effect from data breaches. The thesis investigates data breaches in the more recent period of 2010 - 2019. Accordingly, we want to answer the research question:

"Does negative price effects occur in relation to US listed firms being the target of a data breach?"

This thesis aims to analyze the short-term impact of a data breach on the targeted company`s market value. The purpose of this analysis is to investigate whether there is a price effect resulting from a company being the target of a data breach. According to the semi-strong form of the Efficient Market Hypothesis, the stock price will reflect all publicly available information about a company at a given point in time. Accordingly, one can analyze the reaction in the price of a stock relative to a given event by measuring the impact on the firm's stock price. This thesis utilizes the event study methodology to investigate the market reaction by studying the abnormal returns around the announcement date of a data breach.

Our empirical results suggest that a data breach leads to a reduction in the market value through a statistically significant negative average abnormal return (AAR) the day of and following the announcement of a data breach. There is generally no significant AARs before or after these two days, indicating that the AAR results from the market taking the new information about the data breach into account. Furthermore, the analysis suggests that a data breach, on average, leads to a negative price effect independent of the utilized event window. We believe the reduction in market value following a data breach result from a damaged reputation and reduced trust from customer, resulting in reduced future sales and increased customer turnover.

There are some deviations in the thesis in terms of the market reaction, and all industries except for the consumer industry experience a reduction in the market value following a data breach. Furthermore, the empirical results suggest that a data breach is particularly damaging in the finance industry, which we argue can be because it is highly regulated and stores sensitive customer information. In line with the increased media coverage of data breaches, the results indicate that the market reacts more negatively to data breaches in 2015 - 2019 compared to data breaches in 2010 - 2014. The cross-sectional regression suggests that the data sensitivity, number of records stolen, customer segment, and firm size influence the market reaction.

Lastly, we illustrate that a data breach is associated with increased attention from the average individual in the targeted firm and that a data breach can have large consequences for the management team, as job losses are relatively common.

The remainder of this thesis follows this structure. The remainder of Section 1 gives an introduction to the topic of data breaches. Section 2 presents relevant literature on data breaches and the stock market reaction, while Section 3 describes the dataset underlying our analysis. Section 4 introduces the financial theory that allows us to test for an effect on the stock price, and Section 5 gives a summary of the relevant event study methodology. Section 6 presents the expected findings, followed by Section 7 and the empirical findings from our analysis. Section 8 tests the robustness of the findings before a conclusion is given in Section 9, including a discussion of the limitations and suggestions for further research.

1.1 Importance of datasecurity

To grasp the importance of data security and the potential impact of a data breach, it helps to take a step back in time to look at the rapid development over the last few decades in the way we store and utilize information. When the first working transistor was built back in 1947, one of the building blocks of the third industrial revolution and the modern digital society was in place, and no one could predict the radical development which followed (Riordan, 2004). The first commercially available computer followed soon after, and over the following decades, computing power increased exponentially as the number of transistors increased, as predicted by one of Intel's founders, Gordon Moore, back in 1965 (Gustafson, 2011). This has since then commonly been referred to as "Moore`s law".

One of the significant inventions of today`s digital society is the World Wide Web, which has made a radical change to the way we communicate and conduct business. From 1990 and onwards, the internet's takeover of the communication landscape is unparalleled in a historical context. The number of users has gone from 0.05% to 53.6% of the global population from 1990 to 2019 (The World Bank, 2020; ITU, 2019, p. 3). Furthermore, it is estimated that 66% of the global population will use the internet in 2023 (Cisco, 2020, p. 5).

In combination with digital technology, the emergence of the internet has radically transformed almost every aspect of modern life as we knew it. The transformation has reaped tremendous benefits to a vast number of people, companies, and society as a whole. We are now more socially connected through digital platforms than ever, and it is easier to stay in touch with friends and family. The speed of communication is instant, and it is possible to transfer extensive amounts of data and information at a fast speed. It has revolutionized the way companies conduct business, for example, allowing storage of massive amounts of data in a relatively small space that is accessible from any device that has internet access, and an exponential increase in the computing power allowing employees to perform increasingly complex tasks faster and more precise. The digital revolution is one of the most central facilitation factors for globalization in the last decades.

The amount of data generated today is simply staggering. According to IDC (2020), more than 59 zettabytes of data, which is 59, followed by 21 zeros, will be created, copied, captured, and consumed globally in 2020. To see it from another perspective, if each terabyte in a zettabyte were one kilometer, 59 zettabytes is equivalent to 76 700 round trips to the moon and back.

The amount of data created over the next three years will surpass the amount of data created in the last 30 years (IDC, 2020), and the catchphrases "Data is the new oil" have become more common in recent years and for a good reason. This has been boosted by significantly lower prices and more accessible processing power. So, with all this information in existence and the introduction of new devices and digital technology, it might seem like there only are benefits, but at the same time, there has also been an increased risk.

Cybercriminals exploit IT networks that are vulnerable in many different ways and are always on the hunt for vulnerabilities to exploit, and the typical IT security team must mitigate different cyber threats such as phishing, brute force, and malware[1]. One of the most challenging elements for any IT security team is that there is no universal strategy for optimal IT security, as it is highly dependent on the characteristics of the company and the methods cybercriminals are using. Although it is difficult for companies to have an optimal IT security strategy, it is of high importance as it otherwise can lead to the company being the target of a data breach.

1.2 Definition and consequences of a data breach

A data breach can be defined as an incident where confidential, protected, or sensitive information has been stolen or accessed by an unauthorized individual (Groot, 2020). Data breaches can involve various information, such as financial information, health records, intellectual property, or personal details. Data breaches can involve several millions of records stolen, illustrated by the data breach on Facebook in 2016 that gave Cambridge Analytica access to private information on 50 million Facebook users (Granville, 2018). The variety of information involved in a data breach highlights the reputational loss, reduced market value, lost business, customer turnover, and vast effects on the targeted firm as consumers expect their private details to be safely stored and not misused.

Data breaches have gotten more in the media and consumers' focus in the last 20 years due to increased reliance on digital services and applications. In many cases, they require individuals

[1] The definition of the methods is included in the appendix at page 66.

to submit personal details to utilize them to their full extent, resulting in a high trust in companies and their security systems to keep personal details confidential and protected.

This thesis needs to highlight the difference between a data breach and a vulnerability, which also are known as a security flaw. The primary difference is that a data breach involves stolen or unauthorized access of sensitive, confidential, or protected data while this only was a possibility, and not done, in the case of a vulnerability. This thesis investigates data breaches and the influence on the targeted company.

In most cases, data breaches will severely damage a firm's reputation, which in turn will result in a severe loss of goodwill from its customer and suppliers. New customers might be reluctant to use the company's services, and existing customers might leave the company, resulting in an abnormal customer turnover. According to a study conducted by Deloitte (2018, p. 16), 25% of respondents would trust an organization less if its data was compromised, 70% of respondents identifies a history of data breaches as being a concern and impacting their level of trust, and 17% of participants would stop buying from an organization or using a service that was the target of a data breach. Furthermore, data breaches may lead to system downtime, investigations, and hefty fines from governments, as seen in the Cambridge Analytica data breach in 2016 (Davies & Rushe, 2019). The knock-on effect of all this will decrease the firm's cash flow, profit, and overall value. Another potential detrimental effect on a firm's financial performance resulting from a data breach is losing a competitive edge. This might be a direct effect of the information lost if a company secret is stolen or a more indirect effect by reducing available resources for future investment (Kemal Tosun, 2020, p. 2). Lastly, a data breach can increase future acquisition costs through a diminished reputation in the industry. All of this might create an unwillingness to reveal information about data breaches to the public.

Even though it is hard to collect detailed data for the cost of data breaches, it is clear that the potential cost can be tremendous. Ponemon Institute runs an annual Cost of Data Breach Study for IBM, and the 2020 study estimates that the average total cost of a data breach is $8.64 million in the United States (Ponemon Institute, 2020, p. 23). However, this cost is highly contingent on the company's characteristics and the data breach, as there are vast differences between countries, industries, and the number of records stolen. For example, breaches with 1 to 10 million records stolen had an average cost of more than $50 million, and mega-breaches (more than 50 million records stolen) had an average cost of $392 million (Ponemon Institute, 2020, p. 66). Further, the healthcare and finance industries are the industries with the highest average cost of a data breach (Ponemon Institute, 2020, p. 25).

According to Statista (2020), there has been an increase in the number of data breaches in the US from 157 in 2005 to 1506 in 2019. The increase can be seen in line with companies getting higher workforce mobility through mobile devices and digital applications and becoming more reliant on disruptive technologies such as cloud computing and cloud-based applications. In addition to this, the evolution within the Internet of things proposes a new challenge for companies as individuals are getting more reliant on digital solutions and tools in their everyday life. These elements make it increasingly difficult for companies to deal with data breaches and general IT security risks. This have made data breaches a top concern for companies, their customers, and citizens in general.

The sample analyzed in this thesis is at 46 unique data breaches, which is considerably smaller compared to the numbers presented in the last paragraph. The reason for the large deviation is our inclusion criteria that will be presented in Section 4.1 "Data breaches and criteria of inclusion" as the numbers from Statista include data breaches of all sizes on both listed and private companies in addition to government agencies.

1.3 Data breaches and Covid-19

Covid-19 has caused an economic shock that has disrupted the global economy through changes in international supply chains, closed borders, and social distancing. Further, it has caused major changes in how companies conduct business, and many employees have been forced to work from their home office and communicate through digital applications. This has led to additional security challenges as employees are more reliant on ongoing remote access to a company's internal systems, consequently putting higher pressure on updating security systems and routines to avoid security vulnerabilities that could lead to a data breach. Covid-19 and social distancing have increased home offices' use in many countries, and there has been a considerable push towards digital solutions. This has further increased the interconnectedness between business and technology and the degree to which business and digital solutions are mutually dependent on each other to deliver the same value to customers.

We believe that Covid-19 proposes several new challenges that increase the need to have reliable IT security and data breach prevention as it otherwise can lead to an increased risk of being the target of a data breach. It can be exemplified through potentially more use of free and popular messenger applications for employee communication, more sensitive and confidential data being stored and shared online, and a higher demand for video conferencing. It is also worth noticing that employees need to have a higher focus on tools that protect sensitive information as it is essential to mitigate the increased risk of a data breach. A majority of companies also expect the increased use of remote work to make it more complex and challenging to identify, contain and respond to a data breach (Ponemon Institute, 2020, p. 5).

1.4 Regulation of cybersecurity

Cybersecurity has, over the years, gradually become a more significant concern to regulators. The Securities and Exchange Commission's (SEC) Division of Enforcement, for example, established a Cyber Unit in 2017, focusing on cyber-related activities and providing guidance for companies when dealing with a data breach. SEC also provides an overview of good practices that reduce the probability of being the target of a data breach and how a company should enhance their response in the aftermath of a data breach (SEC, 2020).

Furthermore, from the 25th of May 2018, the General Data Protection Regulation (GDPR) became enforceable in the European Union (EU). GDPR, at a glance, injects a duty on all organizations to report data breaches to supervisory authorities within 72 hours. It also requires companies to inform individuals affected by the data breach if it is "likely to result in a high risk of adversely affecting individuals' rights and freedoms" without an excessive delay (ICO, 2019). However, the GDPR legislation is only valid for companies in the EU or companies that handle European citizens' data.

The United States does not have a direct GDPR equivalent on the federal level. However, different federal laws demand that companies disclose data breaches to the public when financial or healthcare information is stolen (Murciano-Goroff, 2019, p. 2). Further, data breach laws are regulated on the state level in the US. These state laws aim to protect citizens' privacy, data, and digital identity and have been irregularly passed in all states, starting with California in 2002 (Rouse, 2010).

1.5 Preventive measures

Luckily, there are various tools, precautions, and routines that decrease the chance of being the victim of a data breach. One of the most essential is to have employees that are conscious and careful with sensitive data and login details and familiar with the various methods that can be used in a data breach. By being aware of the methods, employees are less likely to enter links in suspicious emails or download software from an unknown entity. Further, it is central to have two-factor authentication when logging in to the internal systems of and applications used by a company and not reuse passwords for multiple login credentials, as it will make it easier for cybercriminals to get access to your account at other systems.

It is also central that companies and employees have good routines and systems for updating software as updates can fix potential security vulnerabilities. An example of a data breach resulting from not updating software is the Equifax data breach in 2017. The root cause of the data breach was a flaw in software for web applications that Equifax, amongst others, used, and although a fix was published shortly after the discovery, Equifax overlooked the discovery and did not update their software (Riley, Robertson & Sharpe, 2017). Finally, companies should have an IT department or support function that employees can contact if they suspect they are the target of a method used in a data breach. The IT department should also be responsible for conducting cybersecurity awareness training for employees to mitigate the risk of being the victim of a data breach.

Companies should be extra aware of cybersecurity and new employees, as new employees, amongst other things, get many emails during their starting period where they need to download software and input login credentials. Although most of the requests will be legitimate, a new employee is a perfect target for a cybercriminal (Fossmark, 2020). This is because they are new to the company, do not know the security routines by heart, and would like to make a good impression on their future coworkers, which for example, can make them vulnerable to phishing.

2. Literature review

Data breaches and cybersecurity are gradually becoming a more significant concern to companies. The negative consequences of being the target of a data breach can be massive, combined with an increasing trend in the number of data breaches in the United States. In this section, a brief literature review will be conducted. There has been some research conducted on the topic previously, but the amount of literature is relatively thin compared to other topics (Kemal Tosun, 2020, p. 6).

2.1 Acquisti, Friedman & Telang (2006)

This paper investigates the effect of privacy breaches on US companies listed on the NYSE or NASDAQ from 1999 to 2006. The paper finds a statistically significant negative impact on the targeted firm's stock value (Acquisti, Friedman & Teland, 2006, p. 12). The cumulative effect increases the day following the data breach's disclosure but then decreases and loses statistical significance (Acquisti et al., 2006, p. 1).

2.2 Goel & Shawky (2009)

This paper investigates the impact of security breaches on publicly traded US companies' market value between 2004 and 2008. The paper finds a statistically significant cumulative abnormal return and abnormal return around the event date for the sample. Further, the results of the paper indicate that the announcement of a security breach has a significant negative impact of about 1% on the targeted firm's market value (Goel & Shawky, 2009, p. 408).

2.3 Amir, Levi & Livne (2017)

Amir, Levi & Livne study if and when managers have the incentive to withhold information on cyberattacks. The paper finds that "the market reaction to disclosed attacks is indeed small, but the market reaction to withheld attacks is negative and significant" (Amir, Levi & Livne, 2017, p. 1205). Further, the paper finds that managers voluntarily disclose less severe cyber-attacks and withhold information from investors if the attack causes significant damage (Amir et al., 2017, p. 1180). The paper concludes that regulators should impose stricter mandatory disclosure rules regarding data breaches if regulators wish to ensure that information about data breaches reach investors (Amir et al., 2017, p. 1205).

2.4 Kemal Tosun (2020)

In his paper, Tosun study how financial markets react to corporate security breaches in the long- and short-term. The paper examines data breaches between 2004 and 2016 on publicly listed US companies and detects a clear downward trend in the realized stock return of the target firm's stock price throughout the event window compared to control firms (Kemal Tosun, 2020, p. 14). Further, the paper detects a significant increase in the trading volume on the day of disclosing the data breach, while there is no relationship before or after the disclosure (Kemal Tosun, 2020, p. 16). As for the long-term effects, the paper shows that the operating performance for all companies in the sample "is not significantly affected up to five years after the event" (Kemal Tosun, 2020, p. 26) and that the policies of the firms "significantly incorporate security breaches by investing more in the existing management" (Kemal Tosun, 2020, p. 26).

2.5 Kamiya, Kang, Kim, Milidonis & Stulz (2020)

The paper study the effect of data breaches on companies and their industry peers. They show that the targeted firm experiences a reputational cost that is manifested through a reduction in sales growth the three years following the data breach and a reduction in the credit rating resulting in a higher cost of debt (Kamiya, Kang, Kim, Milidonis & Stulz, 2020, p. 3). The reputational cost can be of high importance for companies as it reduces the forecasted yearly revenue and increases the bankruptcy risk. The study also shows that a data breach can impact competing firms within the same industry as it can translate into an industry-wide risk of data breaches (Kamiya et al., 2020, p. 29). Lastly, the study shows that the effect on shareholder wealth is highly dependent upon the information stolen in the data breach. Data breaches that involve loss of financial information have a significant shareholder wealth loss, while data breaches that do not include loss of financial information do not have a significant shareholder wealth loss (Kamiya et al., 2020, p. 29).

3. Data

3.1 Data breaches and criteria of inclusion

This section describes the steps when collecting and formatting the data in this thesis. The analysis's starting point will be a dataset on data breaches with more than 30.000 records stolen from 2004 to 2020 that was retrieved from the website of Information is beautiful (Barton, Evans, Geere, McCandless & Starling, 2020). One can argue that a lower boundary of 30.000 stolen records is beneficial since data breaches are getting more common and happening to a greater extent than ever before. By having a lower boundary of records stolen, one has the opportunity to focus on data breaches that affect a relatively large number of the customers, hence, getting a lot of media attention and potentially affecting the market value of the company.

The dataset contains information on the company, year, number of records stolen, data sensitivity, and the method the cybercriminals used in the data breach. All details of the data breaches provided in the dataset were manually cross-referenced with alternative sources such as blogs, news reports, and company statements to ensure that the information was correct. The first step in the data formatting processes was to restrict the dataset to only include companies publicly listed in the US at the announcement date of the data breach. The study focuses on data breaches that happened between 2010 and 2019, as it was considered too time consuming to reliably confirm the announcement date of data breaches before 2010. Further, this study focuses exclusively on data breaches that either was officially confirmed by the targeted company through a press release, statements to the media or confirmed through independent media reports. In total, 81 data breaches fulfill the criteria.

The dataset has multiple methods identified as the method of which the cybercriminals used in the data breach, and they include "poor security," "hacked," "lost device," "inside job," and "oops!". Data breaches where the method is "lost device," "inside job," or "oops!" are excluded to exclusively focus on data breaches that resulted from active measures taken by cybercriminals and exclude data breaches resulting from a mistake or being an inside job. This reduces the number of data breaches in the dataset to 64.

Some data breaches have multiple dates reported as the announcement date in the media, which makes it difficult to determine the correct announcement date. The situation is more complicated in a small number of cases since articles state that the data breach has been known for some time, but not the exact date. Data breaches where this is the case have been excluded due to a high risk of selecting the incorrect announcement date. Further, companies listed less than two years, starting at the data breach's announcement date, were removed due to a potential bias in the beta computation. If a company experienced several data breaches within the estimation window, the most recent data breach was excluded to remove a potential bias in the parameters of the normal return model and computation of abnormal returns.

The criteria discussed above results in 46 unique data breaches. The criteria provide the opportunity to examine the potential influence on the market value of the targeted company for data breaches happening as a result of active measures from cybercriminals. The distribution of data breaches by year is illustrated below in Figure 1. See Table 7 in the appendix for a complete overview of all data breaches included in the sample.

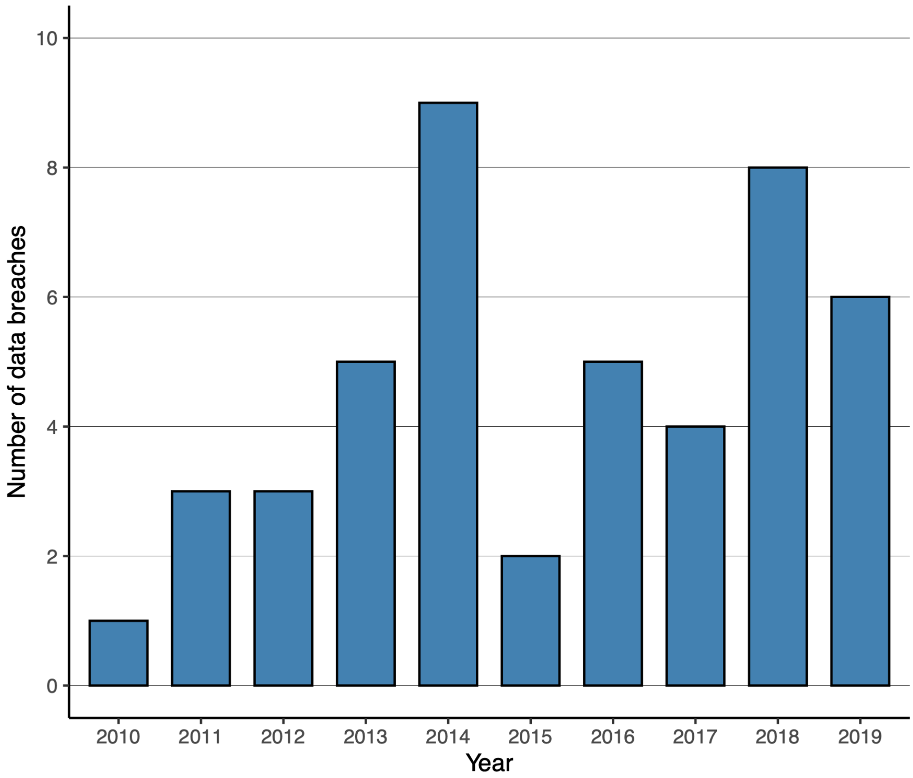


Figure 1
Distribution of data breaches in the sample by year

The announcement date of the data breach is defined as the first appearance of the data breach either in the media or through a statement on behalf of the company. Further, the thesis focuses exclusively on the first appearance of the data breach in either media, blogs, or company statements and disregards follow-up news and updates. The follow-up news may include news about lawsuits from regulators and customers, more information about the number of customers affected, or other relevant information. The additional information revealed is strongly path-dependent in each case and are therefore disregarded.

3.2 Calculation of returns

In addition to the data described above, daily and weekly data on stock prices were retrieved from Yahoo Finance and Wharton Research Data Services, CRSP. Annual data on company characteristics were also retrieved from Wharton Research Data Services, CRSP. All data on stock and index prices are adjusted for dividends and stock splits. Stock and index returns are calculated as the logarithmic change in return. Compared to simple returns, logarithmic returns will have a distribution closer to the normal distribution (Hudson & Gregoriou, 2015, p. 152), and the influence of outliers will be reduced. Logarithmic returns are calculated according to:

$$r_{i,t} = ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \tag{1}$$

Where $r_{i,t}$ is the logarithmic return for index or stock i at time t, and $P_{i,t}$ and $P_{i,t-1}$ denote the value of the index or stock i at time t and time $t-1$, respectively.

3.3 Data frequency

Daily, weekly and monthly data on stock and index prices are available for the analysis. Daily data are used in the event window as it will allow us to analyze the effect of a data breach on a company's stock price on a daily interval and increase the precision of the analysis. It is, however, worth pointing out that daily data for individual stocks tend to depart from normality (Brown & Warner, 1985, p. 4), but that the logarithmic return, which is used in this thesis, will have a distribution closer to the normal distribution (Hudson & Gregoriou, 2015, p. 152).

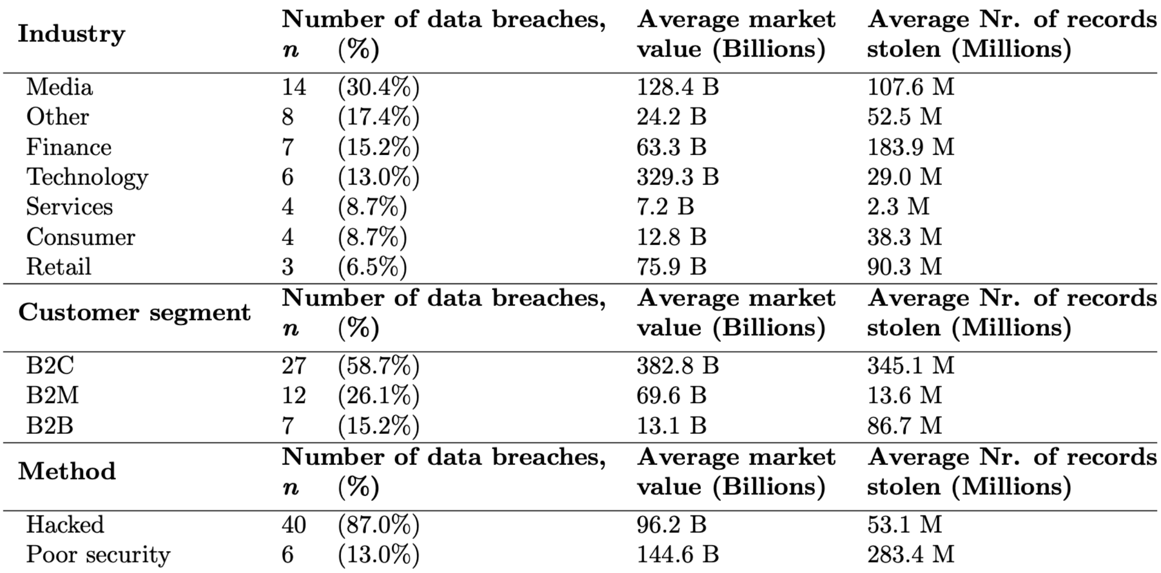
Further, we utilize two years[2] of weekly data to estimate the parameters in the market model and the variance of the abnormal return in the estimation window. Weekly data is beneficial compared to monthly data as it provides more data points and increases the estimate's precision. It is also beneficial compared to daily data in the computation of beta, as it is likely

[2] We have also conducted the same process with one year of weekly data used in the estimation window. The results were relatively similar, and the difference was marginal considering the sample size. The results are available upon request.

to assume that some stocks will have low daily liquidity. The reasoning for using two years of weekly data is due to several factors. The stock market is cyclical, and if only one year is used in the estimation of beta, the beta might be biased due to business cycles. Further, if the estimation of beta is based on four or more years of data, it might be influenced due to a change in a firm's fundamentals. Accordingly, using two years of data can mitigate the potential problems in the computation of beta.

3.4 Industry

Further, the companies targeted in the data breaches are divided into seven industries, which are inspired by the cost of a Data Breach Report 2020, written by the Ponemon Institute, and the definition of the industries can be seen in Table 15 in the appendix. Different industries store different customer information that is of different value for cybercriminals. This can be illustrated by the healthcare and finance industry storing customer information such as social security numbers (SSNs), health records, and credit card numbers that are highly valuable for cybercriminals, while the media industry, in general, store relatively less sensitive information such as name, address and phone number. The firms are also categorized according to their primary customer segment. See Table 1 on the next page for summary statistics of the sample and the distribution between industries, customer segments, and the method used by the cybercriminals.



Industry	**Number of data breaches, *n***	**(%)**	**Average market value (Billions)**	**Average Nr. of records stolen (Millions)**
Media	14	(30.4%)	128.4 B	107.6 M
Other	8	(17.4%)	24.2 B	52.5 M
Finance	7	(15.2%)	63.3 B	183.9 M
Technology	6	(13.0%)	329.3 B	29.0 M
Services	4	(8.7%)	7.2 B	2.3 M
Consumer	4	(8.7%)	12.8 B	38.3 M
Retail	3	(6.5%)	75.9 B	90.3 M
Customer segment	**Number of data breaches, *n***	**(%)**	**Average market value (Billions)**	**Average Nr. of records stolen (Millions)**
B2C	27	(58.7%)	382.8 B	345.1 M
B2M	12	(26.1%)	69.6 B	13.6 M
B2B	7	(15.2%)	13.1 B	86.7 M
Method	**Number of data breaches, *n***	**(%)**	**Average market value (Billions)**	**Average Nr. of records stolen (Millions)**
Hacked	40	(87.0%)	96.2 B	53.1 M
Poor security	6	(13.0%)	144.6 B	283.4 M

Table 1

Summary statistics of the sample

This table present the summary statistics of firms' characteristics affected by a data breach in our sample. Firms in the sample are publicly US listed companies with more than 30 000 records stolen in the data breach that happened between 2010 and 2019. The average market value (billion USD) is defined as the average market value fifteen days before the announcement of the data breach. B2C, B2M and B2B are the business to customer, business to many, and business to business customer segment, respectively.

3.5 Google Trends

Google Trends is a database that provides and analyses search queries in Google across different regions and languages. For the analysis, daily search data were downloaded for all companies in the sample, and the data was collected by searching on the company with "the whole world" as region and "all languages" as language. The data was aggregated around the data breach's announcement date and a period of ten weeks before and after the data breach by averaging the data for all companies. Google Trends rescales the frequency of the searches such that 100 represents the highest search frequency within each sample in the chosen period. For example, if one searches "Facebook" in a given period of time, a value of 100 is given to the day the query was relatively most searched.

4. Theory

4.1 The Efficient Market Hypothesis

The fundamental principle of the Efficient Market Hypothesis (EMH) is that asset prices reflect all relevant and available information and that all market actors behave rationally. This principle implies that investors cannot create trading strategies to gain an abnormal return based on the information available, and only new information will move the stock prices (Bodie, Kane & Marcus, 2018, p. 335).

Inversely, one can argue that a movement in the stock prices must reflect that additional information has been revealed to the market. EMH also assumes that all information is available for free and that there are no transaction costs in the markets. However, in reality, there are few markets where information is available for free to all investors. Even though the competition between financial institutions is highly competitive, there are still some transaction costs in the markets.

The Efficient Market Hypothesis is one of the most debated topics in economics, with literature from prominent scholars on both sides of the argument. Market anomalies such as "Post-Earnings Announcement drift," "The Weather effect," and merely irrational behavior from investors and other similar events are evidence of inefficient markets according to Lakonishok, Shleifer, and Vishny (Bodie et al., 2018, p. 357). On the other side, Fama and French argues that many of these effects are just a manifestation of different risk premiums. So, the empirical reasoning is divided among scholars. However, one of Harvard Business School's most prominent financial professors, Michal Jensen, has stated:

> “I believe there is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Market Hypothesis. That hypothesis has been tested and, with very few exceptions, found consistent with the data in a wide variety of markets.” (Jensen, 1978, p. 96)

There are three forms of the Efficient Market Hypothesis: the weak form, semi-strong form, and strong form. According to the weak form EMH, stock prices incorporates all historical information from the market such as prices and liquidity. The semi-strong form incorporates the weak form in addition to all publicly available information such as fundamentals and earnings forecasts in the stock price. The strong form incorporates both the weak and semi-strong form in addition to all private information in the stock price (Bodie et al., 2018, p. 338).

5. Methodology

5.1 Event study

This thesis uses the event study methodology to measure the price effects of a data breach on the targeted firm. The event study methodology is the standard method of measuring the reaction in the price of a security relative to an event or announcement (Binder, 1998, p. 1) and are commonly used to investigate the effect of economic events or company announcements (MacKinlay, 1997, p. 13). The Efficient Market Hypothesis's semi-strong form states that the market prices fully reflect all publicly available information (Fama, 1970, p. 404). Due to this, an abnormal price effect can result from an unanticipated event that results in new information revealed to the market.

An event study aims to examine the change in stock price relative to a benchmark model of expected returns, such as the market model or the constant mean return model. One critical assumption in an event study is that capital markets respond efficiently to publicly available news and that the impact of the news on the firm will be reflected in the company's stock price according to the semi-strong Efficient Market Hypothesis (Cable & Holland, 1999, p. 332). Event studies, in general, analyze observed reactions in the financial market, and an essential part is hypothesis testing, allowing the user to conclude on the statistical significance of the findings.

According to MacKinlay (1997, p. 13 - 15), an event study consists of the following steps:

1. Definition of the event and event window
2. Estimation of normal returns
   a. Definition of the estimation window
   b. Choice of normal return model
3. Estimation of abnormal returns in the event window
4. Hypothesis testing

In summary, an event study is a suitable methodology for analyzing data breaches as it allows one to test whether and to which degree new information about a data breach influences the stock price of the targeted company.

5.1.1 Definition of the event and event window

The first step in an event study is to identify the event of interest and decide the length of the event window. The event of interest will, in this case, be data breaches for US listed companies. As previously mentioned, the date of the event is identified as the first appearance of the data breach either in the media or through a statement on behalf of the company.

The event window will be the timeframe of the event of interest for a given security. The period before and after the event may also be of interest and is commonly included in the event window (MacKinlay, 1997, p. 15). It is central to expand the window to include the days surrounding the event because the market can have information about the event before the announcement due to information leakage. Furthermore, it is plausible to assume that the market needs some time to digest and react according to the event's information. An event's announcement can happen after the markets have closed, which is an important reason to include the day after the announcement.

The length of the event will, to a certain degree, be a trade-off between capturing the full effect of the event and having the risk of including confounding events that can influence the result. In an efficient stock market, where one does not have leakage of information, it is preferable with a relatively short event window. A short event window also reduces the likelihood of confounding events and increases the validity and reliability of the study (Jong & Naumovska, 2016, p. 1662). However, it is essential to highlight that the event window's length will vary between studies and that there does not exist one single event window suitable for all event studies.

The common practice in the event study literature is to use a two to three-day window around the event date (Jong & Naumovska, 2016, p. 1662), but it is highly dependent on the particular event. It is challenging to select one event window when investigating the topic, and we are using multiple event windows to investigate the effect of data breaches. By including multiple event windows, we can examine if the markets react before the announcement of the data breach and how long it takes the market to react to the new information. The following event windows are used in the thesis [-1, 1], [-1, 2], [-1, 3], [-2, 2], [-3, 3], [-5, 5], and [-10, 10], where the windows refer to the number of trading days relative to the date on which the data breach was disclosed. [-1, 2], for example, refers to an event window beginning the trading day before and ending two trading days after the announcement of the data breach. The complete explanation for all event windows can be found in Table 16 in the appendix.

Lastly, it is central to determine the selection criteria for the inclusion of a given firm in the event study (MacKinlay, 1997, p. 15). Section 4.1 gives a detailed description of the data retrieval process and the selection criteria in this thesis. The final sample consists of 46 unique data breaches on 40 unique companies.

5.1.2 Estimation of normal returns

Definition of the estimation window

In the next step of an event study, the estimation window needs to be established (MacKinlay, 1997, p. 15). The estimation window is central to discover potential abnormal returns and includes several data points used to calculate the company's normal return. It is central that the event and estimation window does not overlap as it otherwise can lead to the returns in the event window having a considerable influence on the normal return model (MacKinlay, 1997, p. 20). As long as companies did not go through extensive changes in the business model or financial metrics during the estimation window, the length of the window is not expected to be of much significance for the results (Krivin, Patton, Rose & Tabak, 2003, p. 3). We use an estimation window of about two years[3] in this thesis, corresponding to 504 trading days.

Further, we utilize a holdout window of about one month, corresponding to 20 trading days. The holdout window is the period between the event and the estimation window and reduces the influence of confounding events. The inclusion of a holdout window reduces the risk of confounding events influencing the normal returns for a company through biasing the estimates. Figure 2 illustrates the event window [-1, 2] and the corresponding holdout and estimation window.

[3] We have also used an estimation window of one year, 252 trading days. The results were relatively similar, and the difference was marginal considering the sample size. The results are available upon request.

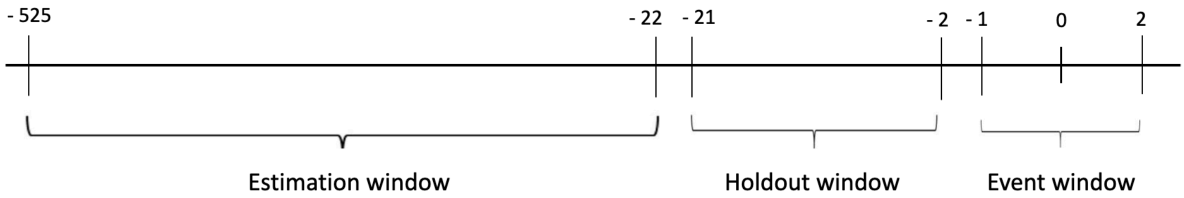


Figure 2
Illustration of event window [-1, 2]
The figure illustrates the estimation window [-525, -22] and holdout window [-21, -2] for the event window [-1, 2]. Day 0 is the announcement date of the data breach. Please note that all numbers are trading days relative to the announcement date of the data breach.

Choice of normal return model

To measure an event's effect, one needs to establish a normal return model to calculate the normal return. The normal return is the return one would expect without considering the specific event. Several models are available for computing the normal return of a security, and MacKinlay (1997, p. 17) divides the models into two groups; statistical and economical. Statistical models build on statistical assumptions and do not depend on economic arguments, while economic models build on assumptions regarding investors' behavior in combination with statistical assumptions (MacKinlay, 1997, p. 17). Some of the models are presented below.

The constant mean return model

The constant mean return model, shown in Equation (2), assumes that the mean return of a security is constant over time and that this mean is the normal return of a security. Even though the constant mean return model possibly is the simplest model to estimate, it often yields results relatively similar to more sophisticated models (MacKinlay, 1997, p. 17). Due to this, the model is often included in event studies.

$$R_{i,t} = \mu_{i,t} + \varepsilon_{i,t} \tag{2}$$

Where $R_{i,t}$ is the normal return for stock *i* at time *t*, $\mu_{i,t}$ is the average return for stock *i* at time *t*, and $\varepsilon_{i,t}$ is the residuals for stock *i* at time *t*.

The market model

The market model, shown in Equation (3), is also known as the single-index model and assumes that we have a stable linear relationship between the return of a security and the market. It is viewed as an improvement over the constant mean return model as it reduces the variance of the abnormal return (MacKinlay, 1997, p. 18).

When applying the market model, one needs to define the market portfolio to use when regressing the stock return on the market return. It is recommended to use a broad market index, such as the S&P 500, as the market portfolio when the event study investigates a large number of stocks (Krivin, Patton, Rose & Tabak, 2003, p. 3). The S&P 500 is a market-weighted index that consists of the top 500 companies in the US and indicates the movement in the US stock market. Due to this, the S&P 500 is selected as our market portfolio.

$$R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t} \tag{3}$$

Where $R_{i,t}$ is the return of stock *i* at time *t*, α_i is the assets excess return relative to the market. $R_{m,t}$ is the return of the market index at time *t*, and β_i is the stocks covariance with the market. $\varepsilon_{i,t}$ is the residuals for stock *i* at time *t*.

Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM), shown in Equation (4), is an economic model in which the expected return of a security is determined by its exposure to systematic risk, measured by beta, in addition to the risk-free rate. CAPM builds on the equilibrium theory developed by Sharpe (1964) and John Linter (1965).

$$R_{i,t} = R_f + \left(R_m - R_f\right) * \beta_i \tag{4}$$

$R_{i,t}$ is the return of stock *i* at time *t*, R_f is the risk-free rate, R_m is the return on the market portfolio, and β_i is the covariance with the market portfolio.

Multifactor models

Another category of models are multifactor models. These types of models have some clear benefits in certain situations; as MacKinlay (1997, p. 18) states in his paper, "Factor models are motivated by the benefits of reducing the variance of the abnormal return by explaining more of the variation in the normal return." The Fama-French three-factor model is one of the best-known multifactor models in finance. On a general form, factor models follow this structure:

$$R_{i,t} = \alpha_i + \beta_1 F_1 + \beta_2 F_2 + \ldots + \beta_n F_n + \varepsilon_{i,t} \tag{5}$$

The β_n in the formula indicate the effect of a given factor, F_n, on the security's return, $R_{i,t}$. According to MacKinlay (1997, p. 19), the potential gain from implementing a multifactor model, for example the Fama-French three-factor model, is limited in the context of an event study. However, if the sample data is skewed, for instance, if one specific industry or category dominates the sample, a multifactor model can reduce the variance of the abnormal returns.

Model selection

This thesis uses two statistical normal return models to investigate the price effect of a data breach on targeted companies. The thesis uses the market model and the constant mean return model, two of the most common normal return models in event studies (MacKinlay, 1997, p. 15; Cable & Holland, 1999, p. 332). The market model will be assigned the most weight in the analysis as it is considered an improvement over the constant mean return model (MacKinlay, 1997, p. 18). However, we comment briefly on the constant mean return model. Due to a potential bias in the computation of beta in the market model, all companies listed less than two years from the data breach's announcement date were excluded from the sample.

We considered economic models such as CAPM, but the decision fell on focusing exclusively on statistical models. The reason is that the result from economic models can be sensitive to the restrictions and assumptions of the chosen model, and the use of CAPM has almost ceased within the event study literature for this reason (MacKinlay, 1997, p. 19). The thesis also finds the use of multifactor models to be of limited interest since we do not have a high concentration of firms in one particular industry or one particular firm characteristic.

5.1.3 Estimation of abnormal returns

Abnormal return is essential to measure the impact of the selected event. The abnormal return is the difference between the actual observed return for a security and the estimated normal return, computed over the days in the event window (MacKinlay, 1997, p. 15). The abnormal return is denoted $\widehat{AR}_{i,t}$ and is calculated by the following formula:

$$\widehat{AR}_{i,t} = R_{i,t} - E\left(R_{i,t}[Normal\ return\ model]\right) \tag{6}$$

Where $R_{i,t}$ is the return observed at time t for stock i, and *E(*$R_{i,t}$ *[Normal return model])* is the expected return at time t for stock i based on a selected normal return model.

To draw overall inference for the selected events, the abnormal return observations must be aggregated across securities and time (MacKinlay, 1997, p. 21). We obtain the cumulative abnormal return, $\widehat{CAR}_{i,t}$, for the respective event windows by aggregating the abnormal returns for each security in each event window in Table 16 in the appendix. Further, by aggregating the abnormal return over the securities in the sample for the respective event windows, we obtain the average abnormal return, *AAR*$_t$, given in Equation (7).

$$AAR_t = \frac{1}{N}\sum_{i=1}^{N}\widehat{AR}_{i,t} \tag{7}$$

Where *N* is the number of firms in the sample and *AAR*$_t$ is the average abnormal return at time *t*.

Lastly, the cumulative average abnormal return, *CAAR*, can be calculated by aggregating the average abnormal return as follows:

$$CAAR = \sum_{t=n}^{t=m} AAR_t \tag{8}$$

Where *n* and *m* are the first and last day in the event window, respectively.

5.1.4 Hypothesis testing

After the computation of *CAAR*, the next step is to test if it is statistically significant from zero which is in this case can be formulated as follows: *H0: CAAR = 0.* The test statistic is computed by the following formula:

$$t(CAAR) = \frac{CAAR}{\sqrt{k * \hat{\hat{\sigma}}^2}} \tag{9}$$

Where *k* is the length of the event window. The variance, $\hat{\hat{\sigma}}^2(CAAR)$, is computed in the estimation window by the following formula:

$$\hat{\hat{\sigma}}^2(CAAR) = \frac{1}{N^2}\sum_{i=1}^{i=N}\hat{\sigma}_i^2(CAR) \tag{10}$$

Where N is the number of firms in the sample.

5.2 Cross-sectional regression

A cross-sectional regression is performed to explain and predict the magnitude of the observed cumulative abnormal returns (CARs) to extend the analysis. The chosen dependent variable is the CARs for the individual companies in the sample, and the independent variables will be firm- and data breach specific characteristics. The regression with CARs from the [-1, 2] event window has been selected as the primary regression. This is due to the general expectation of it taking some time for the market to digest and react according to the new information and that the announcement of the data breach can happen after the markets have closed. There will also be presented regressions with the CARs from the [-1, 1], [-2, 2], [-5, 5] and [-10, 10] event windows. By including firm- and data breach specific characteristics, we can analyze the variance of the CARs, allowing us to dig deeper into which variables affect the CAR and market value of a firm following a data breach.

Generated regressor problem
We have considered the generated regressor problem occurring when running a regression for the CAR's by bootstrapping the standard errors. One of the advantages of bootstrapping is that it does not require any pre-assumed distribution of the data in order to draw inference.

5.2.1 Variable selection

We believe two main types of broad group classification of factors will have an explanatory effect on the CARs observed in the sample. The first category accounts for firm-specific traits, including the logarithm of market capitalization, primary customer segment, and a subsidiary classification. The second category attempts to capture the traits of the individual data breaches, and we have chosen to classify the data sensitivity of the stolen records and control for the number of records stolen.

Market capitalization: To serve as a proxy for firm size, we have included the natural logarithm of the firms' market capitalization 15 days before the data breach expressed in millions of USD. The reasoning for this is to exclude any potential impact of the event. Firm size is one factor with strong empirical support when predicting a firm's return in the stock market and is included in several prominent capital pricing models, such as the Fama-French three- and five-factor model since smaller firms tend to outperform larger firms on average over a period of time. The main argument is that this is compensation for the risk taken by the investors in small firms. Accordingly, if a small firm is the target of a data breach and the data breach's traits are identical, we expect a relatively larger reaction in the market value as the cost of a data breach will represent a relatively larger part of the revenue.

Larger firms generally have a more extensive customer base than smaller firms, increasing the likelihood of more records being stolen in a data breach and the probability of being the target of a mega-breach. We believe that the relative cost in a data breach will be more significant for a small firm, although the absolute cost will be larger for a larger firm. Due to the nature of the stock market, the firm size variable is bounded by zero. Hence, a right-skewness is introduced in the variable. Researchers commonly log-transform the variable to deal with the problem of the right-skewness in the variable, which also is done in this thesis.

Primary customer segment: Primary customer classification is included to investigate if there is a clear distinction between a company's customer segments and the market's reaction to the announcement of a data breach. We believe that the business to business segment (B2B) may be more sensitive to data breaches than the business to consumer (B2C) and business to many (B2M) since a data breach on a B2B company also can influence companies that are customers of the targeted company. Hence, potentially affecting a vast number of end customers. On the other hand, a data breach for a company in the B2C segment will, most likely, affect a more extensive number of customers in itself, and the potential cost is enormous.

Since stocks are normally priced on future expectations and the potential cash flows generated by the business, potential deterioration of the trust in the company may be detrimental for the future cash flows, consequently also the stock price. Since a company in the B2C segment needs to recover the trust of "the masses," which might be troublesome, it will be affected quite hard. However, they might successfully restore their reputation for some customers which can reduce the loss in revenue. Furthermore, we would argue that trust is more deeply ingrained in the relationship between B2B partners. Therefore, a data breach will represent a more severe breach of trust, potentially leading to a higher reduction in revenue for companies in the B2B customer segment.

Subsidiary: The rationale for including an indicator variable for a subsidiary is divided. Partly, there is a diversification effect for firms with several lines of business. Hence, the overall effect of a data breach on the firm will represent a smaller part compared to non-diversified firms, and we believe the effect on the stock price will be more modest. Additionally, there is a minor effect of recognition. The parent company is not always widely known among investors. Consequently, the effect of a data breach might not be fully reflected in the parent companies' share price. The baseline in the regression is that the company targeted by the data breach is not a subsidiary.

The number of records stolen: The number of records stolen can significantly influence a data breach's cost. A higher number of stolen records will generally translate into a more comprehensive data breach. A mega-breach is defined as a data breach with more than 50 million records stolen and has, on average, a higher cost compared to smaller data breaches, which should result in a larger absolute reduction in market value. According to the Ponemon Institute (2020, p. 67), there are huge differences between the average cost of a data breach based on the number of records stolen. For example, data breaches with records stolen in the interval [1 million, 10 million] have seen a cost of about $50 million on average, and mega-breaches have seen a cost of about $392 million on average.

This is the motivation for dividing the records stolen into four categories; (1) under 1 million records stolen, which is set as the baseline for the variable in the regression, (2) between 1 million and 10 million records stolen, (3) between 10 million and 50 million records stolen and (4) more than 50 million records stolen.

Data sensitivity: Kamiya et al. (2020) showed in their study that the effect on shareholder wealth is highly dependent upon the information stolen by the cybercriminal. The loss of reputation, loss of revenue due to system downtime, abnormal customer turnover, and ripple effects following a data breach is expected to be higher in instances where sensitive information was stolen as it represents a greater breach of trust to a company's customers and often is more time demanding to resolve. Accordingly, we have created an indicator variable for data sensitivity, which captures the most severe data breaches in terms of the information stolen. The variable equals one if the information stolen either is SSNs, credit card numbers, health records, or other financial information. The baseline in the regression is that sensitive information was not stolen.

6. Expected findings

Data breaches have the potential to negatively influence the survival and competitiveness of a firm and can result in a diminished reputation, loss of business and revenue due to system downtime, and the loss of the customer's trust. A good reputation and trust can generally be one of the highest-priced assets in the highly competitive business landscape, and the loss of reputation and trust might be the deciding factor for a customer when deciding between one business or its largest rival. According to PWC (2017, p. 2), 87% of consumers will take their business to a competitor if they do not trust a company is responsible when handling their personal data. A data breach is an example of a company not having reliable IT security and not being responsible in the handling of personal data, which illustrates how a data breach may be the beginning of a large future loss of customers and revenue. Furthermore, a data breach is often associated with hefty fines and investigations from regulators that lead to high costs related to cleanup and remodeling of the IT security.

Our hypothesis builds on the elements above and previous research conducted on the topic of data breaches. We expect to find an adverse effect on companies affected by a data breach, which is in line with the Acquisti, Friedman & Telang (2006) and Goel & Shawky (2009) papers. One of the most central elements in our research question *"Does negative price effects occur in relation to US listed firms being the target of a data breach?"* is the reaction in the stock market following the announcement of a data breach. In line with this, the hypothesis in the thesis is:

Are there significant abnormal returns around the announcement date of a data breach for companies listed in the United States.

H0: There are no significant abnormal returns for a US listed company affected by a data breach around the announcement date.

HA: There are significant abnormal returns for a US listed company affected by a data breach around the announcement date.

Kamiya et al. (2020) showed in their study that the effect on shareholder wealth is highly dependent upon the information accessed by the cybercriminal in the data breach. Some industries possess more sensitive customer information, such as credit card information in the finance industry or medical records in the healthcare industry. Furthermore, the potential for litigation and class action lawsuits following data breaches is also higher when sensitive records are stolen. Based on this, we believe that the characteristics of the targeted firm and the individual data breaches to a certain degree will influence the market's reaction.

7. Empirical findings

In this section of the thesis, the empirical findings are presented. First, the findings for abnormal returns for the sample around the announcement date of a data breach will be analyzed. Further, we investigate the variance within the sample by looking deeper into the heterogeneity through the industries of the firms, time period of the data breach and conducting a cross-sectional regression containing various firm- and data breach specific traits. Lastly, we look into the attention a data breach receives from the average individual and the potential managerial consequences of a data breach.

7.1 Abnormal return

7.1.1 Cumulative abnormal return

Figure 3 presents the cumulative abnormal returns (CARs) for each stock included in the sample of data breaches in the US from 2010 to 2019, with more than 30 000 stolen records. The figure uses the market model as the normal return model, and the event window used in the graph is [-1, 2], which we believe is one of the most central windows as some of the events are announced after the market has closed on day 0 (the announcement date of the data breach). Furthermore, the event day has been defined as when the news first appears in a prominent newspaper, blog or an official statement by the company. Hence, it might take some time for the news to spread to a broader audience of investors, and the market possibly needs some time to fully incorporate the effect of a data breach in the stock price.

Figure 3 helps to get an immediate perspective into the sample and provides an overview of the companies with their representative CARs around the data breaches. One can see that a data breach can have a detrimental effect on the stock price. For instance, Equifax experienced a -23.96% CAR after the announcement of its data breach in September 2017. After one week, the data breach had eradicated about $4 billion of shareholder value (Lim, 2017). However, if we look closely at the graph, it becomes evident that data breaches, on average, do not result in a massive decline in the share price, and the average data breach in the sample results in a more modest decline represented through a -1.91% average CAR. Hence, we can see that the stock market seemingly reacts to news about a data breach, or at least in a majority of the cases.

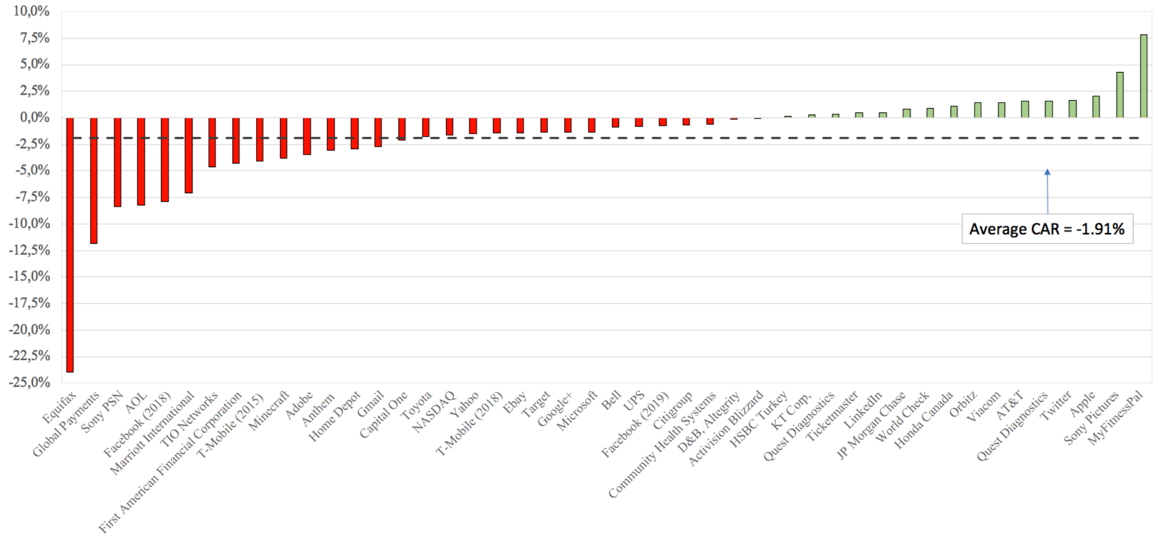


Figure 3
Overview of CARs in the sample
This figure presents the cumulative abnormal returns, CARs (%), for each firm affected by a data breach in our sample. The sample includes data breaches with more than 30 000 records stolen between 2010 and 2019. The event window is one day before to two days after the announcement date of the date breach, [-1, 2]. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days.

Figure 3 illustrates that some data breaches result in a relatively high and positive CAR, although this is a minority of the cases. The most prominent data breaches in terms of positive CAR is MyFitnessPal in 2018 and Sony Pictures in 2014. For MyFitnessPal, Under Armour (the parent company) experienced a reduction in the stock price of about 4% in after-hours trade, the announcement day of the data breach (Statt, 2018). However, the following days were accompanied by a relatively substantial increase in the stock price combined with a reduction in the value of the S&P 500, resulting in a high CAR due to utilization of the market model. Furthermore, MyFitnessPal is only a small part of Under Armour's product portfolio. The data breach on Sony Pictures in 2014 can be said to be rather unique as the cybercriminals only published a small part of the stolen data on November 24th, 2014, the announcement day of the data breach. The majority of the stolen information was published later (RiskBased Security, 2014), potentially explaining the initial reaction in the stock market. It is also worth noticing that Sony experienced a substantial reduction in the stock price in the first half of December 2014, which can result from the data breach as the release of information from the data breach was still ongoing and increasing in terms of importance for Sony.

7.1.2 Cumulative average abnormal return

Figure 4 illustrates the CAAR (%) throughout the ten days before and after the data breaches in the sample. The figure displays a negligible positive CAAR the days before a data breach. Nevertheless, after the announcement date of the data breach, one can see a sharp reduction in the CAAR from zero to six days after the data breach announcement, before the share price stabilizes at a level that is around 2.1% lower than before the data breach in the event window. From the figure, it is apparent that something is happening on an aggregated level as the CAAR begins to decrease the same day as the data breach is announced. Furthermore, the CAAR decreases the most on the day of and following the announcement of the data breach. This indicates that the market view news about a data breach negatively, which supports our expectation of a data breach reducing the reputation, trust, and revenue of the targeted company.

One might experience a problem with the correlation between events in an event study when aggregating several events. In this case, the events do not happen on the same day. Consequently, one does not need to worry about potential correlation between events.

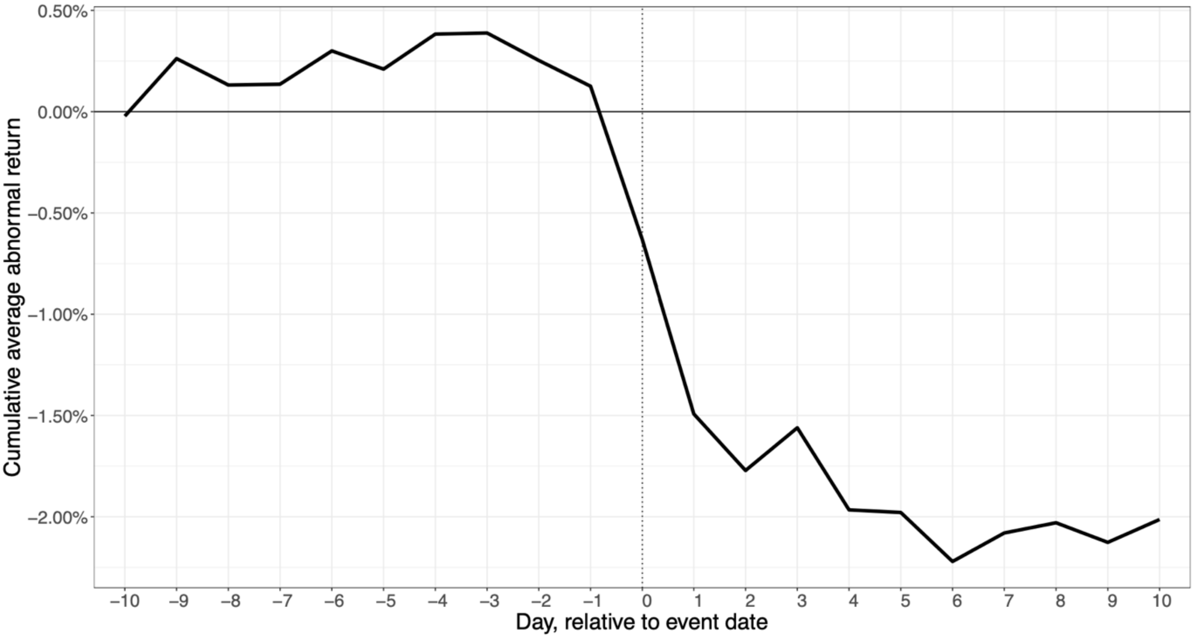
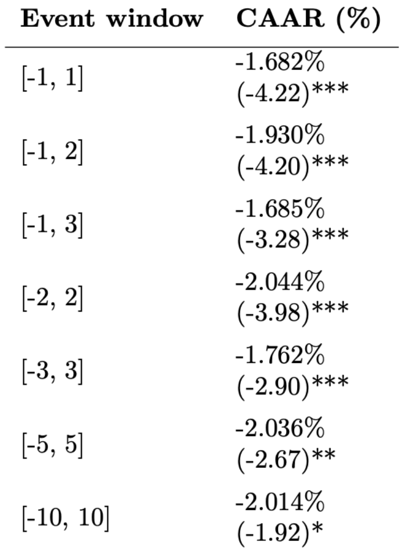


Figure 4
CAAR around the announcement date of a data breach
Cumulative average abnormal return, CAAR (%), for US listed firms affected by a data breach with more than 30 000 records stolen between 2010 and 2019. The announcement date of the data breach is day 0. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days.

The event study has been conducted for several event windows, from the smallest of [-1, 1] to a broader [-10, 10] window, which allows us to investigate if and when the effect of a data breach is shown in the market for the targeted company. The results are presented below in Table 2, and the analysis finds a reduction in the market value for all events windows, represented through a negative CAAR (%). Furthermore, the reduction in market value is statistically significant at the 1% level in several event windows, indicating that the reduction in market value can result from the data breach. These results are in line with the findings of Acquisti, Friedman & Telang (2006) and Goel & Shawky (2009) that find a statistically significant reduction in the share price of targeted companies following a data breach. Based on this, we believe that the market views a data breach as an adverse event, which corresponds with a data breach potentially influencing the targeted company's competitive edge.



Event window	**CAAR (%)**
[-1, 1]	-1.682% (-4.22)***
[-1, 2]	-1.930% (-4.20)***
[-1, 3]	-1.685% (-3.28)***
[-2, 2]	-2.044% (-3.98)***
[-3, 3]	-1.762% (-2.90)***
[-5, 5]	-2.036% (-2.67)**
[-10, 10]	-2.014% (-1.92)*

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 2
CAAR for the event windows included in the analysis
Cumulative average abnormal return, CAAR (%), for US listed firms affected by a data breach with more than 30 000 records stolen between 2010 and 2019. Different event windows are provided, for example, [-1, 1] is indicating an event window of one day before to one day after the announcement of the data breach. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. T-stats and significance levels are provided below the CAARs.

As illustrated by the table, the reduction in market value is greatest in the [-2, 2] event window, indicating that the reduction primarily seems to happen the day of and two days following a data breach. This is further supported by visual evidence in Figure 4. There does not seem to be information leakage before the announcement as the CAAR in Figure 4 does not experience a considerable reduction before the announcement of the data breach.

As of now, the analysis has identified a statistically significant reduction in the market value of a company following a data breach. The economic significance, however, is more up for debate. If one considers the potential effect a data breach can have on a company's reputation, trust, and other potential consequences such as lawsuits and abnormal customer turnover, the knock-on effect will undoubtedly decrease the firm's future cash flow, profit, and overall value. The potentially detrimental effect can be exemplified by a survey from Deloitte where 17% of participants said they would stop buying from an organization or using a service that was the target of a data breach (Deloitte, 2018, p. 16).

The overall reduction of about 2% in the share price on an aggregated level can be said to be significant, although it is somewhat lower than what the authors' believed before performing the analysis. On average, a 2% reduction in the share price corresponds to a reduction in the market value of about $2054 million, $2.054 billion, for the companies in the sample. This can be said to be substantial and reflect the future clean-up cost, redesigning of the IT security, and future customer turnover. Furthermore, it is essential to remember that this only is the short-term cost of a data breach.

7.1.3 Average abnormal return

Table 9 in the appendix presents the average abnormal return (AAR) and CAAR for the sample and allows for an investigation of AARs on a daily level. The AAR for the days -10 to -1 is not of statistical significance, which further supports that there is no information leakage before the announcement. We find a negative and statistically significant AAR on the day of the data breach, day 0, of -0.76%, and the day after the data breach, day 1, of -0.86%. The AAR on days 0 and 1 is significant at the 1% level, indicating that a data breach, on average, leads to a reduction in the market value of the targeted company. The AARs are generally not of statistical significance on days 2 through 10, although the AAR on day 4 is significant at the 10% level. This indicates that the market primarily reacts to the new information the day of and the day following the data breach.

Figure 5 illustrates AARs (%) for the period -10 to 10. The day of and following the data breach (day 0 and 1) stands out by having the greatest negative AAR at -0.76% and -0.86%. The negative and significant AAR on days 0 and 1 supports the hypothesis of a data breach resulting in a loss of reputation, future business, and customers, being manifested through a reduction in the company's market value. Although the AAR is statistically significant at the 1% level for both day 0 and day 1, the reduction in the stock price the day following the announcement of the data breach is greatest. This may result from the fact that the announcement of a data breach can happen after the markets have closed and the expectation of it taking some time for the market to digest and react according to the new information. The results for the constant mean return model have some deviations from the market model in terms of AARs value and significance, but the overall conclusion is the same for both models. The result of the constant mean return model can be seen in Table 9 in the appendix.

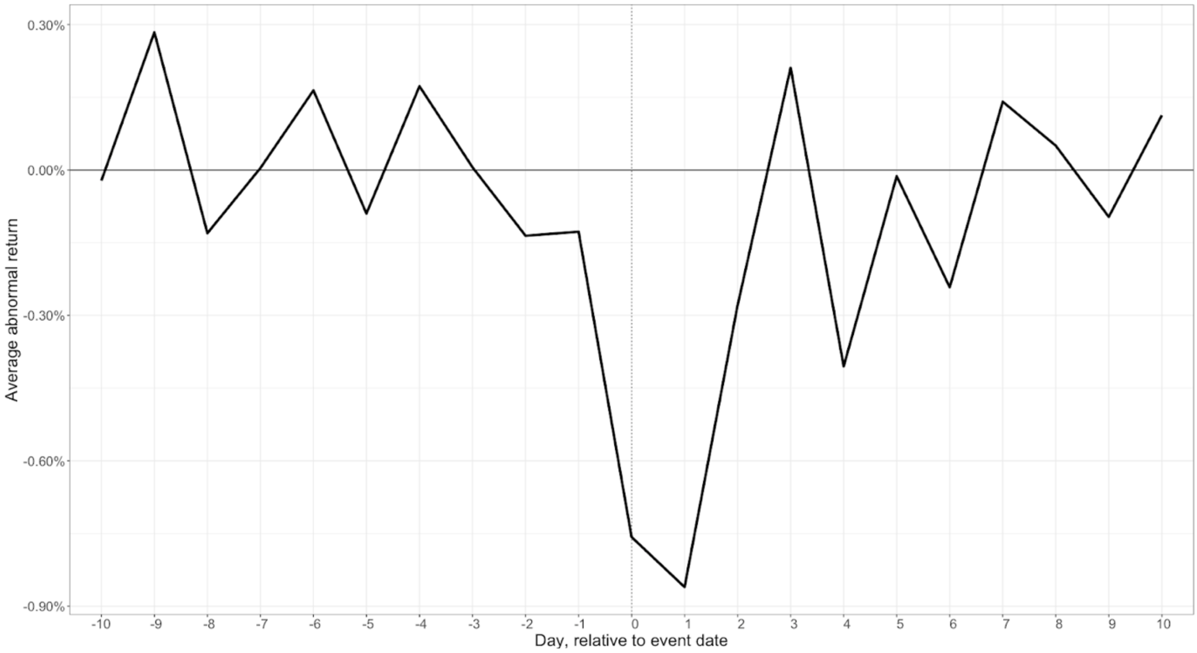


Figure 5
AAR around the announcement date of a data breach
Average abnormal return, AAR (%), for US listed firms affected by a data breach with more than 30 000 records stolen between 2010 and 2019. The announcement date of the data breach is day 0. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days.

7.2 Heterogeneity of data breaches

Figure 3 gives the reader an illustration of the CARs following a data breach for the companies included in the sample. Based on the figure, one can see that there exists a variance within the CARs in the sample from Equifax experiencing a -23.96% CAR to MyFitnessPal experiencing a +7.81% CAR in the [-1, 2] event window. To investigate the variance within the CAR for the sample we look deeper into some characteristics of the companies and the data breaches.

To investigate the sample's heterogeneity, we perform event studies to examine the effect of a data breach in different industries as companies store customer information of diverse sensitivity depending on the industry. Furthermore, in line with the increased media coverage of data breaches (Reitberger & Wetzel, 2017, p. 2), the sample are divided into two subsets based on the announcement year. Lastly, a cross-sectional regression is performed to explain the CARs observed in the sample based on various firm- and data breach specific traits.

7.2.1 Industry

The information companies store on behalf of their customers is, as previously mentioned, highly dependent on the characteristics of the firm, the service provided, and the customer segment. This can be exemplified by the finance industry having one of the highest total costs of a data breach (Ponemon Institute, 2020, p. 25) due to it being a highly regulated industry that stores sensitive customer information, such as credit card information and SSNs. The media industry, for example, stores personal details for customers that also are sensitive, but not as valuable for cybercriminals and sensitive as credit card numbers and SSNs. Accordingly, we believe it is plausible to expect that the variance with regards to the CAR in the sample partly can be explained by the firm's industry.

According to Kamiya et al. (2020, p. 29), the effect of a data breach on a company's market value is highly dependent on the stolen information. Further, the paper finds that data breaches that involve loss of financial information have a significant shareholder wealth loss, while data breaches that do not include the aforementioned have a less significant shareholder wealth loss. Consequently, we believe that the abnormal return following a data breach are influenced by the data a company possesses and that the industry of a company can be an indicator of this.

	CAAR (%)		
Industry	**[-1, 2]**	**[-1, 3]**	**[-10, 10]**
Media	-1.944% (-1.87)*	-1.461% (-1.25)	-1.604% (-0.66)
Other	-0.663% (-0.71)	-0.489% (-0.47)	-0.695% (-0.33)
Finance	-4.734% (-5.72)***	-4.912% (-5.31)***	-8.000% (-4.19)***
Technology	-1.775% (-1.55)	-2.040% (-1.59)	-4.823% (-1.84)
Services	-3.055% (-2.04)	-2.568% (-1.54)	-3.927% (-1.26)
Consumer	1.575% (0.89)	2.727 % (1.38)	8.047% (2.00)
Retail	-1.893% (-1.63)	-2.017% (-1.55)	-1.720% (-0.63)
Entire sample	-1.930% (-4.20)***	-1.685% (-3.28)***	-2.014% (-1.92)*

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 3
CAAR for the industries in the sample
Cumulative average abnormal return, CAAR (%), for the data breaches included in the sample categorized by industry. All data breaches are on US listed firms with more than 30 000 records stolen in the period 2010 to 2019. The market model: $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. T-stats and significance are provided below the CAARs. Different event windows are utilized and [-1, 2] is, for example, indicating an event window of one day before to two days after the announcement of the data breach.

Table 3 illustrates the CAAR in percentage and the statistical significance for the event windows [-1, 2], [-1, 3], and [-10, 10] for the industries in the sample. As illustrated, all industries experience a negative CAAR in the [-1, 2] and [-1, 3] event window except for the consumer industry. The development of the industries between the [-1, 2] and [-1, 3] windows are mixed, where the finance, technology, and retail industries experience a negative development that translates into a further reduction in market value. The other industries experience a positive development in the period that translates into an increase in the market value, which may indicate an initial overreaction in the stock market.

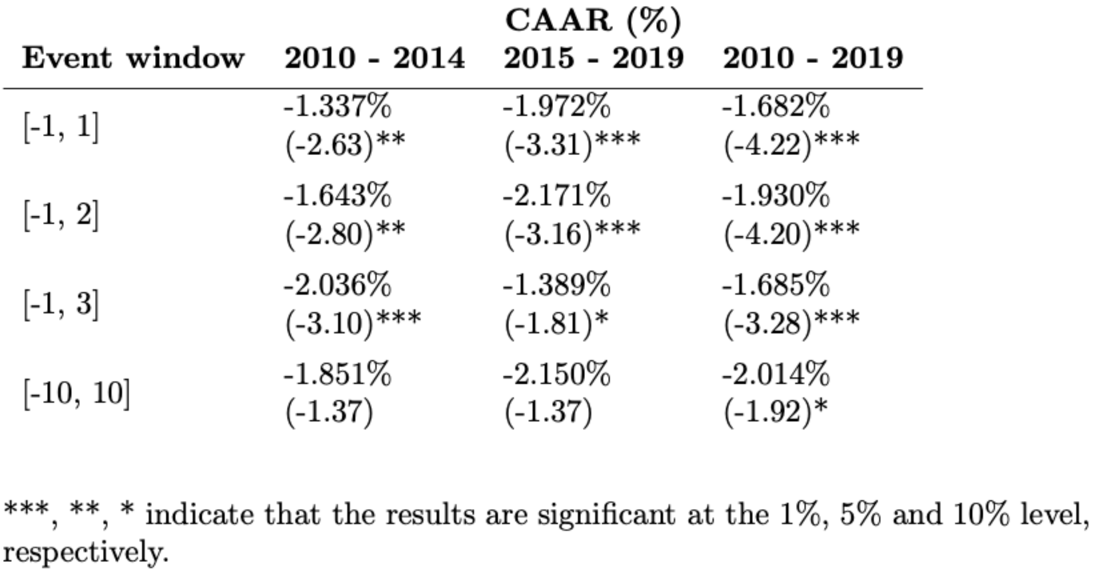
Furthermore, the table illustrates that the finance, services and media industries experience a more considerable reduction in market value in the event window [-1, 2] compared to the average in the sample. In the event window [-1, 3], the situation changes, and the CAAR in the media industry rise above the average while the retail and technology industries sink below the average. In the [-10, 10] event window, the finance, technology and services industries experience a greater reduction in the market value compared to the average in the sample. Based on this, it seems like the market reacts more negatively to a data breach in these industries that may result from the information the industries store on behalf of their customers. Another potential reason can be that investors and customers expect companies in these industries to have a relatively robust IT security and data breach prevention, resulting in a larger decrease in market value following a data breach.

The table also illustrates the statistical significance of the CAAR, and it is worth noticing that the finance industry is the only industry that has a statistically significant CAAR in all event windows (at the 1% level) and also has the highest reduction in market value. We believe this result from the industry storing sensitive information such as credit card numbers, SSNs, and other financial information that can be of great value for cybercriminals. This can be exemplified with the Target data breach, where credit card records were sold for as much as $135 per record (Ablon, Libicki & Golay, 2014, p. 14). Another potential driver may be that the finance industry is one of the most regulated industries, consequently leading to a high trust from customers and regulators in companies having robust IT security and in-depth knowledge in data breach prevention and detection. The media industry also has a statistically significant CAAR at the 10% level in the [-1, 2] event window. However, it is worth keeping in mind that the samples for the industries consist of relatively few observations making the boundaries for statistical significance relatively high.

In summary, there exist differences in the reduction in market value a company experiences after a data breach depending on the industry. All industries except the consumer industry experienced a reduction in market value following a data breach, indicating that a data breach is damaging for a firm. Based on our analysis, one can see that the finance industry seems to be affected to a greater degree compared to the other industries since it experiences the greatest reduction in market value that is significant at the 1% level in several event windows. As previously mentioned, the finance and healthcare industries store highly sensitive information on behalf of their customers, which might be the root cause of the relatively high CAAR. Optimally we would like to draw inference similarly for the healthcare industry, but this is not possible since the sample only has a limited number of data breaches for the industry (consequently being included in the other industry).

7.2.2 Time period of the data breach

In line with the increased media coverage of data breaches over the sample period, it is interesting to investigate whether there has been a change in the market reaction. To investigate this, we divide the sample into two time periods; (1) data breaches that happened between 2010 and 2014, and (2) data breaches that happened between 2015 and 2019. There are 22 data breaches between 2010 and 2014 and 25 data breaches between 2015 and 2019. The rationale behind dividing it into two time periods is that increased media coverage of a data breach will lead to more negative news about the company, consequently reducing the reputation and trust from customers relatively more. We believe this will result in a greater reduction in the market value as it may translate into an increased abnormal customer turnover.



		CAAR (%)	
Event window	2010 - 2014	2015 - 2019	2010 - 2019
[-1, 1]	-1.337% (-2.63)**	-1.972% (-3.31)***	-1.682% (-4.22)***
[-1, 2]	-1.643% (-2.80)**	-2.171% (-3.16)***	-1.930% (-4.20)***
[-1, 3]	-2.036% (-3.10)***	-1.389% (-1.81)*	-1.685% (-3.28)***
[-10, 10]	-1.851% (-1.37)	-2.150% (-1.37)	-2.014% (-1.92)*

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 4

CAAR for the time periods 2010 - 2014 and 2015 – 2019

Cumulative average abnormal returns, CAARs (%), for the companies in the sample divided into the two periods 2010 – 2014 and 2015 – 2019. All data breaches are on US listed firms with more than 30 000 records stolen. Different event windows are utilized and [-1, 1] is, for example, indicating an event window of one day before to one day after the announcement of the data breach. The market model: $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. T-stats and significance levels are provided below the CAARs.

Table 4 illustrates the CAAR and the statistical significance for the two time periods. Looking at the table, it seems as if the market reacts more negatively to data breaches in 2015 - 2019 on the day of and following the news of a data breach as the CAAR in [-1, 1] and [-1, 2] is more negative. Further, the overall reduction in the market value is more noticeable in 2015 - 2019 since the CAAR in the [-10, 10] event window is the most negative at -2.15% relative to

-1.85% in 2010 - 2014. It is worth noticing that the CAAR in the [-10, 10] event window is not statistically significance for either period, but keep in mind that the boundaries of statistical significance are relatively high compared to the full sample due to a decreased sample size in each time period.

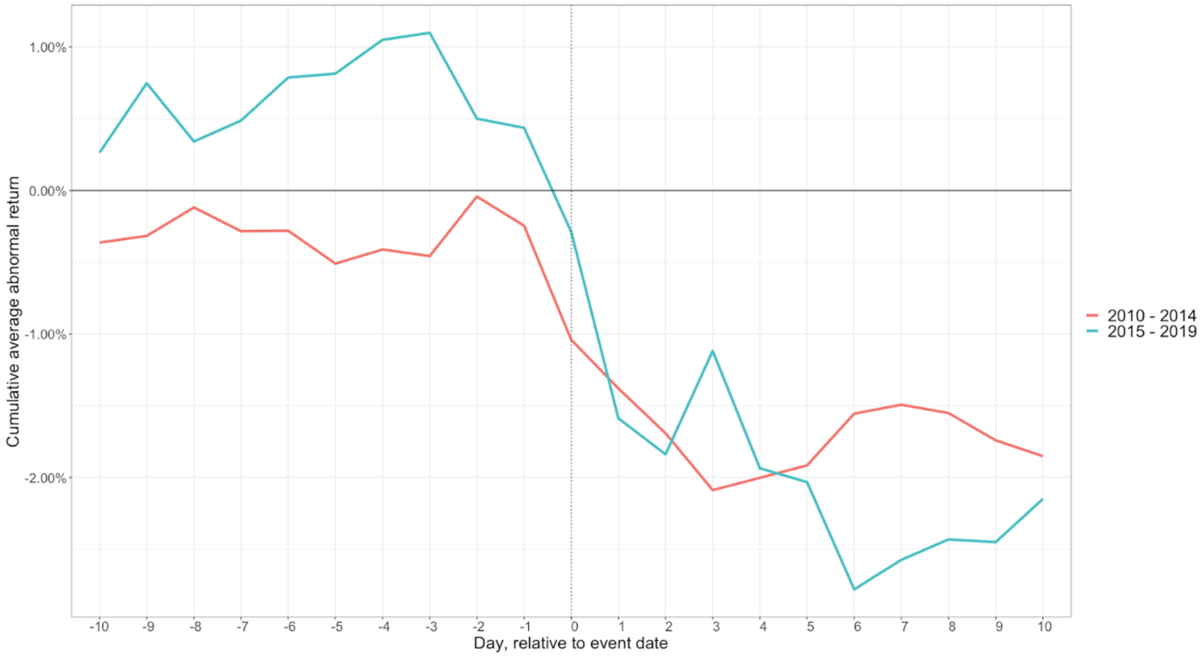


Figure 6
Comparison of CAAR for the time periods 2010 – 2014 and 2015 - 2019
Cumulative average abnormal returns, CAARs (%), for US listed firms affected by a data breach with more than 30 000 records stolen. The sample has been split into two, where the red line is the CAAR for data breaches between 2010 and 2014, while the blue line is the CAAR for data breaches between 2015 and 2019. The announcement date of the data breach is day 0. The market model: $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days

Furthermore, the CAAR is statistically significant at the 1% and 5% level in the period 2010 - 2014 and the 1% and 10% level in the period 2015 - 2019. Table 10 in the appendix illustrates the significance of the AAR for the two time periods. One can see that the AAR on the announcement day of the data breach (day 0) is -0.80% and -0.73%, respectively, for 2010 - 2014 and 2015 - 2019, which can be said to be relatively similar. On the day following the announcement (day 1), the AAR is more negative for 2015 - 2019 at -1.30% against -0.34% for 2010 - 2014, indicating that the market reacts more negatively and for a longer time in 2015 - 2019. It is also worth noticing that the AAR was statistically significant for both day 0 and day 1 in 2015 - 2019, and only significant on day 0 in 2010 - 2014.

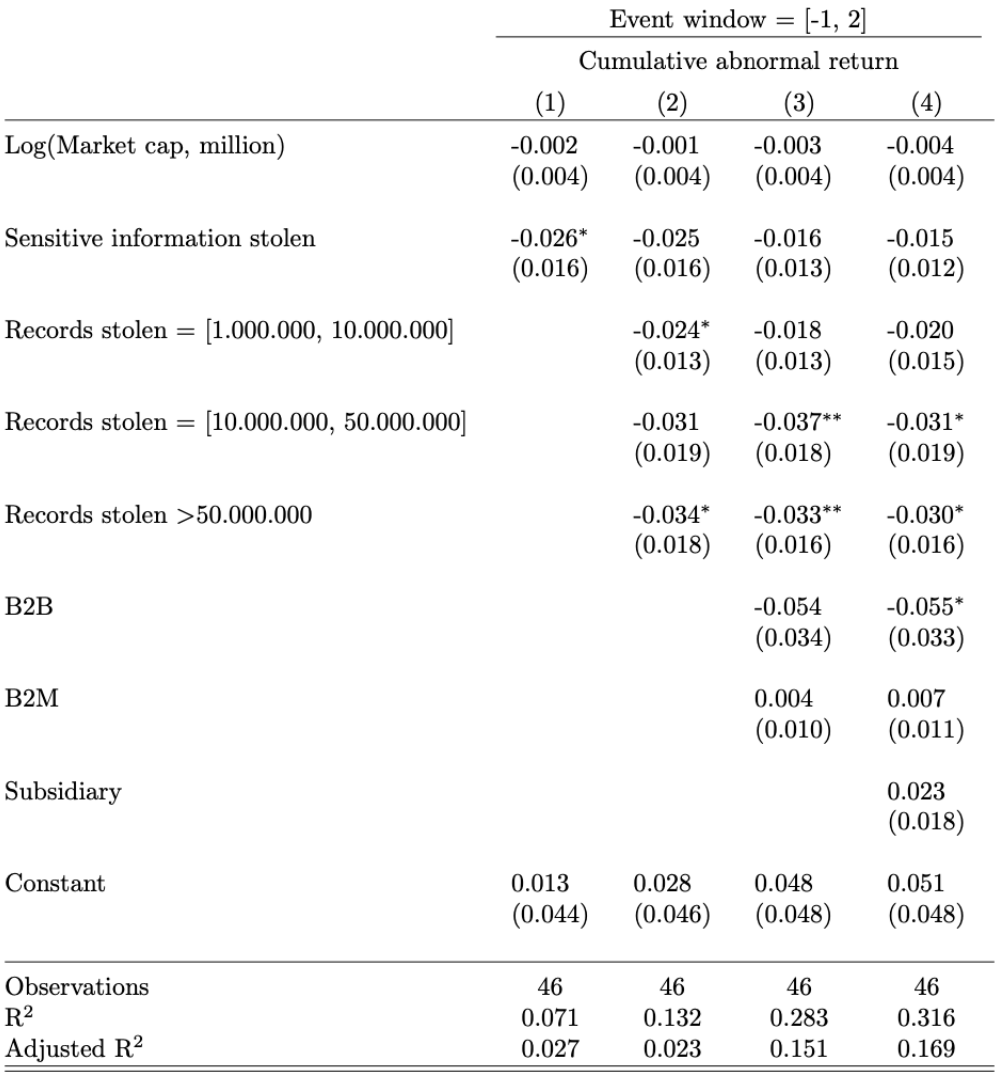
Based on this, companies in both periods experienced a reduction in the market value, but the reduction was more considerable in 2015 - 2019. It seems like the market reacts more negatively to data breaches in this period, which can be illustrated by a higher statistical significance and a more considerable reduction in market value in the [-1, 1] and [-1, 2] event windows in addition to a larger reduction in market value for the event window [-10. 10]. Companies that experienced a data breach in 2015 - 2019, on average, also experienced a greater reduction in the market value the day following the announcement of the data breach.

7.2.3 Cross-sectional regression

A cross-sectional regression is performed with various firm- and data breach specific variables to explain and predict the magnitude of the observed CARs. The rationale behind the variables is discussed in section 5.2.1 “Variable selection”. The results of the cross-sectional regression are presented on the next page in Table 5. We have created four different regressions with the natural logarithm of market capitalization and a variable capturing if sensitive information was stolen in the data breach as the first regression, which is built on with additional variables in the subsequent regressions.

The regressions use the CARs in the [-1, 2] event window as the dependent variable. The regressions with the CARs from the [-1, 2] event window will be emphasized the most due to the expectation of it taking some time for the market to digest and react according to the new information. Furthermore, the announcement of the data breach may happen after the markets have closed on day 0. Regressions based on the event window [-1, 1], [-2, 2], [-5, 5] and [-10, 10] can be found in Table 11 – Table 14 in the appendix.

From Table 5, one can observe that the natural logarithm of market capitalization contradicts our expectation that a smaller firm will experience a more considerable reduction in the market value following a data breach. This might be due to a larger firm having a relatively larger customer base, which can translate into larger abnormal customer turnover through a diminished reputation and customer trust. Another potential reason for this might be that a larger firm will get more attention in the media following a data breach, as readers are more interested in data breaches involving large and well-known firms.



	Event window = [-1, 2]			
	Cumulative abnormal return			
	(1)	(2)	(3)	(4)
Log(Market cap, million)	-0.002 (0.004)	-0.001 (0.004)	-0.003 (0.004)	-0.004 (0.004)
Sensitive information stolen	-0.026* (0.016)	-0.025 (0.016)	-0.016 (0.013)	-0.015 (0.012)
Records stolen = [1.000.000, 10.000.000]		-0.024* (0.013)	-0.018 (0.013)	-0.020 (0.015)
Records stolen = [10.000.000, 50.000.000]		-0.031 (0.019)	-0.037** (0.018)	-0.031* (0.019)
Records stolen >50.000.000		-0.034* (0.018)	-0.033** (0.016)	-0.030* (0.016)
B2B			-0.054 (0.034)	-0.055* (0.033)
B2M			0.004 (0.010)	0.007 (0.011)
Subsidiary				0.023 (0.018)
Constant	0.013 (0.044)	0.028 (0.046)	0.048 (0.048)	0.051 (0.048)
Observations	46	46	46	46
R^2	0.071	0.132	0.283	0.316
Adjusted R^2	0.027	0.023	0.151	0.169

Table 5
Regressions of CAR for the sample in the event window [-1, 2]
Cross-sectional regressions where the dependent variable is the cumulative abnormal return (CAR) for the data breaches in the sample on the event window [-1, 2]. The market capital is defined as the market value fifteen days before the data breach. Sensitive information stolen is equal to one if either SSNs, credit card numbers, financial information or healthcare information were stolen. The records stolen variable is an indicator variable indicating the number of records stolen in the data breach, and the baseline in the regression is that less than 1 000 000 records were stolen. B2B and B2M is an indicator variable indicating if the company are in the business to business or business to many customer segments, respectively. The baseline customer segment in the regression is the business to customer segment. Subsidiary is equal to one if the company are a subsidiary. The standard errors (in parentheses) in the regression are bootstrapped at 10 000 replacements. *, **, *** denote that the variable is statistically significant at the 10%, 5% and 1% levels, respectively.

The variable indicating if sensitive information was stolen influences the CAR according to our expectation as it increases the reduction in market value following a data breach, which is consistent with the findings of Kamiya et al. (2020). We believe this is due to sensitive information being associated with a more severe breach of trust to a company's customers as they consider this information more valuable than personal details as address and phone number. Although the variable is generally not statistically significant in the regressions, we believe it impacts the CAR and that the small sample size influences the significance.

According to our regression, the records stolen variable follows a logical order regarding the impact on the CAR in the second regression, with the lowest number of records stolen having the smallest impact. Generally, it seems like more records stolen in a data breach lead to a larger reduction in market value, although data breaches with records stolen in the interval [10.000.000, 50.000.000] are associated with a larger reduction compared to mega-breaches in regression three and four. We believe this is a bit counter-intuitive, as a higher number of records stolen is an indicator of the breach affecting a larger number of the customers. A possible reason for this might be that it is more likely to have a large company experience a mega-breach, and we believe there exist a correlation between the firm size and the risk of having a mega-breach. Consequently, the fraction of the customer base affected will be smaller compared to a breach in [10.000.000, 50.000.000] records stolen category on a medium-sized company.

It is also worth noticing that the number of records stolen is statistically significant at the 5% or 10% level for several regressions in the [-1, 2] event window, while only the category >50.000.000 is statistically significant in all regressions.

The customer segment of a company seems to be influencing the CAR, as companies in the B2B segment have a more considerable reduction in CAR compared to the baseline (B2C) and B2M. This is consistent with our initial expectation, as we believed trust is more deeply ingrained in the relationship between B2B partners. Accordingly, a data breach will represent a more severe breach of trust and potentially lead to a higher reduction in the overall value.

The regression models' explanatory power varies significantly from 2.3% to 16.9%, which tells us that the models to a varying degree can explain the variation in the CARs in the sample. The overall result is the same for the regressions models for the event windows [-1, 1], [-2, 2], [-5, 5] and [-10, 10] (located in Table 11 - Table 14 in the appendix), although the coefficients have some variation due to different length of the event windows.

7.3 Attention from the average individual

A critical stakeholder in any company is their customers, and although we so far have analyzed the reaction in the stock market following a data breach, it is not given that the reaction reflects the view of their customers. Data from Google Trends allows the user to analyze what individuals are searching for as an event unfolds in real-time (Rogers, 2016), essentially serving as a proxy for the attention an event receives from the average individual and customer relative to the "normal."

One of the underlying assumptions throughout this thesis is that a data breach will increase investors' attention in a firm as a data breach represents a negative shock to a company's reputation, trust, and future growth prospects. Visual evidence of the attention a firm receives following a data breach is provided in Figure 7. One can see a definite and significant spike around day 0, which is the announcement date of the data breach, and indicates that a data breach results in increased attention from the average individual and customer. Further, one can see that the relative search volume and attention received gradually decrease and stabilize at a "normal" level around ten days after the announcement.

To get a more robust confirmation of the assumption, we ran an event study on the data from Google Trends. The study resulted in a statistically significant increase in the search volume around the announcement date of the data breach. Accordingly, it provided evidence towards the fact that a data breach, in addition to reducing the market value, can lead to increased media attention, which further can lead to a negative shift in the reputation of a company and the trust from customers in the long-term.

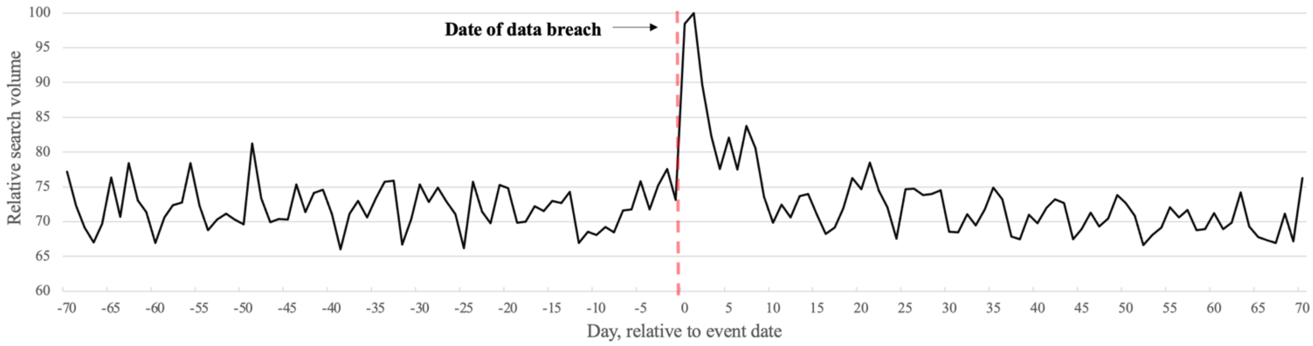


Figure 7
Google Trends - Relative search volume
The figure displays the relative search volume on Google around the announcement date of a data breach and a period of ten weeks before and after the announcement of the data breach. The data have been averaged and aggregated for all data breaches in the sample. The sample consists of data breaches between 2010 and 2019 on US listed firms with more than 30 000 records stolen. Google Trends rescales the frequency of the searches such that 100 represents the highest search volume within each sample for the chosen period.

7.4 Managerial consequences

As of now, we have seen that data breaches can lead to a diminished reputation and trust from customers and a reduction in the market value of a company. Data breaches are also an essential aspect for the management team as a central part of their job is to take care of and secure their stakeholders' interests. An important element is to secure reliable IT security and robust data breach prevention and detection. To illustrate the potential consequences for the management team, we dig deeper into the Equifax data breach in 2017 and the consequences that followed as a result of it.

The data breach on Equifax in March 2017 (disclosed in September 2017) is one of the data breaches with the highest impact in US history as the personal information of at least 143 million people was stolen by cybercriminals (Riley, Robertson & Sharpe, 2017). The drop in the share price in the days following the Equifax data breach was massive and caused significant damage to many of Equifax's stakeholders. In the [-1, 2] event window, Equifax, for example, experienced a CAR of -23.96%, which is, as illustrated in Figure 3, the largest reduction in our sample. The data breach was truly harmful for many reasons, and one of the main reasons is that the stolen data was highly sensitive SSNs, credit card numbers, and other financial information that can have great value to cybercriminals. Although Equifax had invested several million dollars in IT security measures and data breach prevention, they still became the target of a data breach that will echo for years ahead.

In addition to the massive reduction in market value, the data breach led to the CEO leaving the company, effective immediately, three weeks after disclosing the data breach (Equifax, 2017). Equifax's former CEO also lost the annual bonus and a $5 million severance package (Ng, 2017). Furthermore, the chief security officer and chief information officer (CIO) of Equifax left the company the month following the data breach's disclosure (Equifax, 2017b). Another example of managerial consequences is the Target data breach in 2013, where the CEO and CIO lost their job in addition to several board members almost losing their seats on the board (Basu, 2014).

Although it might seem like the departure of key personnel is unique to Equifax and Target, it is quite common, and almost one-third of all data breaches result in job losses (Kaspersky, 2018). The Equifax data breach also led to a massive cleanup job and remodeling of the IT security, which two years after the data breach has cost around $1.4 billion (Hurt, 2019). Based on this, the management team should have an incentive to secure a robust data breach prevention and detection as their job might rely on it.

8. Robustness of findings

Placebo test

To investigate the reliability of the identified event dates and abnormal returns for the entire sample, we have implemented a placebo test where the sample and methodology are equal. However, the event window, estimation window, and event date have been shifted 20 trading days back in time. This translates into about one month (in trading days) before the announcement of the data breach. A placebo test allows testing of whether the results are spurious, or in other words, if the effect of data breaches only is significant for the event dates we are analyzing and not arbitrarily chosen dates.

	CAAR (%)	
Event window	**Initial study**	**Placebo test**
[-1, 1]	-1.682% (-4.22)***	-0.603% (-0.40)
[-1, 2]	-1.930% (-4.20)***	0.131% (0.28)
[-1, 3]	-1.685% (-3.28)***	0.205% (0.40)
[-2, 2]	-2.044% (-3.98)***	0.285% (0.55)
[-3, 3]	-1.762% (-2.90)***	0.215% (0.35)
[-5, 5]	-2.036% (-2.67)**	-0.354% (-0.46)
[-10, 10]	-2.014% (-1.92)*	-0.201% (-0.19)

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 6

CAAR for the initial study and placebo test

Cumulative average abnormal returns, CAARs (%), divided by the initial study and placebo test. All data breaches are on US listed firms with more than 30 000 records stolen between 2010 and 2019. The placebo test and initial study have equal methodology, but the event window, estimation window, and event date have been shifted 20 trading days back in time. Different event windows are used where [-1, 1], for example, is an event window of one day before to one day after the announcement of the data breach. The market model: $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. T-stats and significance levels are provided below the CAARs.

Table 6 shows the results of the placebo test and the initial study for the event windows. Through the table, one can see that the change in market value, CAAR (%), is exclusively negative for the initial study in all event windows while it is more variable for the placebo test with both positive and negative CAARs. Furthermore, the t-stats indicate a statistically significant reduction in market value for the initial study in all event windows, while the change in market value is not statistically significant for any event window in the placebo test. This indicates that the data breaches can be the reason for the observed reduction in the targeted firm's market value following a data breach, and the reduction not being arbitrary which supports the hypothesis and findings of the thesis.

9. Conclusion

9.1 Summary

This thesis aimed to analyze the short-term impact of data breaches on companies listed in the United States through answering the research question: *"Does negative price effects occur in relation to US listed firms being the target of a data breach?"*. The market reaction was measured by studying the abnormal returns around the announcement date of the data breach. Furthermore, we looked deeper into firm- and data breach specific traits to investigate the heterogeneity in the abnormal returns following a data breach to identify characteristics that affected the market reaction. Lastly, we looked at the average individual's reaction and the potential managerial consequences of a data breach.

To examine the abnormal returns, an event study investigating the market return of the targeted companies in data breaches in the period of 2010 to 2019 with more than 30 000 records stolen was conducted. Our empirical results suggest that a data breach leads to a statistically significant reduction in the stock price on the day of and following the announcement of a data breach. Furthermore, on average, a data breach leads to a negative price effect, and all event windows in the study indicated a statistically significant reduction in the market value following a data breach. The results indicated that the market's reaction is primarily occurring on the day of and two days following a data breach. A placebo test was also conducted, indicating that the reduction in market value is not spurious and a result of the data breach. It appears that companies receive more attention after a data breach, which can weaken one of its most priced assets, the reputation.

There are, however, some deviations within the sample in terms of the price effect following a data breach. All industries except for the consumer industry experienced a reduction in the market value following a data breach, indicating that a data breach is damaging. A data breach seems to be particularly damaging in the finance industry. The industry is highly regulated and stores sensitive customer information, leading to a high trust from regulators and customers in having robust IT security and data breach prevention. Further, the market seems to react more negatively to data breaches in 2015 - 2019 relative to data breaches in 2010 - 2014. By conducting a cross-sectional regression, the empirical results suggest that the customer segment, the number of records stolen, data sensitivity, and firm size of the targeted firm also influence the stock market's reaction.

Lastly, we looked at the managerial consequences following a data breach. In addition to reducing the market value, a data breach can lead to the departure of key personnel and have massive consequences for the management team through decreased job security. The thesis also found an increase in the relative search volume for a firm around the announcement of a data breach, indicating increased attention from the average individual.

In conclusion, this study indicates a negative and statistically significant relationship between data breaches and the targeted companies' market value. This thesis is primarily investigating the short-term effects of data breaches, and it has been shown that they definitely can be damaging to the reputation, trust from customers, future revenue and market value of a company. Due to the increased interconnectedness between business and technology and the change in the business landscape due to Covid-19, data breaches most likely will continue to create headlines in the media and influence the business landscape in the years ahead.

9.2 Limitations

There are several methods to approach the research question in this thesis, and the selected methodology can influence the results. This can be exemplified through the selection of a normal return model for computing normal (and abnormal) returns, which is a central element in the event study methodology. Further, different test statistics can be used when investigating the significance of the empirical results.

This thesis focuses on data breaches in the period 2010 - 2019 to analyze the effect of a data breach on the abnormal return of the targeted companies. Although our final sample consists of 46 unique data breaches, it is unfortunate that there were no more data breaches consistent with our inclusion criteria as the limited sample can have influenced the empirical results.

Another potential challenge is the availability of data. For the empirical analysis in the thesis, we relied on a dataset from Information is beautiful that had collected data breaches with more than 30.000 stolen records. Although we did not find any data breaches from the analyzed period that were not included in the dataset, we cannot exclude this entirely. Due to the nature of a data breach, firms may want to keep them secret because of the associated loss of trust, value, and reputation. This may lead to data breaches never being reported, consequently, not included in this thesis. Lastly, the thesis focuses exclusively on companies listed in the United

States at the announcement date of the data breach. This leads to data breaches targeting companies listed in other parts of the world not being included.

9.3 Suggestions for further research

Data breaches are a topic in which previous research is relatively thin compared to other topics. We believe that data breaches and IT security are exciting topics that will become of greater attention for both companies and academics in the future.

Security vulnerabilities are not investigated in this thesis but are, in some respects, relatively similar to data breaches. We believe that the adverse effects of a security vulnerability will, in some respects, be relatively similar to the negative aspects of data breaches. We believe that a natural extension of the thesis could be to investigate security vulnerabilities. However, it is central to highlight that while the characteristics of security vulnerabilities might seem similar to data breaches, they are also different as security vulnerabilities do not involve data being stolen from companies.

Furthermore, the data stolen in data breaches are highly different from incident to incident. In some cases, highly sensitive information such as SSNs, financial information, and passwords are stolen, while other cases include less sensitive information such as an address, email, and other personal details. We believe that the stolen data to a relatively high degree will drive the reduction in market value following a data breach, and at the same time, that the value of the stolen data will be company-specific, and that the same data stolen in two separate data breaches may constitute a different market reaction. In summary, we suggest conducting a future study that focuses on data breaches where the same information was stolen from relatively similar companies. This would lead to a limited sample of events at the current point in time, but as data breaches become increasingly common, the number of events and reliability of a potential future study could increase.

GDPR constitutes a regulatory change in the data breach landscape for European companies or companies that handle European citizens' data. The increased standard of data security following the implementation of GDPR could lead to a more extensive punishment in the form of higher regulatory sanctions, increased reputational loss, and higher abnormal customer turnover, which could translate into a larger reduction in market value following a data breach.

Due to this, we believe a study on European companies and the effect of a data breach could be of high value to the field of research.

Google Trends is a database that allows the user to analyze different search queries in Google Search, which can serve as a proxy for the attention a company receives from the public following a data breach. We believe that it could be an exciting element to include and look more into in future research.

Lastly, we believe the number of days a company uses to contain the data breach or getting the situation "back to normal" could be an interesting variable to look into in an analysis. The rationale is that more severe data breaches generally will take a longer time to contain and translate into a higher total cost associated with the data breach as it would demand a more comprehensive rework of the company's IT security. However, it is difficult to find data on this subject, and we believe it would require qualitative research such as unstructured interviews with the targeted firms to obtain this insight.

10. References

Ablon, L., Libicki, M. C. & Golay, A. A. (2014). Markets for Cybercrime Tools and Stolen Data - Hackers' Bazaar. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR610/RAND_RR610.pdf.

Acquisti, A., Friedman, A. & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 proceedings.* 94. Retrieved from https://aisel.aisnet.org/icis2006/94/.

Amir, E., Levi, S. & Livne, T. (2017). Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of accounting studies, 23*, 1177 – 1206. https://doi.org/10.1007/s11142-018-9452-4.

Barton, P., Evans, T., Geere, D., McCandless, D. & Starling, S. (2020, 25th of May). World's biggest data breaches & hacks. Retrieved 25.08.20 from https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/.

Basu, E. (2014, 15th of June). Target CEO Fired - Can You Be Fired If Your Company Is Hacked? Retrieved 13.11.20 from https://www.forbes.com/sites/ericbasu/2014/06/15/target-ceo-fired-can-you-be-fired-if-your-company-is-hacked/?sh=116989f97c9f.

Binder, J. J. (1998). The Event Study Methodology Since 1969. *Review of Quantitative Finance and Accounting*, 11(2), 111 - 137. https://doi.org/10.1023/A:1008295500105.

Bodie, Z., Kane, A. & Marcus, A. J. (2018). *Investments*. New York: McGraw Hill.

Brown, S., J. & Warner, J. B. (1985). Using daily stock returns - The case of Event Studies. *Journal of Financial Economics, 14*(1), 3 - 31. https://doi.org/10.1016/0304-405X(85)90042-X.

Cable, J. & Holland, K. (1999). Modelling Normal Returns in Event Studies: A Model-Selection Approach and Pilot Study. *The European journal of Finance*, 5(4), 331 - 341. https://doi.org/10.1080/135184799336993.

Cisco. (2020, 9th of March). Cisco Annual Internet Report (2018 – 2023) White Paper. Retrieved from https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

Davies, R. & Rushe, D. (2019, 24th of July). Facebook to pay $5bn fine as regulator settles Cambridge Analytica complaint. *The Guardian*. Retrieved 21.10.20 from https://www.theguardian.com/technology/2019/jul/24/facebook-to-pay-5bn-fine-as-regulator-files-cambridge-analytica-complaint.

Deloitte. (2018). A new era for privacy - GDPR six months on. Retrieved 29.10.20 from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiJwMn08NnsAhXp_CoKHXNYA5kQFjABegQIBBAC&url=https%3A%2F%2Fwww2.deloitte.com%2Fcontent%2Fdam%2FDeloitte%2Fuk%2FDocuments%2Frisk%2Fdeloitte-uk-risk-gdpr-six-months-on.pdf&usg=AOvVaw2f_RkG76Kb79QBOHmohlCz.

Equifax. (2017, 26th of September). *Equifax Chairman, CEO, Richard Smith Retires; Board of Directors Appoints Current Board Member Mark Feidler Chairman; Paulino do Rego Barros, Jr. Appointed Interim CEO; Company to Initiate CEO Search* [Press release]. Retrieved from https://investor.equifax.com/news-and-events/press-releases/2017/09-26-2017-140531280.

Equifax. (2017b, 15th of September). *Equifax Releases Details on Cybersecurity Incident, Announces Personnel Changes* [Press release]. Retrieved from https://investor.equifax.com/news-and-events/press-releases/2017/09-15-2017-224018832.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The journal of finance*, 25(2), 383 - 417. https://doi.org/10.2307/2325486.

Fossmark, E. (2020, 9th of October). Det hjelper ikke å ha drillet medarbeidere i dataangrep, når målet er nye ansatte. Retrieved 25.10.20 from https://karriere360.no/artikler/det-hjelper-ikke-a-ha-drillet-medarbeidere-i-dataangrep-nar-malet-er-nye-ansatte-br/500507.

Goel, S. & Shawky, H., A. (2009). Estimating the market impact of security breach announcements on firm values. *Information and management, 46*, 404 - 410. https://doi.org/10.1016/j.im.2009.06.005.

Granville, K. (2018, 19th of March). Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. Retrieved 20.10.20 from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html.

Groot, J. D. (2020, 5th of October). *The history of Data Breaches*. Digital Guardian. Retrieved 13.10.20 from https://digitalguardian.com/blog/history-data-breaches.

Gustafson J. L. (2011) Moore's Law. In: Padua D. (eds) Encyclopedia of Parallel Computing. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-09766-4_81.

Hudson, R., S. & Gregoriou, A. (2015). Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis*, 38, 151 - 162. https://doi.org/10.1016/j.irfa.2014.10.008.

Hurt, E. (2019, 10th of May). Equifax Says Cybersecurity Breach Has Cost $1.4 Billion. Retrieved 04.11.20 from https://www.wabe.org/equifax-says-cybersecurity-breach-has-cost-1-4-billion/.

ICO. (2019, 22nd of May). Personal data breaches. Retrieved 28.09.20 from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/personal-data-breaches/.

IDC. (2020, 8th of May). IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data. Retrieved 28.10.20 from https://www.idc.com/getdoc.jsp?containerId=prUS46286020.

ITU. (2019). Measuring digital development: Facts and figures 2019. Retrieved 17.11.20 from https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx.

Jensen, M. (1978). Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics, 6*(2-3), 95 -101. https://doi.org/10.1016/0304-405X(78)90025-9.

Johansen, A., G. (2019, 27th of November). What is ransomware and how to help prevent ransomware attacks. Retrieved 22.10.20 from https://us.norton.com/internetsecurity-malware-ransomware-5-dos-and-donts.html.

Johansen, A., G. (2020, 24th of July). What is a Trojan? Is it a virus or is it malware? Retrieved 22.10.20 from https://us.norton.com/internetsecurity-malware-what-is-a-trojan.html.

Jong, A. & Naumovska, I. (2016). A Note on Event Studies in Finance and Management Research. *Review of Finance*, 20(4), 1659-1672. https://doi.org/10.1093/rof/rfv037.

Kamiya, S., Kang, J., K., Kim, J., Milidonis, A. & Stulz, R. M. (2020). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, 1 - 31. https://doi.org/10.1016/j.jfineco.2019.05.019.

Kaspersky. (2018). From data boom to data doom: the risks and rewards of protecting personal data. Retrieved from https://www.kaspersky.com/blog/data-protection-report/23824/.

Kaspersky. (2020). Brute Force Attack: Definition and Examples. Retrieved 22.10.20 from https://www.kaspersky.com/resource-center/definitions/brute-force-attack.

Kemal Tosun, O. (2020). Cyber Attacks and Stock Market Activity. Working paper, Cardiff Business School. Retrieved 27.10.20 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3190454.

Krivin, D., Patton, R., Rose, E. & Tabak, D. (2003). Determination of the Appropriate Event Window Length in Individual Stock Event Studies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.466161.

Lim, J., P. (2017, 12th of September). Equifax's Massive Data Breach Has Cost the Company $4 Billion So Far. Retrieved 03.11.20 https://money.com/equifaxs-massive-data-breach-has-cost-the-company-4-billion-so-far/.

MacKinlay, C.A. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, 35(1), 13 - 39. Retrieved from https://www.jstor.org/stable/2729691.

Murciano-Goroff, R. (2019). Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure? https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_33.pdf

Ng, A. (2017, 26th of September). Equifax CEO steps down in wake of massive data breach. Retrieved 11.11.20 from https://www.cnet.com/news/equifax-ceo-steps-down-in-wake-of-massive-data-breach/.

Norton. (2019, 9th of August). What is spyware? And how to remove it. Retrieved 22.10.20 from https://us.norton.com/internetsecurity-how-to-catch-spyware-before-it-snags-you.html.

Ponemon Institute. (2020, July). 2020 Cost of Data Breach Study. Retrieved from https://www.ibm.com/security/data-breach.

Porter, K. (2020, 25th of September). What is phishing? How to recognize and avoid phishing scams. Retrieved 22.10.20 from https://us.norton.com/internetsecurity-online-scams-what-is-phishing.html.

Rouse, M. (2010, May). California Security Breach Information Act. From https://searchcio.techtarget.com/definition/California-Security-Breach-Information-Act

PwC. (2017). Consumer Intelligence Series: Protect.me. Retrieved 10.11.20 from https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/cybersecurity-protect-me.html.

PwC. (2018, 28th of October). Dataangrep rammer 7 av 10 bedrifter. Retrieved from https://www.pwc.no/no/publikasjoner/cybercrime-survey-2019.html.

Reitberger, G. & Wetzel, S. (2017). Investigating the impact of media coverage on data breach fatigue. *2017 IEEE 38th Sarnoff Symposium*. https://doi.org/10.1109/SARNOF.2017.8080399.

Riley, M., Robertson, J. & Sharpe, A. (2017, 29th of September). The Equifax Hack Has the Hallmarks of State-Sponsored Pros. *Bloomberg Businessweek*. Retrieved 04.11.20 from https://www.bloomberg.com/news/features/2017-09-29/the-equifax-hack-has-all-the-hallmarks-of-state-sponsored-pros.

Riordan, M. (2004, 30th of April). The Lost History of the Transistor. Retrieved 28.10.20 from https://spectrum.ieee.org/tech-history/silicon-revolution/the-lost-history-of-the-transistor.

RiskBased Security. (2014, December 5th). A Breakdown and Analysis of the December, 2014 Sony Hack. Retrieved 15.11.20 from https://www.riskbasedsecurity.com/2014/12/05/a-breakdown-and-analysis-of-the-december-2014-sony-hack/.

Rogers, S. (2016, 1st of July). What is Google Trends data – and what does it mean? Retrieved 10.11.20 from https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8.

SEC. (2020, 23rd of October). Spotlight on Cybersecurity, the SEC and You. Retrieved 29.10.20 from https://www.sec.gov/spotlight/cybersecurity.

Sharpe, W., F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance, 19* (3), 425 – 442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Spring, T. (2016, 6th of October). Web-Based Keylogger Used to Steal Credit Card Data from Popular Sites. Retrieved 10.11.20 from https://threatpost.com/web-based-keylogger-used-to-steal-credit-card-data-from-popular-sites/121141/.

Statista. (2020, October). Annual number of data breaches and exposed records in the United States from 2005 to 1st half 2020. Retrieved 20.10.20 from https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/.

Statt, N. (2018, 29th of March). Under Armour says 150 million MyFitnessPal accounts compromised in data breach. *The Verge*. Retrieved 15.11.20 from https://www.theverge.com/2018/3/29/17177848/under-armour-myfitnesspal-data-breach-150-million-accounts-security.

The World Bank. (2020). Individuals using the Internet (% of population). Retrieved 08.11.20 from https://databank.worldbank.org/source/world-development-indicators#.

Tunggal, A., T. (2020, 2nd of October). What is a Computer Worm? Retrieved 22.10.20 from https://www.upguard.com/blog/computer-worm.

van der Kleut, J. (2020). What is a keylogger and how do I protect myself against one? Retrieved 22.10.20 from https://us.norton.com/internetsecurity-malware-what-is-a-keylogger.html.

11. Appendix

Definition of selected methods that can be used in a data breach.

Phishing*:* Phishing is a type of social engineering attack commonly used to steal login credentials and user data. Cybercriminals pretend to be a trusted entity and often mimic a target site, which can deceive individuals to give away sensitive or confidential information such as login credentials (Porter, 2020). Phishing can be targeted at random individuals or more targeted in the form of some selected individuals or companies. According to PwC (2018), phishing is one of the most common and widespread methods used in a data breach. Based on our experiences, phishing is a method that is being used to a greater extent, and there has been an increase in the use in the last years.

Brute force*:* Brute force is when a cybercriminal uses a combination of letters, numbers, and unique signs to guess an individual's password through trial-and-error (Kaspersky, 2020). As this is a time-consuming process, cybercriminals have begun using lists of passwords and emails instead. The lists are commonly based on breach compilations, which are disclosed passwords and emails in former data breaches.

Malware*:* Malware is a relatively broad category and includes several methods such as spyware, trojans, computer worms, keyloggers and ransomware. Spyware is a broad category of software that infiltrates devices and are commonly used to steal sensitive and protected data from computers and other devices (Norton, 2019). Trojans are malware disguised as legitimate software that potentially can take control over computers and are designed to generally inflict harmful actions, such as stealing digital information (Johansen, 2020). Computer worms are malicious software designed to steal sensitive data, corporate espionage, or exploit known security vulnerabilities. They pose a big threat as they can self-replicate to computers on the same network (Tunggal, 2020).

Keyloggers are used to track and log the keys users strike on the keyboard, consequently capturing any typed information. The data captured can, for example, be login details or financial information such as credit card numbers (van der Kleut, 2020; Spring, 2016). Ransomware is a type of malicious software that infects a computer or system and demands individuals to pay a ransom (often in a virtual currency) to restore access. However, it is worth noticing that paying the ransom will not necessarily ensure restoration of access and that the payment can lead to the cybercriminals demanding an even higher ransom (Johansen, 2019).

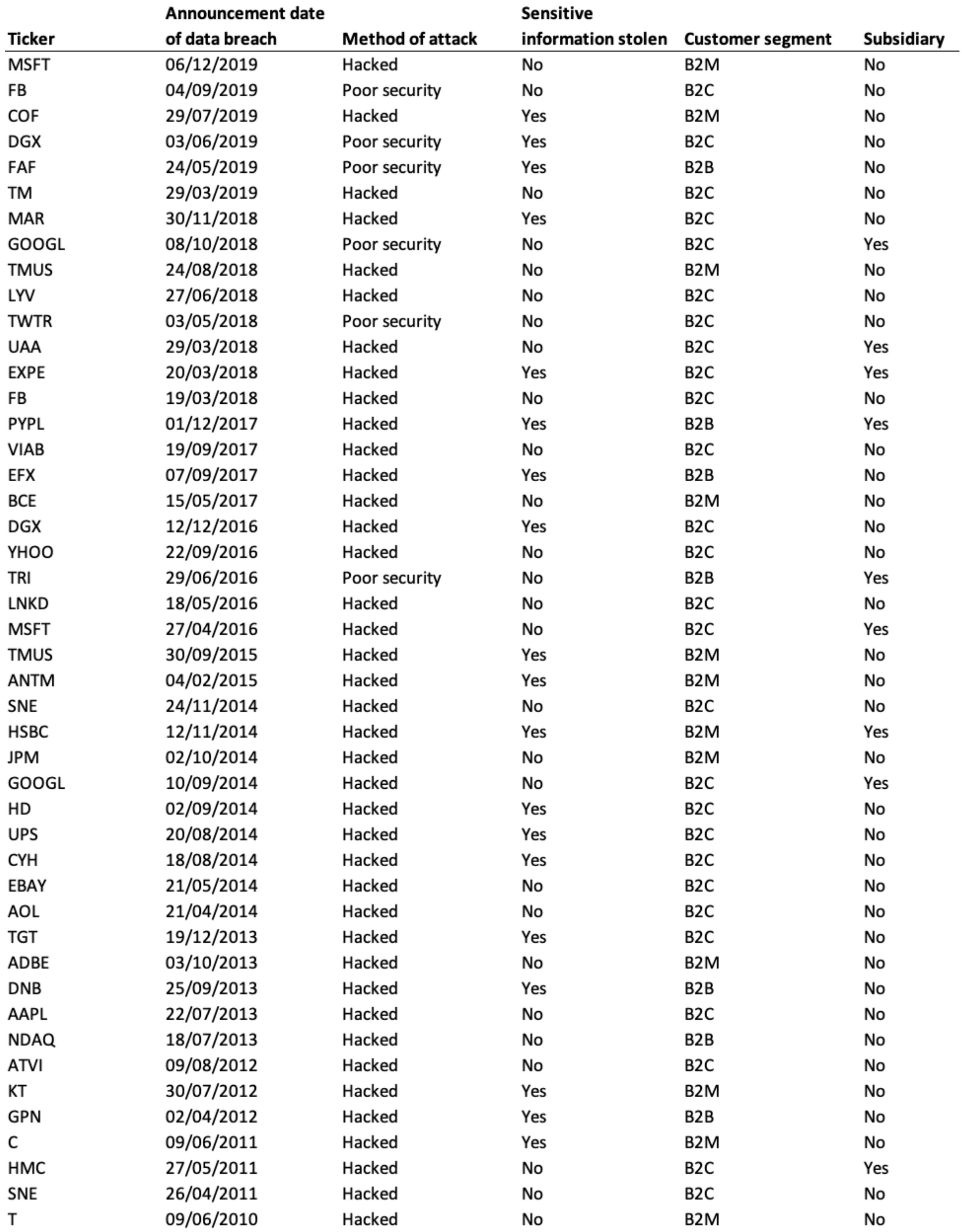
Table 7: Data breaches on US listed firms, 2010 – 2019.

This table illustrates data breaches included in the sample in addition to the number of records stolen, industry and announcement date of the data breach.

Target company	Ticker	Industry	Records stolen	Announcement date of data breach
Microsoft	MSFT	Technology	44 000 000	06/12/2019
Facebook	FB	Media	419 000 000	04/09/2019
Capital One	COF	Finance	100 000 000	29/07/2019
Quest Diagnostics	DGX	Pharmaceuticals	11 900 000	03/06/2019
First American Financial Corporation	FAF	Finance	885 000 000	24/05/2019
Toyota	TM	Consumer	3 100 000	29/03/2019
Marriott International	MAR	Hospitality	383 000 000	30/11/2018
Google+	GOOGL	Media	52 500 000	08/10/2018
T-Mobile	TMUS	Media	2 000 000	24/08/2018
Ticketmaster	LYV	Entertainment	40 000	27/06/2018
Twitter	TWTR	Media	330 000 000	03/05/2018
MyFitnessPal	UAA	Consumer	150 000 000	29/03/2018
Orbitz	EXPE	Services	880 000	20/03/2018
Facebook	FB	Media	50 000 000	19/03/2018
TIO Networks	PYPL	Media	1 600 000	01/12/2017
Viacom	VIAB	Media	3 000 000	19/09/2017
Equifax	EFX	Finance	143 000 000	07/09/2017
Bell	BCE	Consumer	1 900 000	15/05/2017
Quest Diagnostics	DGX	Pharmaceuticals	34 000	12/12/2016
Yahoo	YHOO	Media	500 000 000	22/09/2016
World Check	TRI	Communication	2 200 000	29/06/2016
LinkedIn	LNKD	Media	117 000 000	18/05/2016
Minecraft	MSFT	Technology	7 000 000	27/04/2016
T-mobile	TMUS	Media	15 000 000	30/09/2015
Anthem	ANTM	Finance	80 000 000	04/02/2015
Sony Pictures	SNE	Technology	10 000 000	24/11/2014
HSBC Turkey	HSBC	Finance	2 700 000	12/11/2014
JP Morgan Chase	JPM	Finance	76 000 000	02/10/2014
Gmail	GOOGL	Media	5 000 000	10/09/2014
Home Depot	HD	Retail	56 000 000	02/09/2014
UPS	UPS	Transportation	4 000 000	20/08/2014
Community Health Systems	CYH	Healthcare	4 500 000	18/08/2014
Ebay	EBAY	Retail	145 000 000	21/05/2014
AOL	AOL	Media	2 400 000	21/04/2014
Target	TGT	Retail	70 000 000	19/12/2013
Adobe	ADBE	Technology	36 000 000	03/10/2013
D&B, Altegrity	DNB	Services	1 000 000	25/09/2013
Apple	AAPL	Technology	275 000	22/07/2013
NASDAQ	NDAQ	Services	500 000	18/07/2013
Activision Blizzard	ATVI	Entertainment	14 000 000	09/08/2012
KT Corp.	KT	Media	8 700 000	30/07/2012
Global Payments	GPN	Services	7 000 000	02/04/2012
Citigroup	C	Finance	360 083	09/06/2011
Honda Canada	HMC	Consumer	283 000	27/05/2011
Sony PSN	SNE	Technology	77 000 000	26/04/2011
AT&T	T	Media	114 000	09/06/2010

Table 8: Characteristics of data breaches on US listed firms, 2010 – 2019.

This table illustrates the characteristics of the data breaches included in the sample.

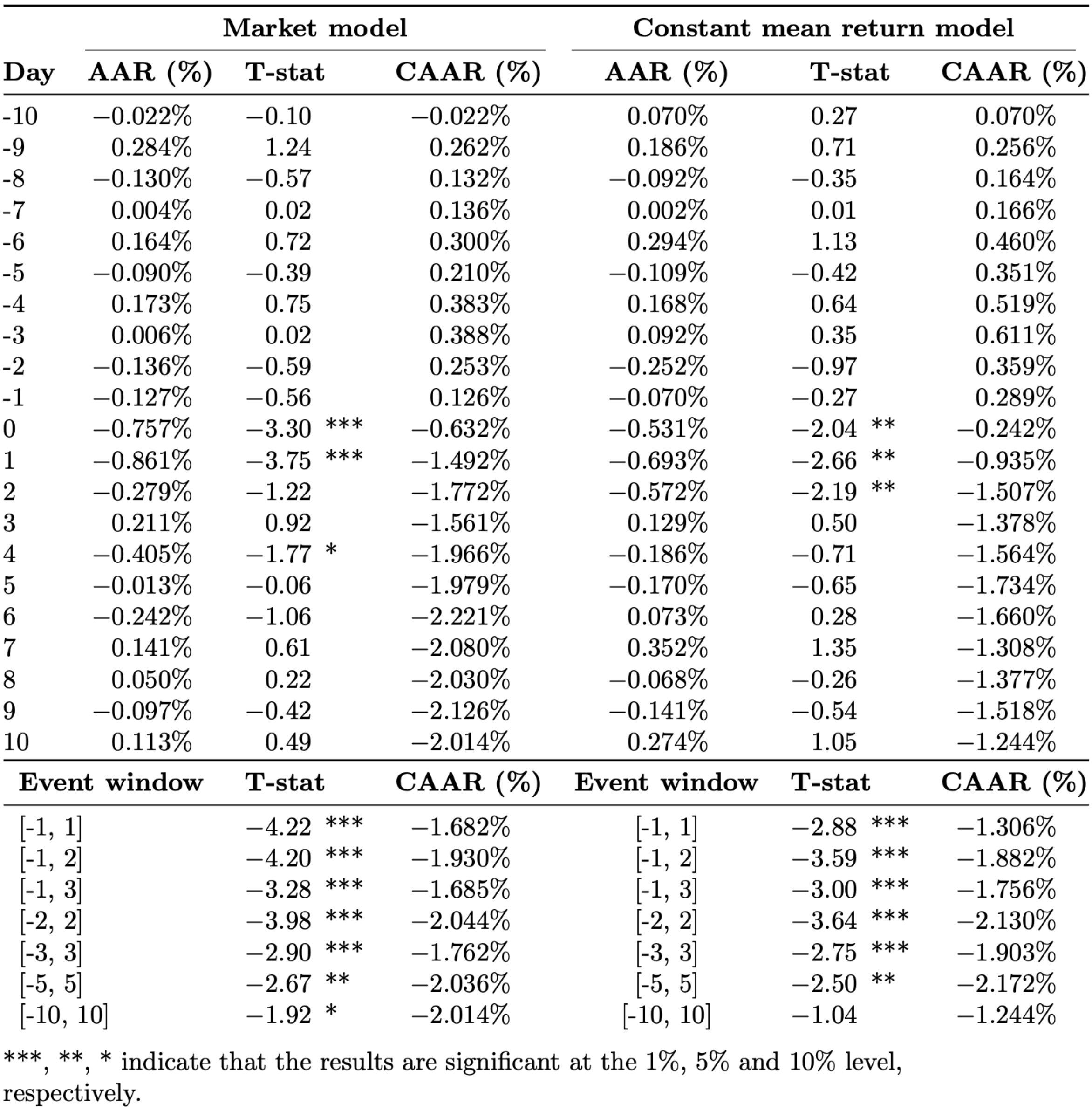


Ticker	Announcement date of data breach	Method of attack	Sensitive information stolen	Customer segment	Subsidiary
MSFT	06/12/2019	Hacked	No	B2M	No
FB	04/09/2019	Poor security	No	B2C	No
COF	29/07/2019	Hacked	Yes	B2M	No
DGX	03/06/2019	Poor security	Yes	B2C	No
FAF	24/05/2019	Poor security	Yes	B2B	No
TM	29/03/2019	Hacked	No	B2C	No
MAR	30/11/2018	Hacked	Yes	B2C	No
GOOGL	08/10/2018	Poor security	No	B2C	Yes
TMUS	24/08/2018	Hacked	No	B2M	No
LYV	27/06/2018	Hacked	No	B2C	No
TWTR	03/05/2018	Poor security	No	B2C	No
UAA	29/03/2018	Hacked	No	B2C	Yes
EXPE	20/03/2018	Hacked	Yes	B2C	Yes
FB	19/03/2018	Hacked	No	B2C	No
PYPL	01/12/2017	Hacked	Yes	B2B	Yes
VIAB	19/09/2017	Hacked	No	B2C	No
EFX	07/09/2017	Hacked	Yes	B2B	No
BCE	15/05/2017	Hacked	No	B2M	No
DGX	12/12/2016	Hacked	Yes	B2C	No
YHOO	22/09/2016	Hacked	No	B2C	No
TRI	29/06/2016	Poor security	No	B2B	Yes
LNKD	18/05/2016	Hacked	No	B2C	No
MSFT	27/04/2016	Hacked	No	B2C	Yes
TMUS	30/09/2015	Hacked	Yes	B2M	No
ANTM	04/02/2015	Hacked	Yes	B2M	No
SNE	24/11/2014	Hacked	No	B2C	No
HSBC	12/11/2014	Hacked	Yes	B2M	Yes
JPM	02/10/2014	Hacked	No	B2M	No
GOOGL	10/09/2014	Hacked	No	B2C	Yes
HD	02/09/2014	Hacked	Yes	B2C	No
UPS	20/08/2014	Hacked	Yes	B2C	No
CYH	18/08/2014	Hacked	Yes	B2C	No
EBAY	21/05/2014	Hacked	No	B2C	No
AOL	21/04/2014	Hacked	No	B2C	No
TGT	19/12/2013	Hacked	Yes	B2C	No
ADBE	03/10/2013	Hacked	No	B2M	No
DNB	25/09/2013	Hacked	Yes	B2B	No
AAPL	22/07/2013	Hacked	No	B2C	No
NDAQ	18/07/2013	Hacked	No	B2B	No
ATVI	09/08/2012	Hacked	No	B2C	No
KT	30/07/2012	Hacked	Yes	B2M	No
GPN	02/04/2012	Hacked	Yes	B2B	No
C	09/06/2011	Hacked	Yes	B2M	No
HMC	27/05/2011	Hacked	No	B2C	Yes
SNE	26/04/2011	Hacked	No	B2C	No
T	09/06/2010	Hacked	No	B2M	No

Table 9: Price effects around the announcement date of a data breach, 2010 – 2019.

Average abnormal return (AAR) and cumulative abnormal return (CAAR) over the [-10, 10] event window. The sample consists of 46 data breaches with more than 30 000 records stolen on US listed firms between 2010 and 2019. The announcement date of the data breach represents day 0. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ and the constant mean return model $R_{i,t} = \mu_{i,t} + \varepsilon_{i,t}$ is used as the normal return models with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. The abnormal return is computed as:

$\widehat{AR}_{i,t} = R_{i,t} - E(R_{i,t}[Normal\ return\ model]).$

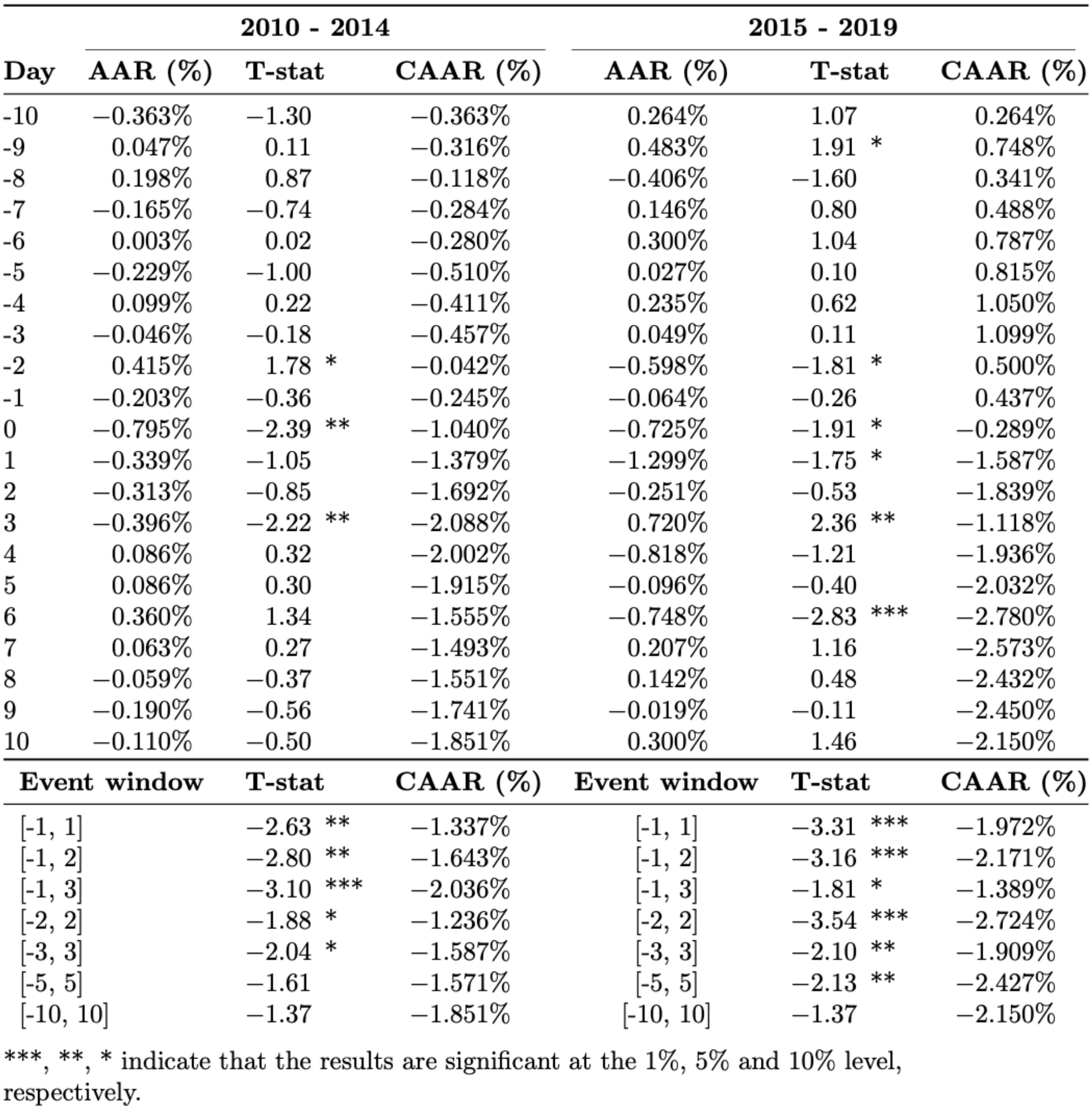


	Market model			Constant mean return model		
Day	AAR (%)	T-stat	CAAR (%)	AAR (%)	T-stat	CAAR (%)
-10	−0.022%	−0.10	−0.022%	0.070%	0.27	0.070%
-9	0.284%	1.24	0.262%	0.186%	0.71	0.256%
-8	−0.130%	−0.57	0.132%	−0.092%	−0.35	0.164%
-7	0.004%	0.02	0.136%	0.002%	0.01	0.166%
-6	0.164%	0.72	0.300%	0.294%	1.13	0.460%
-5	−0.090%	−0.39	0.210%	−0.109%	−0.42	0.351%
-4	0.173%	0.75	0.383%	0.168%	0.64	0.519%
-3	0.006%	0.02	0.388%	0.092%	0.35	0.611%
-2	−0.136%	−0.59	0.253%	−0.252%	−0.97	0.359%
-1	−0.127%	−0.56	0.126%	−0.070%	−0.27	0.289%
0	−0.757%	−3.30 ***	−0.632%	−0.531%	−2.04 **	−0.242%
1	−0.861%	−3.75 ***	−1.492%	−0.693%	−2.66 **	−0.935%
2	−0.279%	−1.22	−1.772%	−0.572%	−2.19 **	−1.507%
3	0.211%	0.92	−1.561%	0.129%	0.50	−1.378%
4	−0.405%	−1.77 *	−1.966%	−0.186%	−0.71	−1.564%
5	−0.013%	−0.06	−1.979%	−0.170%	−0.65	−1.734%
6	−0.242%	−1.06	−2.221%	0.073%	0.28	−1.660%
7	0.141%	0.61	−2.080%	0.352%	1.35	−1.308%
8	0.050%	0.22	−2.030%	−0.068%	−0.26	−1.377%
9	−0.097%	−0.42	−2.126%	−0.141%	−0.54	−1.518%
10	0.113%	0.49	−2.014%	0.274%	1.05	−1.244%
Event window		T-stat	CAAR (%)	Event window	T-stat	CAAR (%)
[-1, 1]		−4.22 ***	−1.682%	[-1, 1]	−2.88 ***	−1.306%
[-1, 2]		−4.20 ***	−1.930%	[-1, 2]	−3.59 ***	−1.882%
[-1, 3]		−3.28 ***	−1.685%	[-1, 3]	−3.00 ***	−1.756%
[-2, 2]		−3.98 ***	−2.044%	[-2, 2]	−3.64 ***	−2.130%
[-3, 3]		−2.90 ***	−1.762%	[-3, 3]	−2.75 ***	−1.903%
[-5, 5]		−2.67 **	−2.036%	[-5, 5]	−2.50 **	−2.172%
[-10, 10]		−1.92 *	−2.014%	[-10, 10]	−1.04	−1.244%

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 10: Price effects around the announcement date of a data breach in the time periods 2010 – 2014 and 2015 – 2019.

Average abnormal return (AAR) and cumulative abnormal return (CAAR) over the [-10, 10] event window in the two time periods 2010 – 2014 and 2015 – 2019. The sample consists of 22 data breaches between 2010 and 2014 and 25 between 2015 and 2019 for US listed firms with more than 30 000 records stolen. The announcement date of the data breach represents day 0. The market model $R_{i,t} = \alpha_i + \beta_i * R_{m,t} + \varepsilon_{i,t}$ is used as the normal return model with parameters estimated in the estimation window using weekly data over two years and a holdout window of 20 trading days. The abnormal return is computed as: $\widehat{AR}_{i,t} = R_{i,t} - E\left(R_{i,t}[Normal\ return\ model]\right)$.



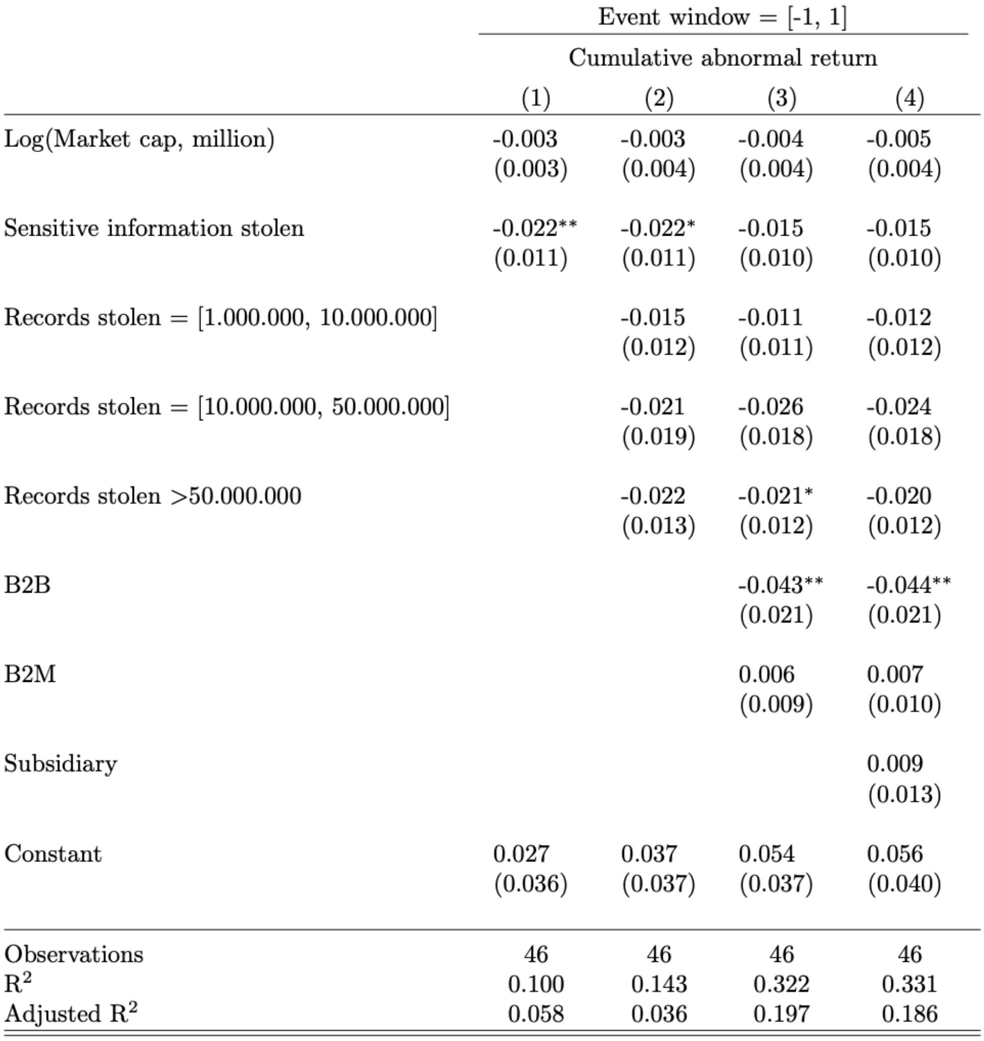
	2010 - 2014			2015 - 2019		
Day	**AAR (%)**	**T-stat**	**CAAR (%)**	**AAR (%)**	**T-stat**	**CAAR (%)**
-10	−0.363%	−1.30	−0.363%	0.264%	1.07	0.264%
-9	0.047%	0.11	−0.316%	0.483%	1.91 *	0.748%
-8	0.198%	0.87	−0.118%	−0.406%	−1.60	0.341%
-7	−0.165%	−0.74	−0.284%	0.146%	0.80	0.488%
-6	0.003%	0.02	−0.280%	0.300%	1.04	0.787%
-5	−0.229%	−1.00	−0.510%	0.027%	0.10	0.815%
-4	0.099%	0.22	−0.411%	0.235%	0.62	1.050%
-3	−0.046%	−0.18	−0.457%	0.049%	0.11	1.099%
-2	0.415%	1.78 *	−0.042%	−0.598%	−1.81 *	0.500%
-1	−0.203%	−0.36	−0.245%	−0.064%	−0.26	0.437%
0	−0.795%	−2.39 **	−1.040%	−0.725%	−1.91 *	−0.289%
1	−0.339%	−1.05	−1.379%	−1.299%	−1.75 *	−1.587%
2	−0.313%	−0.85	−1.692%	−0.251%	−0.53	−1.839%
3	−0.396%	−2.22 **	−2.088%	0.720%	2.36 **	−1.118%
4	0.086%	0.32	−2.002%	−0.818%	−1.21	−1.936%
5	0.086%	0.30	−1.915%	−0.096%	−0.40	−2.032%
6	0.360%	1.34	−1.555%	−0.748%	−2.83 ***	−2.780%
7	0.063%	0.27	−1.493%	0.207%	1.16	−2.573%
8	−0.059%	−0.37	−1.551%	0.142%	0.48	−2.432%
9	−0.190%	−0.56	−1.741%	−0.019%	−0.11	−2.450%
10	−0.110%	−0.50	−1.851%	0.300%	1.46	−2.150%

Event window	**T-stat**	**CAAR (%)**	**Event window**	**T-stat**	**CAAR (%)**
[-1, 1]	−2.63 **	−1.337%	[-1, 1]	−3.31 ***	−1.972%
[-1, 2]	−2.80 **	−1.643%	[-1, 2]	−3.16 ***	−2.171%
[-1, 3]	−3.10 ***	−2.036%	[-1, 3]	−1.81 *	−1.389%
[-2, 2]	−1.88 *	−1.236%	[-2, 2]	−3.54 ***	−2.724%
[-3, 3]	−2.04 *	−1.587%	[-3, 3]	−2.10 **	−1.909%
[-5, 5]	−1.61	−1.571%	[-5, 5]	−2.13 **	−2.427%
[-10, 10]	−1.37	−1.851%	[-10, 10]	−1.37	−2.150%

***, **, * indicate that the results are significant at the 1%, 5% and 10% level, respectively.

Table 11: Regressions of CAR for the event window [-1, 1].

Cross-sectional regressions where the dependent variable is the cumulative abnormal return (CAR) for the data breaches in the sample on the event window [-1, 1]. The market capital is defined as the market value fifteen days before the data breach. Sensitive information stolen is equal to one if either SSNs, credit card numbers, financial information or healthcare information were stolen. The records stolen variable is an indicator variable indicating the number of records stolen in the data breach, and the baseline in the regression is that less than 1 000 000 records were stolen. B2B and B2M is an indicator variable indicating if the company are in the business to business or business to many customer segments, respectively. The baseline customer segment in the regression is the business to customer segment. Subsidiary is equal to one if the company are a subsidiary. The standard errors (in parentheses) in the regression are bootstrapped at 10 000 replacements. *, **, *** denote that the variable is statistically significant at the 10%, 5% and 1% levels, respectively.



	Event window = [-1, 1]			
	Cumulative abnormal return			
	(1)	(2)	(3)	(4)
Log(Market cap, million)	-0.003 (0.003)	-0.003 (0.004)	-0.004 (0.004)	-0.005 (0.004)
Sensitive information stolen	-0.022** (0.011)	-0.022* (0.011)	-0.015 (0.010)	-0.015 (0.010)
Records stolen = [1.000.000, 10.000.000]		-0.015 (0.012)	-0.011 (0.011)	-0.012 (0.012)
Records stolen = [10.000.000, 50.000.000]		-0.021 (0.019)	-0.026 (0.018)	-0.024 (0.018)
Records stolen >50.000.000		-0.022 (0.013)	-0.021* (0.012)	-0.020 (0.012)
B2B			-0.043** (0.021)	-0.044** (0.021)
B2M			0.006 (0.009)	0.007 (0.010)
Subsidiary				0.009 (0.013)
Constant	0.027 (0.036)	0.037 (0.037)	0.054 (0.037)	0.056 (0.040)
Observations	46	46	46	46
R^2	0.100	0.143	0.322	0.331
Adjusted R^2	0.058	0.036	0.197	0.186

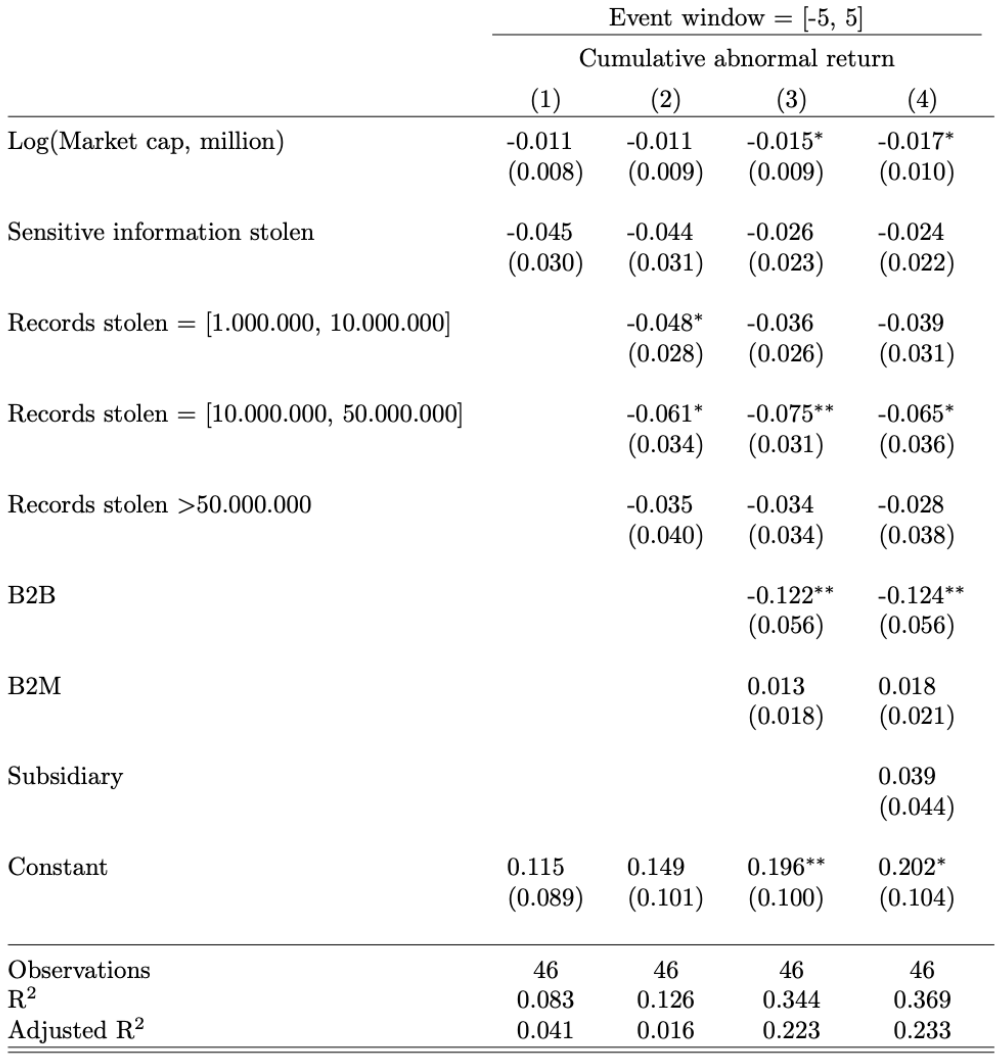
Table 12: Regressions of CAR for the event window [-2, 2].

Cross-sectional regressions where the dependent variable is the cumulative abnormal return (CAR) for the data breaches in the sample on the event window [-2, 2]. The market capital is defined as the market value fifteen days before the data breach. Sensitive information stolen is equal to one if either SSNs, credit card numbers, financial information or healthcare information were stolen. The records stolen variable is an indicator variable indicating the number of records stolen in the data breach, and the baseline in the regression is that less than 1 000 000 records were stolen. B2B and B2M is an indicator variable indicating if the company are in the business to business or business to many customer segments, respectively. The baseline customer segment in the regression is the business to customer segment. Subsidiary is equal to one if the company are a subsidiary. The standard errors (in parentheses) in the regression are bootstrapped at 10 000 replacements. *, **, *** denote that the variable is statistically significant at the 10%, 5% and 1% levels, respectively.

	Event window = [-2, 2]			
	Cumulative abnormal return			
	(1)	(2)	(3)	(4)
Log(Market cap, million)	-0.005	-0.004	-0.006	-0.007
	(0.004)	(0.005)	(0.005)	(0.005)
Sensitive information stolen	-0.032*	-0.031*	-0.020	-0.019
	(0.017)	(0.017)	(0.013)	(0.012)
Records stolen = [1.000.000, 10.000.000]		-0.030**	-0.023*	-0.025*
		(0.015)	(0.013)	(0.015)
Records stolen = [10.000.000, 50.000.000]		-0.040**	-0.047**	-0.041**
		(0.020)	(0.019)	(0.020)
Records stolen >50.000.000		-0.035*	-0.035**	-0.032*
		(0.020)	(0.017)	(0.018)
B2B			-0.069**	-0.070**
			(0.032)	(0.032)
B2M			0.0002	0.003
			(0.010)	(0.011)
Subsidiary				0.021
				(0.020)
Constant	0.040	0.061	0.085*	0.088*
	(0.046)	(0.052)	(0.050)	(0.052)
Observations	46	46	46	46
R^2	0.101	0.167	0.374	0.396
Adjusted R^2	0.059	0.063	0.258	0.266

Table 13: Regressions of CAR for the event window [-5, 5].

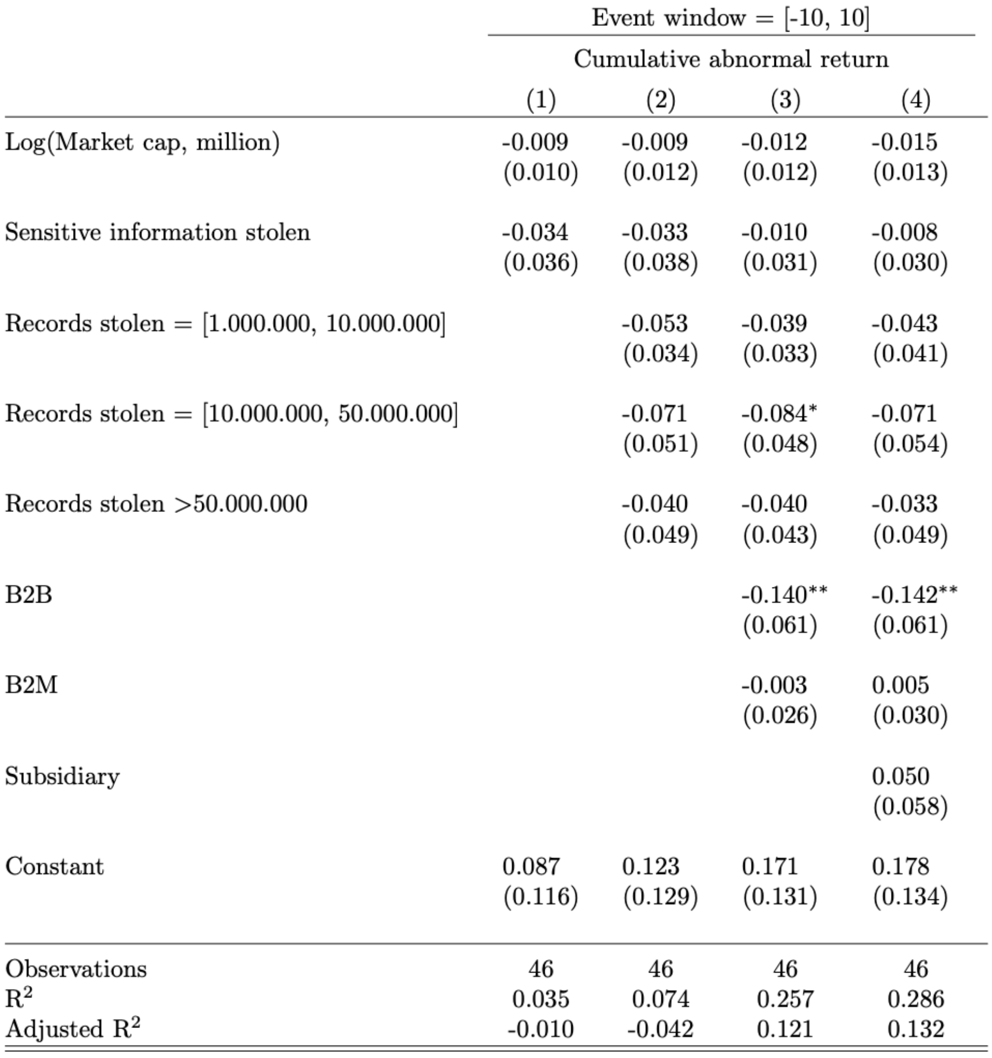
Cross-sectional regressions where the dependent variable is the cumulative abnormal return (CAR) for the data breaches in the sample on the event window [-5, 5]. The market capital is defined as the market value fifteen days before the data breach. Sensitive information stolen is equal to one if either SSNs, credit card numbers, financial information or healthcare information were stolen. The records stolen variable is an indicator variable indicating the number of records stolen in the data breach, and the baseline in the regression is that less than 1 000 000 records were stolen. B2B and B2M is an indicator variable indicating if the company are in the business to business or business to many customer segments, respectively. The baseline customer segment in the regression is the business to customer segment. Subsidiary is equal to one if the company are a subsidiary. The standard errors (in parentheses) in the regression are bootstrapped at 10 000 replacements. *, **, *** denote that the variable is statistically significant at the 10%, 5% and 1% levels, respectively.



	Event window = [-5, 5]			
	Cumulative abnormal return			
	(1)	(2)	(3)	(4)
Log(Market cap, million)	-0.011	-0.011	-0.015*	-0.017*
	(0.008)	(0.009)	(0.009)	(0.010)
Sensitive information stolen	-0.045	-0.044	-0.026	-0.024
	(0.030)	(0.031)	(0.023)	(0.022)
Records stolen = [1.000.000, 10.000.000]		-0.048*	-0.036	-0.039
		(0.028)	(0.026)	(0.031)
Records stolen = [10.000.000, 50.000.000]		-0.061*	-0.075**	-0.065*
		(0.034)	(0.031)	(0.036)
Records stolen >50.000.000		-0.035	-0.034	-0.028
		(0.040)	(0.034)	(0.038)
B2B			-0.122**	-0.124**
			(0.056)	(0.056)
B2M			0.013	0.018
			(0.018)	(0.021)
Subsidiary				0.039
				(0.044)
Constant	0.115	0.149	0.196**	0.202*
	(0.089)	(0.101)	(0.100)	(0.104)
Observations	46	46	46	46
R^2	0.083	0.126	0.344	0.369
Adjusted R^2	0.041	0.016	0.223	0.233

Table 14: Regressions of CAR for the event window [-10, 10].

Cross-sectional regressions where the dependent variable is the cumulative abnormal return (CAR) for the data breaches in the sample on the event window [-10, 10]. The market capital is defined as the market value fifteen days before the data breach. Sensitive information stolen is equal to one if either SSNs, credit card numbers, financial information or healthcare information were stolen. The records stolen variable is an indicator variable indicating the number of records stolen in the data breach, and the baseline in the regression is that less than 1 000 000 records were stolen. B2B and B2M is an indicator variable indicating if the company are in the business to business or business to many customer segments, respectively. The baseline customer segment in the regression is the business to customer segment. Subsidiary is equal to one if the company are a subsidiary. The standard errors (in parentheses) in the regression are bootstrapped at 10 000 replacements. *, **, *** denote that the variable is statistically significant at the 10%, 5% and 1% levels, respectively.



	Event window = [-10, 10]			
	Cumulative abnormal return			
	(1)	(2)	(3)	(4)
Log(Market cap, million)	-0.009	-0.009	-0.012	-0.015
	(0.010)	(0.012)	(0.012)	(0.013)
Sensitive information stolen	-0.034	-0.033	-0.010	-0.008
	(0.036)	(0.038)	(0.031)	(0.030)
Records stolen = [1.000.000, 10.000.000]		-0.053	-0.039	-0.043
		(0.034)	(0.033)	(0.041)
Records stolen = [10.000.000, 50.000.000]		-0.071	-0.084*	-0.071
		(0.051)	(0.048)	(0.054)
Records stolen >50.000.000		-0.040	-0.040	-0.033
		(0.049)	(0.043)	(0.049)
B2B			-0.140**	-0.142**
			(0.061)	(0.061)
B2M			-0.003	0.005
			(0.026)	(0.030)
Subsidiary				0.050
				(0.058)
Constant	0.087	0.123	0.171	0.178
	(0.116)	(0.129)	(0.131)	(0.134)
Observations	46	46	46	46
R^2	0.035	0.074	0.257	0.286
Adjusted R^2	-0.010	-0.042	0.121	0.132

Table 15*:* Definition of the industries in the sample.

The industries in the table below refer to the industries in the sample and the sectors they include. The industries are inspired by the cost of a Data Breach Report 2020 written by the Ponemon Institute.

Industry	Definition
Consumer	Manufacturers and distributors of consumer products
Finance	Banking, insurance, investment companies
Media	Television, satellite, social media, Internet
Retail	Brick and mortar and e-commerce
Services	Professional services such as legal, accounting and consulting firms
Technology	Software and hardware companies
Other	Communication, entertainment, healthcare, hospitality, pharmaceuticals transportation

	Industries within Other
Communication	Newspapers, book publishers, public relations and advertising
Entertainment	Movie production, sports, gaming and casinos
Healthcare	Hospitals, clinics
Hospitality	Hotels, restaurant chains, cruise lines
Pharmaceuticals	Pharmaceuticals and biomedicine
Transportation	Airlines, railroad, trucking and delivery companies

Table 16: Overview and definition of event windows

The table below gives an explanation of the event windows analyzed in the thesis. The event windows refer to the number of trading days relative to the announcement date of the data breach. [-1, 2] for example, refer to an event window beginning the trading day before and ending two trading days after the announcement of the data breach.

Event window	Definition
[-1, 1]	One trading day before to one trading day after the data breach.
[-1, 2]	One trading day before to two trading days after the data breach.
[-1, 3]	One trading day before to three trading days after the data breach.
[-2, 2]	Two trading days before to two trading days after the data breach.
[-3, 3]	Three trading days before to three trading days after the data breach.
[-5, 5]	Five trading days before to five trading days after the data breach.
[-10, 10]	Ten trading days before to ten trading days after the data breach.