ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialisation Financial Economics

Investors' Anticipation Of Data Breaches Effects On Public Stock Prices

Author: Y. Van Dongen
Student number: 407420
Thesis supervisor: Dr. J.J.G. Lemmen
Finish date: August 2016

PREFACE AND ACKNOWLEDGEMENTS

First of all, I would like to thank the companies in this study and the newspapers who made the data breach public in order to make this study possible. I would also like to recommend the companies to disclose their data breaches in order to help others to fight the data breach war. This is a fight where everybody has to stand together and help each other so that companies can better protect themselves and their customers.

I would also like to thank my supervisor, Dr. Jan Lemmen, for all the precious advice that he has given to me in order to improve my thesis and all the time he has accorded me in providing new ideas and articles related to data breaches. I would never have been able to fulfil this thesis without his help.

Special thanks also to the staff and teachers from Erasmus University of Rotterdam who have helped me through the years by giving me the knowledge necessary to achieve this thesis.

Last but not least, I would like to thank my family for all the mental support. This support has given me strength to be able to fulfil things to the best of my abilities.

ABSTRACT

This paper examines the effects of data breaches on stock prices in order to investigate if investors account for those breaches in stock price. This study has been performed with a sample of 107 firms breached and 128 events between 2006 and 2015 through an event study. Different firm characteristics have been taking into consideration in this study to observe if investors pay more attention to a particular size or industry. The results show that investors do take into account data breaches in general, however they still punish firms in certain specific sectors such as the financial industry and firms whose protection is not evolving at the same speed as their business. The reason for this is that investors have seen the data breaches rising over the last years and have finally decided that they cannot neglect cybercrime. However, some specific characteristics are not anticipated by the investors due to the fact that they are too important and too sensitive to simply account for it in the stock price.

JEL Classification: G14

Keywords: Data breaches, Cybercrime, Event study, Efficient market hypothesis, Cumulative abnormal return

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

Internet has revolutionized our time. It makes our world smaller, information easier and more accessible to obtain. Internet does not only help individuals, but also companies by making transactions easier to pursue wherever in the world their clients could be. Therefore, internet has largely contributed to globalization. Such accessibility and freedom to a "new world" is very convenient, but is automatically linked with new security threats.

There are about 3,5 billion internet users in the world today. People nowadays cannot live without internet anymore and have often more than one device connected to the internet, which increases even more the risk of becoming a victim of "cybercrime". The dictionary defines Cybercrime as a "crime (as theft, fraud, intellectual property violations, or distribution of child pornography) committed electronically" (Merriam-Webster, 2016). A hacker is the one who commits the act of cybercrime, and this would lead to a data breach for a company. A data breach as Harvard's business review defines it, is an "incident in which sensitive, protected, or confidential data has potentially been viewed, stolen, or used by an individual unauthorized to do so" (Savoie, 2012). Committing cybercrime has two main goals to the hacker. His first goal is to hamper the company and disrupt their ability to provide services to their customers and secondly is to cause economic losses. On the companies' side, this is not the finish line. It has to cope with the loss of trust and reputation and also go through diverse lawsuits. Examples of data breaches are theft and modification of computer software's, manipulation and destruction of data, shut down of computer services and illegally accessing passwords. Those situations do not always have to be intentional, but can also occur accidentally, loss of laptop for example. The perpetrator does not always have to be someone external to the firm, but can also be someone working already inside the firm (Power, 2002; Howard, 1997).

Lyne (2013), a hacking specialist, says that these threats are increasing every day, with about 250.000 new viruses daily and around 300.000 contaminated websites daily. The creation of viruses never stops and becomes easier and more available to the public market. The methods of hacking also change constantly. The first type of attacks was called "phishing" attacks. Kapersky lab (2013), famous for their research against cybercrime, defines phishing as follows: "Phishing is an email or electronic communications scam targeted towards a specific individual, organization or business". For example, an email arrives from someone's bank or a trustworthy source, which provides a link to a website filled with malware, with the goal to obtain bank codes or direct access to the data on the computer. However, this trend has gone down. In 2013, 1 out of 300 emails was a phishing email, while in 2015 this is only 1 out of 3000 (Symantec, 2016). Symantec states that this method is easy to get access to, since it costs between 2 USD and 10 USD and does not require any technical skills.

Subsequently, the hackers used "email malware". Malware are viruses, Trojan horses, bots… Norton (2016), a cybersecurity firm, defines it as "a category of malicious code that includes viruses, worms, and Trojan horses. Destructive malware will utilize popular communication tools to spread, including worms sent through email and instant messages". Therefore, email malware is the transmission of a malware through an attachment. Symantec states in their report that most malware are obtained by opening corrupted every day files (word files and excel files). Nowadays, this remain a medium effective way for cybercriminals to hack into firms. The average number of emails corrupted by email malware is 1 out of 200 between 2013 and 2015. However, this method is not as easy as phishing, since the hacker has to perform social engineering[1] on the victim, so he would agree on opening the attachment. For example, the hacker would first identify his target on social media, then call the person and make some advertisement about something the target likes, to finally send the victim the "corrupted brochure" by email.

Nowadays we are in a trend which is called "APT" (Advanced Persistent Threat), those are attacks zero-day based, which have never been seen before. Symantec defines this as "widespread, persistent, and sophisticated attacks against government organizations and business of all sizes pose greater risk to national security and the economy". Some of the victims of those attacks have been the White House, the Pentagon and the German Bundestag who lost 21.5 million personnel files. The amount of those attacks are very low, because these are very specific and strong attacks, however the amount has doubled from 24 in 2014 to 54 in 2015. Precautions have been taken by engaging cyber security firms that help the victims to respond as quickly as possible to those attacks. They have been able to reduce the time of exposure from 295 days in 2014 to 7 days in 2015. Those attacks were mainly due to false updates on the adobe flash player.

Companies try to protect themselves against those threats through cyber protection. This means they engage specialists, sometimes hackers in order to see if their system is easy to hack into. A lot depends on this, since a breached company loses on average 2.1% of its market capitalization per breach (Cavusoglu, 2004).

The investors should therefore be highly interested in knowing if the company invested in is well protected against cybercrime. The research on this very topic has been very scarce. Most of the literature links directly the stock prices to cybercrime announcement dates and observe which factors impact the stock price and by what weight. This paper is based on those findings and will analyse if

[1] A method of persuasion in order to obtain information from people exerting key positions within the company. (Hadnagy, 2010). "The use of psychological tricks, the manipulation of behavior often through deception, by cybercriminals on unsuspecting users to gain "access information"."(Chantler, 2006)

investors anticipate cybercrime attacks and account for the consequences of those attacks in the stock price. The main question of this paper is:

Do investors, when investing in a public firm, take into account the risk of the firm's data breaches?

In order to answer this question, different variables from previous literature will be taken into account. Authors mostly do not agree upon the factors and the timing of the research make some variables more important than others as cybercrime evolves. The main variables found in the literature are the types of attack, the industry and the size of the firm. However, this will be extended by adding more variables such as continent, worldwide economic state, year, the number of times attacked and the recovery time, which might as well influence the investors.

The results prove that the investors do take data breaches into account when investing in public stocks. However, certain specific characteristics are not taken into account by the investors. This is because they do not want to anticipate the fact that sensitive information can be breached. When important information gets stolen from the financial industry this cannot stay unpunished and therefore the investors do not account for it before the breach. Even though some specific characteristics are purposely not anticipated, we still observe that the recovery time is one element they did not think about taking into account. This characteristic, which has a significant negative impact on the CAAR, could be anticipated by looking at the average time a peer in the same industry needs to solve the data breach.

This paper will be constructed as follows: First a review of the literature in Chapter two. Chapter three will present the hypothesis development. Chapter four presents the methodology and data that will be used to test if investors take data breaches into account. Chapter five shows the findings of the study. Chapter six summarizes the findings, concludes and gives directions for future research.

CHAPTER 2 Literature review

Research in the field of cybercrime has developed over time and the results have changed significantly. This is due to the fact that cybercrime is a very recent event and is constantly evolving. At the start of this phenomenon in 1990, it was not seen as a severe threat. However, over time the data leakages and information theft contained more important information and started to create fear for the customers. There is proof that breaches starting from 2000 affected the firms more negatively than before (Andobaidoo, 2010). The literature helps to understand the effect of the variables that affects the stock price due to data breaches over time.

2.1 Data breaches have only a short time negative effect

Kannan, Rees and Sridhar (2007) studied if the abnormal negative return of a data breach is only a short term effect or also a long term effect. The authors also wanted to observe if the following variables had an impact on the stock price: the size of the firm, the type of attack and the industry. They computed the CAR over 3-day, 8-day and 30-day windows to get the overall impact of security breach announcements. For such events, 3-days are seen as the short term effect while 8-days and 30-days are seen as the long term effect. They used a 30-day window since empirical rejection rates were found to exceed theoretical rates in the long run. Therefore, they used the Lyon and Barber (1999) methodology. This consists of a bootstrapped skewness-adjusted t-statistic.

They started with calculating the return for the sample using regression. The parameters were estimated using a 50-day window before the announcement date. Then, they calculated the three different CAR periods and verified using t-statistics if the mean CAR was significantly different from 0.

They concluded that the size of the firm on all those event windows were insignificant. This means that investor reactions do not differ between firm sizes. An explanation for this could be that investors are not aware of the firm's security investments, and it is hard to imagine for them what loss they could face or how secure the firm is. They also did not find any significant difference between the different attack types (Confidentiality, Integrity and Availability). Two reasons are given for this last observation. First, the sample size was very small, but a second reason could also be that the investors are indifferent towards the type of attack. Finally, no significant abnormal returns across any industry was observed.
They did not observe any negative reaction in the market in the long run (30-day window) on average over the period of 1997-2003. The negative cumulative abnormal returns ended at the 21st day after the announcement.

However, subsequently, lots of studies oppose this last finding by observing significantly different negative cumulative abnormal returns between firm characteristics and attack types. These studies have been made on U.S. companies only, by obtaining their data sample through newspapers: *Wall Street Journal*, *The New York Times*, *Financial Times*, *The Washington Post*, and *USA Today*. (Campbell et al., 2003; Morse et al., 2011; Andobaidoo et al., 2010)

2.2 Methods of attack

The first relevant study observing a significant different negative cumulative abnormal return between attack types has been done by Campbell, Gordon, Loeb, and Zhou (2003). They studied if the announcement of data breaches is related with negative abnormal return and looked if the amplitude changes by type of attack. Two types of attack have been defined. The first one is unauthorized access to confidential information, which corresponds to an external party who hacks into the company. On the other hand, there is non-confidential information, which represents simply a disruption of the current business without stealing anything.

In order to find this difference, they performed first the standard OLS methodology to estimate what the stock return would have been in the absence of the event. They also assumed that the daily stock returns are consistent with the CAPM (capital asset pricing model) and used a 120-day estimation period. This started 121 days before the security breach and ended 2 days before the announcement. They used a 120-day window based on their previous literature research in order to retain the most observations in their sample. As proxy for the market return they used the equally weighted NYSE/AMEX/NASDAQ market index return. Once those parameters were defined, they calculated the CAR based on a 3-day event window. They estimated that CAR (-1, 1) would be the best choice, because of possible information leakage, and also because a DOS (Denial of Service) attack might begin before the market closes and would then only be posted in the newspaper a day later. Finally, they used as well the SUR (seemingly unrelated regressions) methodology. They used this method because there is clustering of events in their sample (several companies were affected by the same virus the same day), as well as industry clustering. This leads to a violation of the standard OLS method. They obtained a data sample composed of public companies in the U.S. between 1995 and 2000 and obtained a final data set of 43 cases.

The main result of this paper is that not all data breaches have the same effect. There is no significant market reaction when the breach is not related to confidentiality, but when confidentiality is involved (theft of customer database) in the breach, then there is a significant negative market reaction. This negative market reaction leads on average to a loss of 5,5% of the firm value. This paper clearly underlines that breadth of press coverage is not the only driver of the stock market's reaction.

Hovav and D'Arcy (2003) focused on one type of attack which was the DOS attack. DOS attack is an attack that paralyzes the company by not being able to do anything, but does not involve any theft. Their study was to verify if DOS attacks leads to a decrease in market value of firms.
The authors decided to first estimate the returns on the firm's stocks victims of a DOS attack between 1998 and 2002 using the CAPM. They used the S&P 500 as the index of the market and decided to use 200 daily returns in order to estimate the market model. It starts 201 days before the announcement and ends 2 days before the press release. This 200 daily return window has been observed through past studies. In order to obtain their result, they calculated the CAR over 5 different periods: (-1, 0), (-1, 1), (-1, 5), (-1, 10) and (-1, 25).

Their findings were that a DOS attack does not lead to negative cumulative abnormal returns in general. However, there were some indication that such attacks did have an impact on companies that have the Web as their core business (internet companies). More specifically, internet companies had negative abnormal returns during five days after the announcement of the breach. On the other hand, the small internet companies (less than 100 employees) had a positive abnormal return after the announcement of the breach. The authors assumed that this could be due to the fact, that it gave "free publicity" to the company or could have made people think that they belong to the "major internet players". The effect of a DOS attack on an internet specific company led on average to a decrease of 1% of the firm value.

As previous literature states, attack types influence the effect on firm value and stock prices. Industry has also become an important factor, since internet specific firms are more penalized by those attacks. Therefore, more studies have been performed in order to see which industries are the most impacted by the data breaches.

2.3 The industry of the firm

Moving on to literature where industry of the firm is an important factor in explaining the difference in stock price drops due to data breaches. Cavusoglu, Mishra, and Raghunathan (2004) agreed with the previous literature that firm type plays a role. They verified if the 3 specific variables, referred to in previous literature, have an impact on the stock price after the announcement of a data breach. Those variables are firm type (internet firm or not), firm size, attack type. This study has been realized over the period of 1996 until 2001.

Similar to previous studies, they also started with an OLS regression in order to determine the intercept and slope parameters and used the NASDAQ composite index as market index. The size of the estimation window and event window has been based on previous studies. Since the previous

observed event windows were 120 and 200 days they decided to choose an average estimation window of 160 days, which started 160 days before announcement and ended 1 day before. Then the CAR has been calculated with an event window of (0, 1), since the authors state that it is rare that leakage from data breaches occur one day before the announcement by the press.

The first conclusion that the authors draw, is that there is a clear negative abnormal return in general after the announcement of the data breach. One day after the release of this information, the firm loses on average 2.1% of market value. Based on this data set, it corresponds on average to a loss of 1.65 billion American dollars. Going more into detail, this study proved that internet specific firms lose more per data breach. They lose, on average, 2.83 % of the firm value. This paper showed as well that smaller firms lose, on average, 1.5% more market value than bigger firms. But in contrast to previous literature, the authors found that all attack types had the same effect on the stock market.

One of the most recent studies has been done by Morse, Raval, Wingender (2011) who looked at the market price effect of data security breaches and therefore could observe if investors take the data breaches into consideration. Their dataset constituted worldwide companies between 2000 and 2010. Just as in previous literature, they started with the OLS method in order to be able to estimate the parameters in normal times. Their estimation period deviates from the others and they decided to take a whole year into account which corresponds to 255 days and used a window of 505 days before the data breach until 251 before the event. This corresponds to a year of daily trades one year before the announcement of the breach. To find the market returns, they used the value-weighted CRSP (Centre for Research in Security Prices) Index. They adjusted this for possible autocorrelation using GARCH (1, 1). They moved on by calculating the average abnormal returns which led to the CAAR (cumulative average abnormal return). Finally, the authors finished by performing a Z-test in order to be able to observe if the CAAR is equal to 0.

To see the industry differences, they decided to use CAAR (0) and CAAR (0, 1). Different attack types have also been reconsidered and have been observed over different time periods after the event. They calculated the CAAR (0), CAAR (0, +1), CAAR (+1, +5) and CAAR (+1, +10).

The first result from this paper is that from the 3 industries analysed (financial services, industrial and retail), only the financial services and the industrial industries have significant negative average abnormal returns. A breach in a financial service firm would lead on average to a loss of 0.34% of market price one day after the event. The reason for this is that those companies are the core of the payment card industry and detain a lot of confidential information related to money. This means that a failure in this industry would lead to wider impact and leads to contagion. A leak in the financial sector will lead to a problem in other businesses affiliated with the particular financial company

breached. Investors expect to have more security in the financial sector and therefore expect this to be a one-time event. However, in the industrial industry, the average negative effect is stronger with a decrease of 0.67% the day of the event. This might be due to the greater volume of funds flowing through payment cards. Finally, there is not a significant effect on retailers.

Type of attack has been defined here again in three areas, which are: stolen laptop, fraudulent access and hacking. The first observation is that a stolen laptop leads to a long term effect, having an impact over at least 10 days and results in a loss of 1% of the market value. Fraudulent access has a shorter term effect, which is only negative on the day of announcement, resulting in a loss on average of 0.86 % of market price. Those two types of breaches are inexcusable by investors and punish the market heavily. However, when a company gets hacked there is no significant impact on the market price. This means that investors agree that this is something that can happen to a company and nothing can be done to avoid it.

The difficulty for investors is to judge if they should accept that the data breach happened or not. Investors cannot figure out what the companies are doing on cyber security level. It is clear that a company cannot be 100% protected and has to have a plan B in case something goes wrong. Therefore, investors have to make sure that the company has at least one of those three plan B's. The first option is to take an insurance against cybercrime, which can reduce the overall cost of the breach (Shackelford, 2012). The second option would be to build a special department for cybercrime within the organization. This is mostly the case for big companies. The third option is to ask help from cyber security firms, who will constantly have an eye on the company such as Fireeye for example.

There is clear evidence that data breaches have different impacts on different industries and therefore the industry the company is in, impacts more or less the stock prices. We have seen that the financial, internet and industrial industries are punished the hardest. However, this trend can change over time. Previous studies have given an ambiguous view if the size of the company really plays a role on the stock price after a data breach, therefore the following literature will clarify if this variable has an important role.

2.4 The size of the company

Gatzlaff and McCullough (2010) studied an event period of 2 years between 2004 and 2006. They tried to find a relation if the firm type, the breach characteristics and the time of breach play a role on the negative impact of the data breach.
They used the OLS methodology, estimating the parameters based over the estimation window (-252,-7), which corresponds to the returns one year before the announcement of the event. In order to

estimate the returns of the market, they used the value-weighted CRSP index. They then performed the calculation of the CAR on the event period of (0, 1). The (0, 0) should perfectly capture the stock market's reaction, but since the announcement could be at any time of the day, the authors also included a second day.

They find that on average the firms lose 0.84% of their market value the day after the announcement. Three other main findings were outlined by this paper. First they observed that if the company refuses to disclose more information about the breach, it would lose on average 3 % of their market capitalization. The second finding is that stock market response to a breach in a small firm is more severe than in a big one. Finally, they observe that if the subsidiary of a firm is hacked, the parent firm will not be penalized by it. On the other hand, they observed no significant difference from 0 for the types of attack and for the number of records exposed by the breach. Therefore, the authors state that characteristics of the breach do not play a role in the stock market's response after a data breach.

A good example for this is Bitcoin, which got hacked in 2014 (Frisby, 2014). The responsible of communication, Mark Kapeles, stated just after the announcement of the hack: "I would like to kindly ask that people refrain from asking questions to our staff: they have been instructed not to give any response or information". This type of answer given by a company will have a bigger negative impact then if the company is directly honest and states what happened.

Table 1 gives a summary of previous research done in this area.

Based on the literature review, the announcement of the data breach leads to a decrease in firm value within three days, therefore this study will move on based on an event window of (-1, 1). For the estimation window, a window of 120 days will be used, which will be (-121, -2).

Furthermore, the financial industry, industrial firms and the online companies are the most affected by a data breach. Finally, there is evidence that size and attack type matters. This paper will try to find out if these factors still matters as of today, and if investors anticipate those attacks. If the investors do anticipate the attack for the characteristics (i) firm size, (ii) industry, (iii) attack type, this should result in a positive or close to zero cumulative abnormal return. New variables will be taken into account in order to establish a new link to the stock price effect after a data breach. Therefore, we could think that the following variables might as well influence the stock prices after a breach: (iv) year of the breach, (v) continent, (vi) economic state of the world, (vii) the recovery time and (viii) number of times hit.

Table 1. Summary of research in cybercrime

Researchers	Region	Time Period	Estimation window	Event window	Variables for CAR	Results
Campbell, Gordon, Loeb, Zhou (2003)	US only	1995-2000	(-121, -2)	(-1, 1)	Attack type	Breaches by unauthorized access to confidential information leads to loss of 5,5% of firm value.
Hovav, D'Arcy (2004)	US only	1988-2002	(-201, -2)	(-1, 0); (-1, 1); (-1, 5); (-1, 10); (-1, 25)	Firm type	Online companies are more penalized.
Cavusoglu, Mishra, Raghunathan (2004)	US only	1996-2001	(-160, -1)	(0, 1)	Firm type, Firm size, Attack type	CSAR not significantly negative.
Kannan, Rees, Sridhar (2007)	US only	1997-2003	(-52, -2)	(-1, 2); (-1, 7); (-1, 30)	Firm type, Firm sector, Attack type	No significant negative market return in the long run.
Goe, Shawky (2009)	US only	2004-2008	(-124, -5)	(-5, 5)	Breach type	Significant impact on financial performance.
Gatzlaff, McCullough (2010)	Not specified	2004-2006	(-252, -7)	(0, 1)	Firm type, breach characteristics, Time	Stronger negative reaction for growth opportunity companies and if no information is disclosed about the breach.
Andobaidoo, Amoakogyampah, Osei-Bryson (2010)	US only	1997-2003	(-120, -2)	(-1, 1)	Firm type, Firm size, Time	More negative impact after February 2000
Morse, Raval, Wingender (2011)	Global	2000-2010	(-505, -251)	(0); (0, 1); (1; 5); (1, 10)	Industry, Attack type	Investors do pay attention. It has a medium term effect not only temporary. Investors more sensitive to financial sector.
Das, Mukhopadhyay, Anand (2012)	Indian / US firms	2000-2012	(-122, -2)	(-1, 1); (-1; 3)	attack type, damage potency, firm type, size, performance	If subsidiary breached, then the parent company is not affected.

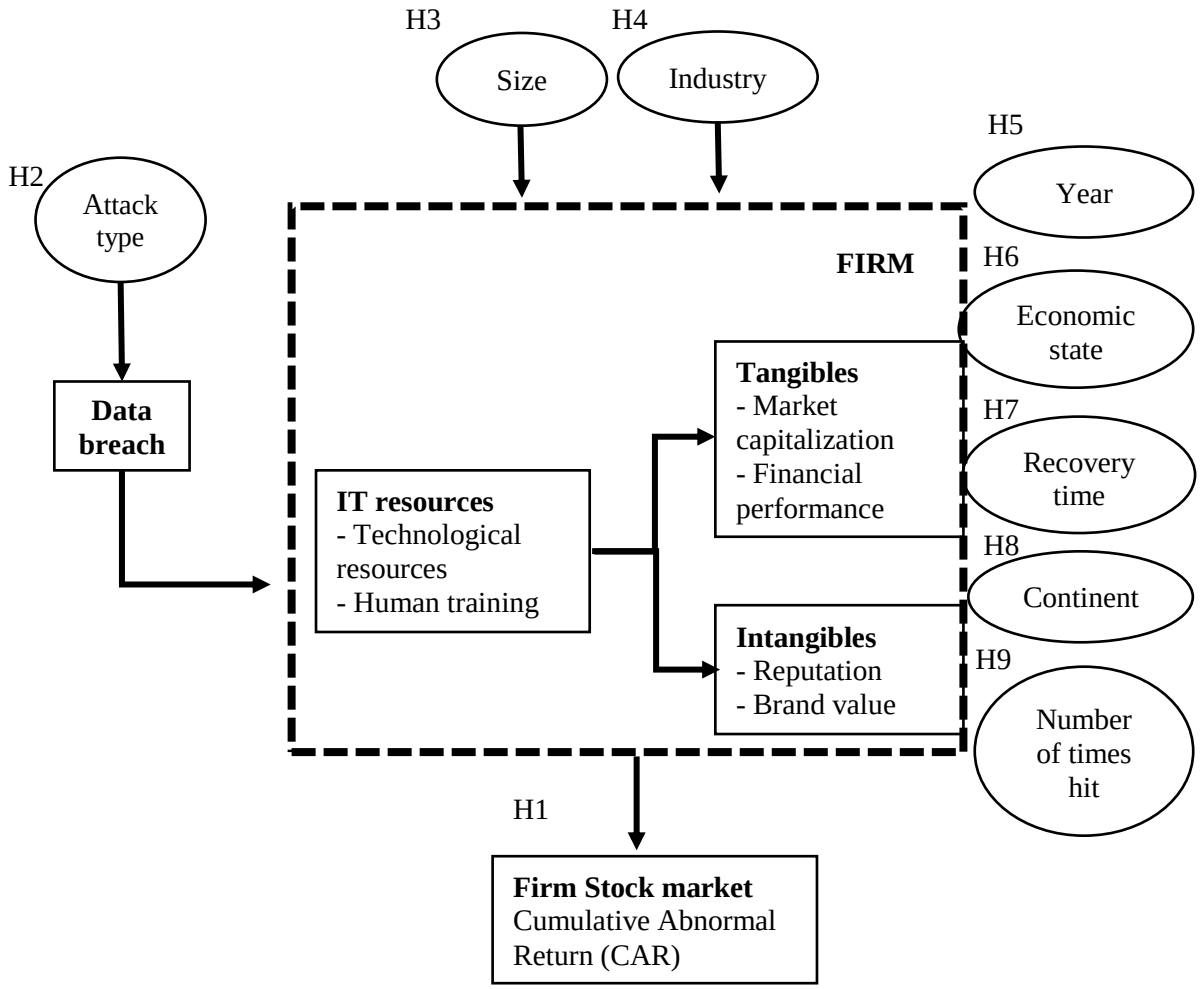
CHAPTER 3 Theoretical Framework

In order to be able to develop the hypotheses, it is useful to first describe the different types of costs a breached firm can face. Once the costs have been established, the model will then explain how this directly affects the firm's valuation. This section will be built on investors' beliefs and firm valuation based on the efficient market theory.

3.1 The cost of data breaches

The efficient market hypothesis suggests that financial markets are informationally efficient, and that stock prices reflect all publicly available information (Fama, 1969). Therefore, the impact of security breaches is measured by the change of market value of firms in response to the announcement of a data breach. However, since investors should be aware now that data breaches can occur, a premium might be taken into account to offset the negative effect a breach could have. The investors should be aware of the costs stated in Figure 1.

Figure 1. Characteristics influencing data breaches and costs related to cybercrime



As outlined in figure 1, the impact of a data breach does not only depend on the type of attack but also on the IT resources. The IT resources depend on the size of the company and its industry. For example the technological resources of a small financial firm are not the same as a big industrial firm. Small firms cannot spend the same amount of money on those resources. However, the human training should be the same for all kinds of firms. Human training corresponds to the trainings the firms give to their employees in order to update them on the new sorts of viruses. The importance of cybercrime has to be underlined and employees should be trained every 6 months on the new upcoming threats. They should also be educated about what they can and cannot do on the office computers. For example, they have to change their passwords monthly, not go on social media or free time websites on office computers, and should not connect any personal USB sticks or mobile phones to the computer. Previous study show that employees are the biggest problem to the data breaches, because they do not receive enough trainings about the new threats or by not applying what they learned in those sessions. The competition within a firm between co-workers make the employees consider cyber-security as the least of their problems. Delivering the work as soon as possible and doing it as efficiently as possible are their biggest priorities. This comes mostly at the cost of not being "cyber-secure" (Van Dongen, 2014).

H1 until H9 outlined in Figure 1 correspond to hypothesis 1 until hypothesis 9, which will be discussed in the hypothesis development part 3.2.

Once the damage is done, the question is what is the impact of the breach, and how much will it cost (Kelly, 1999). This is called risk assessment. The most difficult step in risk assessment is to evaluate the real financial cost of the breach, because:

a. Companies are unwilling or unable to quantify their financial losses (Power, 2001).
b. Lots of breaches are not reported. This is because the firms are afraid of future crimes, negative publicity and management embarrassment (Hoffer and Straub, 1989). Companies do not want their competition to know, otherwise they could exploit this attack and gain a competitive advantage.
c. Finally, they want to avoid at all cost a drop in the firm's stock price by not disclosing it. (Sprecher and Pertl, 1988).

However, even if the companies cannot put a real number onto the data breach costs, a clear distinction can be made by the factors impacted. There are tangible and intangible costs. Potential tangible costs are (D'Amico, 2000):

a. Staff and material costs related to the detection of the breach, the containment and the repair of it.
b. Public relations costs for statements to the press and answer possible customer questions
c. Lawyer costs in order to defend the company in case it has to go to court, because of its failure to deliver assured information and services.
d. Loss of productivity of the employees, who work in a bad situation or do not work at all.
e. Loss of financial performance and market capitalization, due to the loss of productivity and disruption of work conditions.

The tangible costs are the costs that are the easiest to value. These are costs that are observable and the impact of the breach can be seen straight ahead in the expenses of the company. However, the intangibles are the hardest to value. Example of intangible costs are:

a. A loss of competitive advantage. This leads to an opportunity for their competitors who would certainly take advantage of it.
b. Loss of brand value. When a company gets breached, people will rarely forget it and always have their doubt about it.
c. Loss of reputation. People lose trust and faith in a company and are afraid the company might be hit again.

Finally, all these costs will end up affecting the stock price and the cumulative abnormal return.

3.2 Hypothesis development

From previous literature, we have seen that attack type (Campbell et al., 2003), industry (Cavusoglu et al., 2004; Morse et al., 2011) and size (Gatzlaff and McCullough, 2010) affect the CAR. The size and industry are factors that are clearly observable and can be taken into account while anticipating a breach. However, attack types are more difficult to anticipate, since investors have no information on that. Brounen and Derwall (2010) noticed that price reaction differs across the countries and therefore I decided to implement the continent variable into this study.

In this study, 4 additional variables will be taken into account. Previous literature did not mention those variables, but it is possible those might influence the CAR as well. The year and the economic state could explain if the investors are more risk averse in economic distress and take a bigger premium into account. The recovery time will tell us how long it takes for the firm to solve the data breach. Finally, the last variable is the number of times hit. This will tell us if the data breach effect has a more negative impact when the company is breached more than once.

This suggests the following hypothesis:

Hypothesis 1A: Investors take into account data breaches nowadays, which leads to a CAR close to zero.

Hypothesis 1B: Investors do not take data breaches into account, this will result in a negative CAR.

We have seen from all the previous literature that over the past years, latest 2012, the overall CAR after the press release of a data breach has been negative. However, this study is going to verify if it is still the case as of today. The investors might start to understand that being hacked is inevitable and are softer about this for certain types of companies and industries. Therefore, our hypothesis is that over time the negative CAR effect after the announcement of a data breach is decreasing and could be very close to 0 as of today. Campbell et al.(2003) found that firms lose on average 5.5% of firm value after the announcement of the breach, while a later study from Gatzlaff et al. (2010) shows that this effect leads to an average loss of 0.84 % of firm value.

Hypothesis 2: Investors take into account specific attack types only.

Previous literature has made a difference between confidential breaches and non-confidential breaches (Campbell, 2003; Hovav, 2004). There has been clear evidence that attacks which simply disrupt the company's functioning (non-confidential attack) have no impact on the CAR. Therefore, this study will go deeper into the confidential breaches and differentiating "employee mistakes" and "technological failure".

To illustrate this, an "employee mistake" would correspond to the case of Target's breach. "The company was warned when the hackers attacked in 2013, but it ignored multiple alerts that something was wrong and continued selling to consumers. As a result, millions of people continued to swipe their credit cards and their information continued to be sent to hackers."(Manworren and al., 2016). This is something that employees within Target could have avoided, but instead were more focused on keeping the business running.

On the other hand, there is the breach at Siemens which evolved zero-day based attack by the name of Stuxnet. This is one of the zero-day based type attacks initially built to destroy Iran's nuclear program, but is now used against companies. Siemens press release contained the following information: "[…] these vulnerabilities were discovered while working under special laboratory conditions with

unlimited access to protocols and controllers."(McMillan, 2011). These are the types of attacks that companies can hardly avoid, even sometimes hardly detect.

Therefore, this hypothesis will find out if investors punish those two types of attack differently.

Hypothesis 3: Investors take into account the risk for small firms only.

Large firms have more resources then small firms, but at the same time large firms have more people so there is a higher probability that a "human mistake" occurs. Gatzlaff et al. (2010) found that the market response after a breach in a small firm is more severe than in a big firm. It seems that smaller firms are more prone to be victims of real hackers, while big firms are more often victims of employee mistakes. In the data breach investigations report from Verizon 2012, it is illustrated that the hackers are mostly focused on smaller companies, especially those between 11 and 100 employees. 450 breaches out on the 750 reported worldwide in 2010 have been on companies having between 11 and 100 employees, which correspond to 60% of worldwide breaches. In 2011, this trend has increased. During that time, 580 data breaches were in small firms out of 860 reports, which is about 67%.

However, there is more money to steal from big companies, but those are also more secured. In a market where obtaining hacking devices is quite easy, even for the most "novice" people, being a start-up company or just a small company having less than 100 employees puts you in the spotlight for hackers.

Small firms are the companies that investors like to invest into, because stocks of small companies have higher incidences of price volatility and mispricing, and therefore this is an increased opportunities for investors to earn excess returns. However, the statistics from Verizon show that they are the most dangerous ones. It is also known that returns for small firms exceed returns for large firms. Fama and French (1992) say that this is due to distress risk or liquidity risk in small firms.

Therefore, it will be evaluated if investors observe the difference between the firm size and who they punish the most.

Hypothesis 4: The investors take into account the risk in certain industries only.

Evidence has been found that in the past, different industries had different impacts on the CAR. The industries that had a significant negative impact on the CAR were the financial services, industrial firms and internet specific companies. (Cavusoglu et al., 2004; Morse et al., 2011). Trends have changed nowadays and the industries affected by data breaches have changed. Figure 2 from Trend

Micro (Huq, 2015) shows how often an industry is attacked by cybercriminals between 2005 and 2015.

As we can see here, the industry mostly attacked by data breaches is the healthcare industry. This doesn't mean that it is the industry with the most sensitive data and the one that loses the most with those breaches. This is why further industry analysis has been done by Statista, a static portal, regrouping the average annual costs caused by cybercrime in the United States as of August 2015. The financial companies are the ones who lost the most with an annualized average cost of 13.5 million USD. They have the most sensitive data even though they aren't hit as often as the other industries, they lose more than all other industries.

Figure 2. Industries affected by data breaches

Source: Trend Micro, 2015

By combining those two sources, we obtain a reliable set of industries that will be investigated into. Therefore, in our analysis we will be focusing on the industries that are hit the most and lose the most which are (in brackets are the average annualized costs of data breaches in million USD):

- Financial services (28.33)
- Utilities and energy (27.62)
- Technology (16.45)
- Web Services (12.93)
- Retail (11.96)

Government, education and healthcare are industries where no stock prices are available and therefore will not be taken into account.

Hypothesis 5: Over time the investors do not punish the firms as hard as before when they are victim of a data breach.

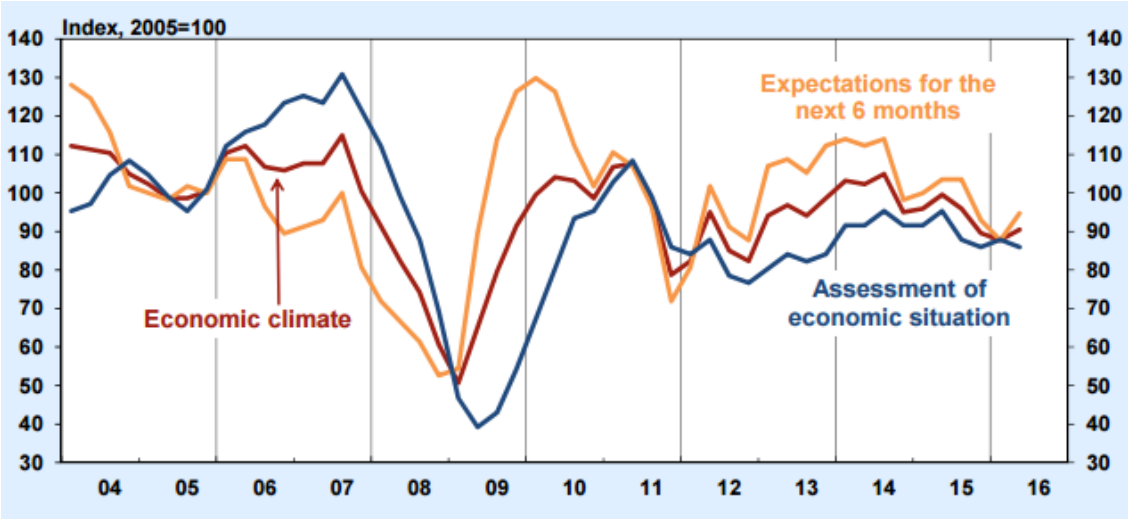
As discussed earlier, the trend of hacking and companies being victim of data breaches started already in 1990. Campbell et al. (2003) showed that investors were initially punishing the firm very hard after it got breached, while later on this negative impact has decreased (Gatzlaff et al., 2010). This study is going deeper on the year impact and the evolution of the CAR over the years 2006 until 2015.

Hypothesis 6: The investors pay more attention to firms being hacked during distress periods.

During our area of study (2006 until 2015), the worldwide economic situation and expectation vary over time. There is no clear evidence yet, if this has impacted the valuation of the investor on the stock prices. Therefore, two states of the world will be defined in order to observe if there is relation. A "bad" economic situation, where the world's economic expectation was above its actual situation. The centre of economic studies (Garnitz et al, 2016) shows us that the periods of distress are from 2009 until 2011 and from 2012 until today. On the other hand, the "better" economic situation where the world's economic situation is above economic expectation. These periods are between 2006 until 2009 and from 2011 until 2012.

Figure 3 shows the period of "bad" and "better" situation by observing the difference between the expectation and the actual situation. The economic climate corresponds to the average of those two factors. Defining the economic state of the world will help us understand if there is a relation between the economic state and the CAR.

Figure 3. Worldwide economic situation over the years



Source: World Economic survey, 2nd quarter 2016, 11 May 2016

Hypothesis 7: The longer the recovery time, the harder the investors punish the firm.

Previous study has shown that it takes on average 21 days for the stock price to recover from a data breach (Kannan et al., 2007). Later on, Morse et al. (2011) proved that the average time for the stock price to recover was about 10 days, which corresponds already to a decrease of 10 day compared to the study before. Therefore, this study will try to find out if the companies have a faster response time after it got breached.

Hypothesis 8: Investors are aware that some continents are more prone to become victims of cybercrime.

The main focus of this study will be on 3 continents which are Europe, America and Asia. Those locations have been chosen, because most data breaches occur in those areas. Gemalto, a cyber security firm, provided information about the share and number of attacks per continent in 2015. The continent that has been hit the most is America with 1,293 incidents over the year, with 1,222 only in the United States. This represents 77% of the worldwide incidents in 2015. Europe has been victim of 209 data breaches, or 12% worldwide attacks, while Asia only has been breached 131 times in that year.

It can be clearly seen that there is a geographical difference in attacks. This factor has not been taken into account in previous studies and the main goal is to observe if investors do observe this geographical trend.

Hypothesis 9: Investors punish firms harder when a second data breach occurs.

In the last hypothesis, we will be observing if the CAAR for the firms is more negative when they get breached more than one time. No previous literature has covered this matter and it could be interesting to find out if the number of times attacked is related with a different effect on the CAAR.

CHAPTER 4 Data & Methodology

This chapter describes how the data have been collected in order to obtain a sample of data breaches. It will continue with the methodology part, which has been divided into two parts. The first part consists of a detailed versions where all the elements within the variables will be analysed in order to observe if any of them have an impact on the CAR. The second part consists in a cross-sectional regression to give a bigger picture to see which variables investors really take into account. Finally, a full explanation of the variables and the descriptive statistics will be given in order to have an overview of the data.

4.1 Sample selection

The sample has been collected from various sources in order to obtain worldwide data from publicly traded companies. First of all, I identified data breaches by electronically searching in the following newspapers: *Financial Times*, *New York Times*, *Washington Post*, *Wall Street Journal* and *USA Today*. In those newspapers, I have looked for the following terms: ”data breach”, “computer attacks”, “hacker” and “cyber-attack”. The newspapers are considered as being the most important sources of data since this is the main source for the investors’ community to get information. Therefore, if there has been a stock market reaction, this is the place where it is most likely to be found. This brought me to a dataset based on 90 listed firms. However, those newspapers contain mostly big, highly visible international companies, while our testing should also include small sized firms. Therefore two other databases have been used in order to make up for the differences in the firm sizes.

The second source that has been used in order to obtain more data is the Privacy Rights Clearinghouse. This website offered a database of data breaches from 2005 until 2015 mostly concentrated in America, containing small and big companies. This provided me with 41 additional observations.

Finally, the last source that has been used is the database from Breach level index. This database consists of worldwide companies and government breaches from 2013 until today. This was very helpful in order to get international data. This database helped to gather 24 additional breached companies.

The sample is made of 155 events worldwide from January 1, 2006 until December 31, 2015. However, for 27 companies the estimation window of 120 days was not available on the Centre for Research in Security Prices (CRSP) and are therefore dropped out of the sample. The final sample is made of 128 events from 107 different firms. The names of the firms used in this database can be found in appendix B. The sample size is small since a lot of companies are private, governmental or

educational, where stock prices are unavailable. I believe that this sample size is representative of the breached firms market and provides reliable outcomes.

4.2 Methodology

Two methods have been considered in order to see if the variables play a role in the investors' decisions after a breach. First of all, an event study has been performed in order to see if the factors within the different variables play a role. For example for the industry variable, the event study will observe if the financial industry lead to a CAR significantly different from zero, and replicate this for every single industry stated in the hypothesis. When those effects within have been observed, a cross-sectional regression will give a broader overview of the situation. For instance, if financial industry plays a significant role on the CAR, this does not specifically mean that the whole industry variable has an impact on the CAR. I will start to explain the methodology of the event study.

4.2.1 Event study

In order to test all the hypotheses individually, I use an event study methodology to observe the stock performance around the announcement of data breaches. Event studies allow us to measure the value effect of the announcement of the breach under the assumption of market rationality, which confirms the fact that investors' assessment of the firm value is reflected in the stock prices. Therefore, any abnormal return observed is interpreted as a measure of the impact of the data breach.

First, I estimate the firm's stock return in absence of the event in order to later estimate the effect of the announcement of a data breach. This effect will be given in a percentage of the market value and also calculated back in US dollars. If the effect is significant, this will give us the average loss in US dollars per data breach. We assume that the daily stock returns are consistent with the Capital Asset Pricing Model (CAPM).

The estimation of the market model has been set over a 120- day period, which corresponds to a half year of daily trades before the event. Various different time periods have been used in previous literature. However, most of them have used a 120-day estimation period (Campbell et al., 2003; Goe et al., 2009; Andobaidoo et al., 2010; Das et al., 2012). This is the shortest of the accepted estimation periods and helps me to retain as many observations as possible. The market model is estimated as follows:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \quad (1)$$

where

R_{it} = the return for the firm *i* for the period *t*,

α_i = the intercept of the returns for the firm *i*,

β_i = the market model slope parameters for firm *i*,

R_{mt} = the return on the CRSP equally-weighted portfolio for the period *t*,

ε_{it} = the residual error term for firm *i* for period *t*.

To find the slope and intercept (alpha and betas), the ordinary least square and the method of Scholes and Williams (1977) are used for each firm in the data set. The parameters found in equation (1) are then used to find the expected return, which is finally necessary to find the abnormal returns. The abnormal returns are the difference between the actual returns (R_{it}) and the expected returns. This is calculated as follows:

$$AR_{it} = R_{it} - (\alpha_i + \beta_i R_{mt}) \tag{2}$$

Once the abnormal return for each firm for time period *t* has been defined, I calculated the average abnormal return (AAR_t) in order to get an overview of the average of the whole samples' abnormal return. This can be calculated as the average AR_{it} for N events:

$$AAR_t = \frac{\sum_{i=1}^{N} AR_{it}}{N} \tag{3}$$

Finally, to obtain the cumulative average abnormal return ($CAAR_{T1,T2}$), I summed up all the AAR between the event period. $T1$ is defined as the start of the event period and $T2$ is defined as the end of the event period. This can be calculated as follows:

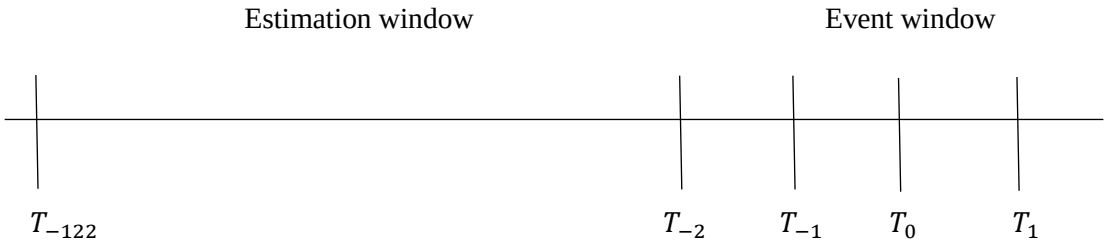
$$CAAR_{T1,T2} = \sum_{t=T1}^{T2} AAR_t \tag{4}$$

The cumulative average abnormal return will tell us if the data breaches have an effect on the stock market. The main test will be to see if the CAAR is significantly different from zero. In order to do so the effect of the data breach has to be estimated within a reasonable time window, this is called the event window. It is the amount of days an investors needs to absorb the impact of a news event. A short event window is likely to be more reliable to test the data breach effect since according to the efficient market hypothesis, new information is directly incorporated into the stock price. This has also been proven by previous literature on data breaches, since a longer window period would lead to a distraction by other possible external factors (Hovav, 2004; Kannan, 2007). Therefore, event periods

of (0, 0), (0, 1) and (-1, 1) have been assumed to be relevant. The effect should be covered in the CAAR (0, 0), however sometimes there might be information leakage before the announcement date and thus I took also into account the day before the event.

Figure 4 summarizes the estimation window (-122, -2) and event window (-1, 1) taken into account for this study.

Figure 4. Estimation and event window



To test if the CAAR is significantly different from zero, I use the standardized residual method also known as the Patell Z test (Patell, 1976). In this test, the event window's abnormal returns are standardized by the standard deviation of the estimation window. There are two main assumption with the Patell Z. The first one is that it assumes cross-sectional independence of the abnormal returns. The second assumption is that the abnormal returns are normally distributed. In order to be able to perform the Patell test, we first have to calculate the standardized abnormal returns (SAR_{it}):

$$SAR_{it} = \frac{AR_{it}}{S_{it}} \tag{5}$$

In order to obtain the square root of the estimated forecasted variance of the firm *i* for the period *t* (S_{it}), the following calculation is made:

$$S_{it}^2 = S_i^2 \left(1 + \frac{1}{D_i} + \left[\frac{(R_{mt} - \bar{R}_m)^2}{\sum_{t=1}^{D_i}(R_{mt} - \bar{R}_m)^2}\right]\right) \tag{6}$$

where D_i corresponds to trading days used in the estimation period (120 days) for the firm *i* and $\bar{R}_m$ is the average market return over the estimation period. Finally the estimated forecasted variance of the firm *i* (S_i^2), is calculated as follows:

$$S_i^2 = \frac{\sum_{t=1}^{D_i} AR_{it}^2}{D_i - 2} \tag{7}$$

This brings us finally to the Patell Z test equation (8), for which the null hypothesis is that the $CAAR_{T1,T2}$ is equal to zero, which means in our research that investors do take into account the data breach risk.

$$Z_{T1,T2} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{T1,T2}^i \tag{8}$$

where

$$Z_{T1,T2}^i = \frac{1}{\sqrt{Q_{T1,T2}^i}} \sum_{t=T1}^{T2} SAR_{it} \tag{9}$$

and

$$Q_{T1,T2}^i = (T2 - T1 + 1)\frac{D_i - 2}{D_j - 4} \tag{10}$$

The last test performed in this event study is the robustness check. In order to verify the robustness of the results I used a non-parametric test called the Corrado's rank test (1989). This test is also based on standardized returns and is very competitive and often called superior in comparison to the Patell Z (1976) and the Boehmer et al. (1991). The Corrado's rank test has proven to be not only robust against event-induced volatility, but also to cross-correlation due to event-day clustering, which might be the case in this sample. Clustering might be an issue, since one hacker could have attacked more companies at the same time with the same virus, leading to a cluster of data breaches around the same day. The Corrado's rank test has been extended later by Campbell and Wasley (1993) in order to get a rank test for multi-day event periods. This is calculated as follows:

$$T_{Corrado} = \sqrt{t_1 - t_{-1}}\ \frac{\overline{K}_{event} - (t_1 - t_{-2} + 2)/2}{\sqrt{\sum_{t=t_{-2}}^{t_1} (\overline{K}_t - (t_1 - t_{-2} + 2)/2)^2 / (t_1 - t_{-2} + 1)}} \tag{11}$$

where

$(t_1 - t_{-2} + 2)/2$ = the mean rank

$(t_1 - t_{-2} + 1)$ = the largest cumulative average abnormal return rank

$\overline{K}_{event}$ = the average rank across N stocks

$t_1 - t_{-1}$ = the days of the event window

K_{it} represents the rank of the cumulative average abnormal in a period of $(t_1 - t_{-2} + 1)$ cumulative average abnormal return of stock *i*. $\overline{K_t}$ is the average rank across N stock for the period *t*, which corresponds to the estimation and event period.

Once those tests have been performed and once the effect of each element on the CAR has been verified, a broader approach will be considered. In order to have a bigger picture and see which variable really influence the CAR, the cross-sectional regression approach will be used, which we will explored in the next section.

4.2.2 Cross-sectional regression

To test the hypothesis overall, a cross-sectional regression will be performed. This will help to zoom out and to see if every single variable considered has an impact on the CAR. Since the most relevant event period used in previous literature has been (-1, 1), this period will be used. Therefore, the independent variable will be CAR (-1, 1). The dependent variables will be all the variables discussed in the hypothesis, which will be explained in more detail in part 4.3.

As Mackinlay (1997) suggests, this implies regressing directly the cumulative abnormal return on different events and company characteristics. This is shown by equation 12:

$$\begin{aligned} CAR(-1,1) = \ & \alpha + \beta_1(attack\ type) + \beta_2(firm\ size) + \beta_3(industry) \\ & + \beta_4(year) + \beta_5(economic\ state) + \beta_6(recovery\ time) \\ & + \beta_7(continent) + \beta_8(number\ of\ times\ hit) + \varepsilon \end{aligned} \tag{12}$$

$$E(\varepsilon) = 0 \tag{13}$$

The main assumption to test in the cross-sectional regression is that there is no multicollinearity. If there is multicollinearity this might lead to distortion of the results. The test of multicollinearity is performed with the Pearson Correlation test. Feldman and Santangelo (2008) estimate that if the Pearson Correlation reaches a coefficient of 0.80, there is evidence for multicollinearity. Table 2 presents the Pearson Correlation test.

Table 2. Pearson Correlation

	Continent	Economic situation	Year	Industry	Attack type	Firm size	Recovery time	Times Hit
Continent	1							
Economic Situation	0.263	1						
Year	0.398	0.745	1					
Industry	0.036	-0.018	0.007	1				
Attack type	0.263	0.247	0.368	-0.04	1			
Firm size	0.176	0.05	0.12	0.083	0.058	1		
Recovery time	-0.194	-0.18	-0.21	-0.109	0.071	-0.2	1	
Times Hit	0.09	0.247	0.272	-0.133	0.224	0.29	-0.018	1

We can see that there is no significant evidence of multicollinearity in this sample. However, the correlation between the year and economic situation is very high, which is normal since the economic situation is based on the year.

Another important assumption is to verify that there is no heteroscedasticity. Heteroscedasticity is observed when the variance of the semi-elasticities is influenced by the magnitude of the independent variable. Table 3 presents the results of the Breusch-Pagan test.

There is evidence that there is heteroscedasticity in this sample. This can be seen by the value of Breusch-Pagain which is highly significant. Therefore, in order to prevent the coefficient to be biased, robust standard errors will be used during the cross-sectional regression.

Table 3. Breusch-Pagan heteroscedasticity

This table presents the Breusch-Pagan heteroscedasticity test with dependent variable CAR (-1, 1). Independent variables include Attack type, which corresponds to a dummy: 0 if the company faced a technological failure and 1 if the company faced a human mistake. Firm size is also a dummy variable, where a small firm corresponds to 0 and a large firm to 1. The industry variable is an ordinal variable. The following numbers have been assigned per industry: 1 corresponds to the financial industry, 2 the technological industry, 3 the retail industry, 4 the web industry and 5 the utility and energy industry. The variable year has been assigned in the same way, chronologically, where 2006 corresponds to 1 and 2015 to 10. The economic state of the world has been defined as a dummy variable, 0 if the economic state is good and 1 if there is an economic depression. The recovery time corresponds to the time the company takes to solve the problem within the company, 1 is a short recovery time (shorter than a week), 2 is a medium recovery time (between a week and a month) and 3 is a long recovery time (longer than a month). The 3 continents have been assigned the following values: 1 for America, 2 for Europe and 3 for Asia. Finally, the number of times hit is also a dummy variable, where 0 is the dummy for being hit only once and 1 if the company got hit more than once. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement. The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

	Coefficient	Std. Error	T statistic
Constant	-0,3444	1,6802	-0,2047
Attack type	0,9841	0,7715	1,276
Firm size	0,0930	0,8992	0,1034
Industry	-0,0941	0,2240	-0,4203
Year	0,0858	0,1852	0,4632
Economic state	0,0288	1,0351	0,0278
Recovery time	0,0461	0,4748	0,0971
Continent	0,7925	0,8816	0,8989
Times hit	-1,0759	0,8323	-1,293
Observations	128		
Breusch-Pagan	42,0004***		

4.3 Explanation of variables

Independent variable

The main independent variable is the Cumulative abnormal return for the event period (-1, 1) for the cross-sectional regression, while the main independent variables for the event study are the cumulative average abnormal returns for different event periods: (0, 0), (0, 1) and (-1, 1).

Table 4 confirms that (-1, 1) is the best event window that can be used for this research. From the AAR, we cannot really observe any significant effect, except for AAR (2). For AAR (2) we observe a significant positive effect for a 1% significance level. This effect is positive and could correspond to the counter effect of the value loss after the data breach. An additional study of the event windows over the CAAR has been done in order to observe which would be the best window to use. We

observe with the Corrado rank test, which is robust, that the event window (-1, 1) would be the best predictor for this study.

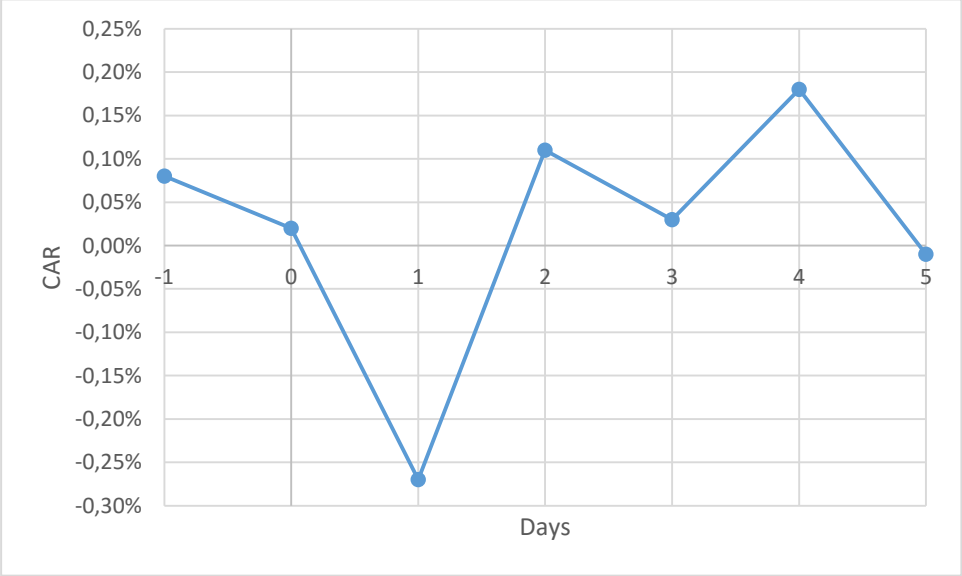
Table 4. Event window

This upper part of this table presents the average abnormal returns from 1 day before until 2 days after the breach. The second part of the table presents different event windows (0), (0, 1). (0, 2), (-1, 1) and (-1, 2) in order to see which event window is the best to use in this event study. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement. The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event day	N	Mean AR (%)	Positive:Negative	Patell Z	Corrado Rank test
-1	128	0,08	56:72	-0,992	-1,1
0	128	-0,06	60:68	-0,151	-0,518
1	128	-0,29	60:68	-1,003	-0,943
2	128	0,38	72:56	2,918*	1,984**
Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank test
(0)	128	-0,06	60:68	-0,151	-0,519
(0, 1)	128	-0,35	61:67	-0,816	-1,03
(0, 2)	128	0,03	66:62	1,018	0,302
(-1, 1)	128	-0,28	60:68	-1,239	-1,476***
(-1 , 2)	128	0,38	66:62	0,386	-0,289

Figure 5 provides an overview of the CAAR from 1 day before the event until 5 days after in order to observe if there could be an overall medium term effect. The graph confirms that the event window (-1, 1) fits the best to the answer of the data breach event and no clear medium effect can be observed apart from the counter effect from the breach.

Figure 5. Overview of the overall CAR of the sample



Dependent variable

I will now outline the different dependent variables, where each variable corresponds to one hypothesis presented in part 3.2.

Attack type

To test hypothesis 2, two different attack types are considered. First there will be the "technological failures", which consists of all intrusion that have been made purely by hacking. Those are attacks that are based on the computer defence gear, which might not be sufficient or can be simply unavailable since the technology to avoid those hackings is not there yet. The second type of intrusion is the "employee mistake". This corresponds to all types of hacking that are done only because of an employee within the firm. Four categories are grouped within the "employee mistake" part:

- Insider: This is someone already working for the firm who hacks or collects information from the inside of the company.
- Lost media: An employee who lost his laptop or cell phone for example.
- Stolen media: An employee's laptop or cell phone which got stolen at a hotel for example.
- Accidently published: This is very rare, but sometimes employees who are not very good with computers published client information unintentionally.

Firm size

The European Union's European Commission and the U.S. Small Business Administration (2010) have defined that a small company would have a maximum turnover of 10 million Euro up to 50 million Euro for medium businesses. This has also been considered in this study and the rest has been stated as being large firms.

Industry

For hypothesis 4, five industries have been considered. Those five industries have been chosen based on the number of times they are hit and how much annual costs each industry faces. More detailed description will now be given in order to better identify the five industries:

- Financial services: In this industry, all companies related to the financial world have been taken into account, such as banks, consulting companies and payment companies (American Express for example).

- Utilities and Energy: Energy companies are firms working around oil, gas and electricity, while utility companies regroup a lot of various activities such as travel agencies, telecommunication firms, and transportation companies.
- Technology: The gap between technology and web services is very small. However, technology companies are companies where the main business is to innovate and bring new items to the market.
- Web Service: Web service companies are firms that are only allocated on internet and do not have any physical customer point. This can be social media, dating websites or specialized internet companies.
- Retail: The retail industry is composed of supermarkets and hypermarkets which can be local or worldwide.

Year

The period used for this study is from 2006 until 2015. This time period has been used to give a better overview over time how the abnormal returns evolves and 2006 was the furthest data that I was able to collect. 2016 has not been taken into account, because the data is not yet available on Eventus.

Economic state

As described in hypothesis 6, two periods are defined: "good" and "bad economic state. The "good" economic state has been defined to be from 2006 until end 2008 and then from June 2011 until March 2012. On the other hand, the "bad" economic state are set to be the periods in between, thus, from 2009 until June 2011 and from March 2012 until 2015.

Recovery time

The recovery time corresponds to the time it takes for a company to recover its data and close the breaches. In the case of a loss of laptop this corresponds to the time it takes for the company to recover the laptop, while for hacking this corresponds to the time until the breach is closed. Three different periods have been defined. First, the short period which corresponds to a recovery time between 0 and 7 days after the event. Then there is the medium recovery time which corresponds to a time between 7 days and 30 days after the event. Finally, the long period is assumed to be more than 30 days after the event took place.

Continent

For hypothesis 8, 3 continents are considered: America, Europe and Asia. Those are the continents where the data breach took place which does not specially mean that the companies are quoted on the stock market in the place of the event. All the data used for this research are quoted on the American stock market even though the events happened in other places.

Number of times hit

The number of times hit will be defined in two categories. There are companies that are hit only one time and those who are hit two times or more. The outcome of this test will help us to understand if the investors punish the companies more if they are hit more than once by data breaches.

4.4 Descriptive statistics

Since all the variables are nominal or ordinal variables, it is better to analyse the number of observations within each category and understand the disparities within the sample, which sometimes will make it difficult to draw significant conclusion from. Table 5 presents the number of observations for each variable.

Most of the observations are more or less equal to each other, however it will be difficult to extrapolate the results from the variables firm size and continent. Previous research has shown a very small sample size for small firms as well (Kannan et al., 2007).

The paper draws conclusions about the sample size, however, this should be taken with caution. From table 5 we can also observe that 87.50 % of the sample is made of breaches on American soil. Even if the sample size of breaches in Europe and Asia are very close to reality in percentage (10 % and 2% of the sample size respectively), the results need to be taken with caution as well, because I was able to collect only 13 European and 3 Asian breaches.

Table 5. Number of observations within each variable

	Counts		Percentage of total
	Total	Amount	
Attack type	128		
Technological failures		43	33.59%
Employee mistakes		85	66.41%
Firm size	128		
Small / Medium		25	19.53%
Large		103	80.47%
Industry	128		
Financial services		37	28.91%
Retail		33	25.78%
Web services		16	12.50%
Technology		15	11.72%
Utility and Energy		27	21.09%
Year	128		
2006		13	10.16%
2007		12	9.38%
2008		12	9.38%
2009		10	7.81%
2010		13	10.16%
2011		12	9.38%
2012		13	10.16%
2013		16	12.50%
2014		14	10.94%
2015		13	10.16%
Economic state	128		
Good		45	35.16%
Bad		83	64.84%
Recovery time	128		
Short		82	64.06%
Medium		27	21.09%
Long		19	14.84%
Continent	128		
America		112	87.50%
Europe		13	10.16%
Asia		3	2.34%
Times hit	128		
One time		95	74.22%
More than once		33	25.78%

Table 6 provides a breakdown of the countries taken in this sample. The sample is mainly made of companies attacked in the United States. They represent 71.09% of the sample. The results might be very US bias. The second most important player in this sample is Canada with 15 observations and as third, the UK representing about 5% of the total sample.

Table 6. Countries in which the firms have been breached

	Total	**Percentage of continent**	**Percentage of total**
Continent	**128**		
America	**112**		**87.50%**
USA	91	81.25%	71.09%
Canada	15	13.39%	11.72%
Mexico	3	2.68%	2.34%
Brazil	2	1.79%	1.56%
Argentina	1	0.89%	0.78%
Europe	**13**		**10.16%**
UK	6	46.15%	4.69%
Germany	5	38.46%	3.91%
Finland	1	7.69%	0.78%
France	1	7.69%	0.78%
Asia	**3**		**2.34%**
Japan	2	66.67%	1.56%
Hong-Kong	1	33.33%	0.78%

Apart from size and continent, all the other variables will be analyzed and can be extrapolated to the whole population.

CHAPTER 5 Results

This chapter will present the results of this study with the methodology described in the previous part. First, in part 5.1, the event study will give a detailed picture of the factors which impact the CAAR. After that, a bigger picture will be shown to see which variables directly affect the CAR in part 5.2.

5.1 Event study

In order to test the 8 hypothesis a detailed view will be presented in order to see which factors play an important role on the CAAR. We will first see if the CAAR is significantly different from 0, and then observe what impact the different factors have.

5.1.1 Overall CAR effect and attack type

The main hypothesis (H1) is to observe if the investors take date breaches into account. This means that the Patell Z should be insignificant. The goal of the Patell Z test is to verify if the CAAR is significantly different from zero. We observe that we cannot reject the Patell Z for any event period. This means that the CAAR is not significantly different from zero. This leads me to accept the hypothesis 1A and therefore means that investors do take into account data breaches. This is proven by the fact that the announcement of data breaches does not significantly impact the stock prices and therefore we can assume that investors take those already into account when investing in them.

However, the type of attack does matter for the CAAR, especially the technological failures. The investors do not anticipate technological failures and punish the firms more harshly if they encounter these. This is opposed with previous literature, which stated that employee mistakes are less forgiven and are punished more harshly, while technological failures are unavoidable. This leads, on average, to a decrease of 0.11% of the firm value around 3 days of the event (-1, 1) with a 5% significance level. This result is proven to be robust with the Corrado rank test. In real figures, a technological breach will lead on average to a loss of 13,044,181 US dollars in market capitalization.

The reason for this might be that investors want the companies to protect themselves well and that they do everything in their power to protect themselves more. The investors could think that companies are not doing enough and that they need to invest not only in a professional in house cyber protection department but also in an external cyber protection firm.

On the other hand employee mistakes do not have any significant effect on the CAAR. This could be due to the fact that every human makes mistakes; this is unavoidable. Humans use computers.

Therefore the logical conclusion from this is that humans can make mistakes on computers, which leads to data breaches. Most of the time, this is hard to avoid for companies, therefore the investors account for it, when they take the necessary steps after the event. For example firing the employee. The results can be found in table 7.

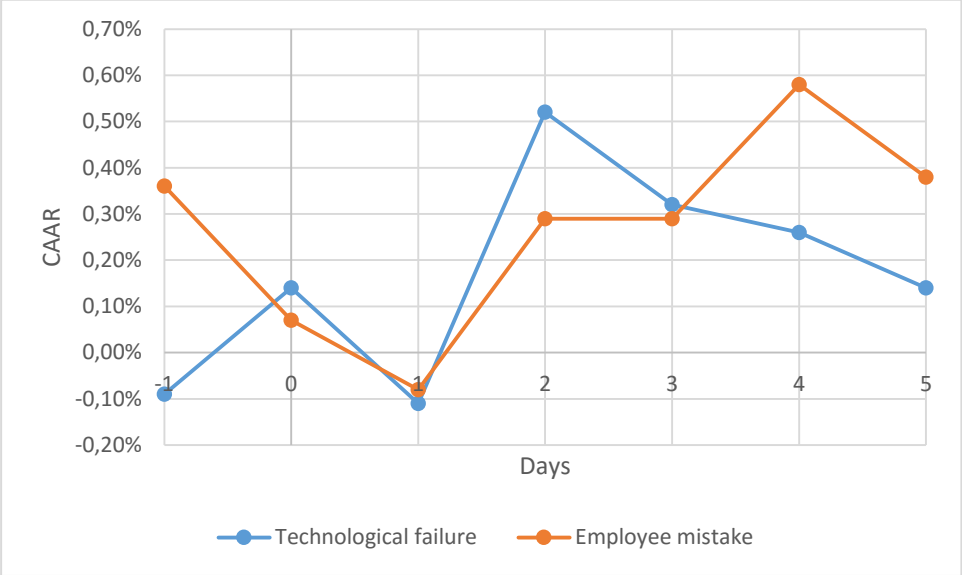
Table 7. Overall CAR and attack type

This table presents first the effect that the announcement of data breaches has on the overall CAAR, which corresponds to "All types of attacks". Then this table goes deeper into the attack type part by differentiating "Technological failure" and "Employee mistakes". The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
All types of attacks	128				
(0)		-0.06	60:68	-0.151	-0.519
(0, 1)		-0.35	61:67	-0.816	-1.03
(-1, 1)		-0.28	60:68	-1.239	-1.476*
Technological failures	43				
(0)		0.23	19:24	-0.674	-1.137
(0, 1)		-0.02	19:24	-1.416*	1.722**
(-1, 1)		-0.11	18:25	-1.807**	-2.026**
Employee mistakes	85				
(0)		-0.29	42:43	0.199	0.148
(0, 1)		-0.45	47:38	0.205	0.186
(-1, 1)		-0.08	43:42	0.166	0.099

Figure 6 shows that this effect lasts until the first day after the announcement of the breach and is directly offset the day after. Concerning employee mistakes, we also observe that there is a sharp decrease the day before the announcement until the day after. For the technological mistakes this effects is only observable between the announcement day and the day after and is significant as is shown in table 7.

Figure 6. Evolution of CAAR per attack type



5.1.2 Firm size

Moving to the third hypothesis, where the question is to find out if different firm sizes have any impact on the CAAR. This result might be mitigated since the number of observations is very small for the small sized companies, therefore the results might be biased. However, for the event period of (-1, 1) there is evidence that small companies are punished after a data breach. This leads to a decrease of firm value of 1.18% within 3 days of the event with a significance level of 1% and robust with the Corrado Rank test. The small companies lose on average 6,099,161 US dollars of firm value after a data breach.

This result is in line with previous findings. Most of the time, the small companies are more punished compared to the big firms and therefore investors do not take into account the risk of data breach when investing in small firms. The cyber security has to grow with the size of the firm. This is often the failure of a lot of start-ups which get blown up by the lack of cyber security evolution. Another reason for this is that small companies often have less protection and therefore are easier targets for hackers. They also have a slower time to solve the problem. Therefore this leads in the end to a bigger data loss.

A data breach in bigger firms does not have any impact on the CAAR on all the event periods. This is due to the fact that the problems are solved very quickly and the protection is better for those firms. We have seen that employee mistakes do not affect the CAAR. This is usually the biggest issue for the bigger firms and their main source of data breach. Therefore we can say that this is in line with our previous findings. The result are displayed in table 8.

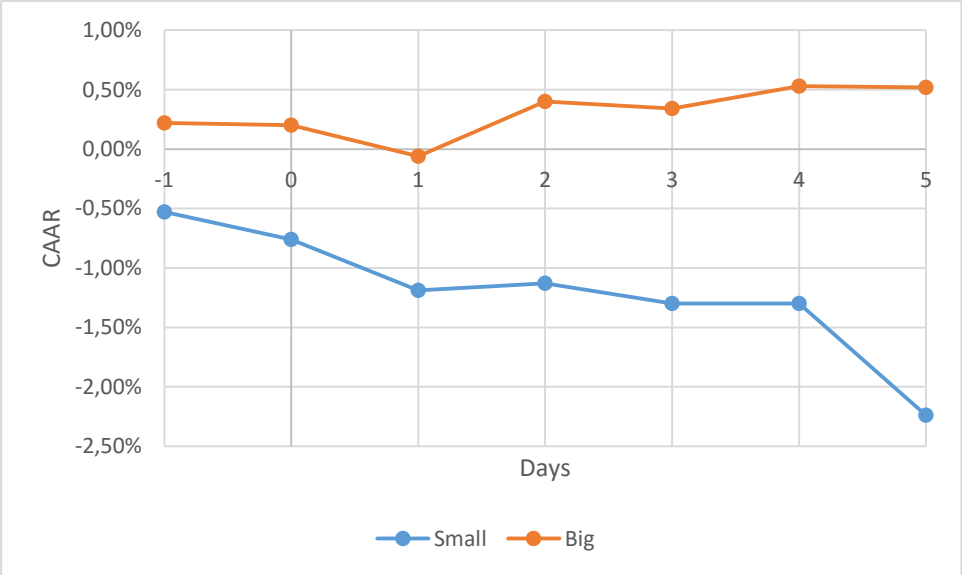
Table 8. Firm size

This table presents the effect that the announcement of data breaches has on firm value for small and big firms. Small firms correspond to a turnover of maximum 50 million Euro. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
Small company	25				
(0)		-0.23	11:14	-1.301*	0.126
(0, 1)		-0.66	11:14	-2.171**	-0.95
(-1, 1)		-1.18	9:16	-2.685***	-1.388*
Big company	103				
(0)		-0.02	49:54	0.472	-0.597
(0, 1)		-0.28	50:53	0.159	-0.632
(-1, 1)		-0.05	51:52	-0.059	-0.895

Figure 7 shows that small and big firms have a totally opposite effect after the announcement of the data breaches. The result of a data breach for a small firm leads to a decrease of the firm value. In this graph it is also observable that there might be a medium term effect for the firm size. This is because the firm value continue to decrease after the (-1, 1) event window. On the other hand, the big sized firm faces mostly a positive CAR between 0 and 0.5%. However, table 8 has shown that this is not significantly different from 0.

Figure 7. Evolution of CAAR per firm size



5.1.3 Industry

The fourth hypothesis tests which industries are mostly affected by a data breach. The strongest significant negatively impact on the CAAR is the financial services industry. This effect is instantly there the day of the event and decreases the market value of the firm on average by 0.96% with a 1% significance. This is also supported by the Corrado rank test. Previous literature also came up with the same result and explained this by the fact that this industry has highly classified information and detain the source of living, money. This does not particularly mean that investors are not aware of this, but means that investors simply react harsher on those breaches since they do not want to allow those breaches to happen. Investors cannot close their eyes for the financial sector, because too much sensitive information is in it. The financial industry loses 35,363,201 US dollars after a data breach on average.

From previous study, we have also seen that the retail industry is an industry highly targeting due to the amount of fluctuation of money or transition of money from customers. However, this industry does not have very sensitive information. The information they detain are mostly about individuals and less about firms. The reason why there is a significant positive effect on the CAAR with 5% significance on the day of announcement is probably because investors overestimated the impact of such breach and underestimate the speed of solution to those problems. Since these industries have less confidential data, one would think to protect this industry less. But this industry has proven to be effective on this side, because they still have a marketing strategy called “Big data consumer analytics”. This phenomenon is becoming big and thanks to what the customers buy and the things they look at, the retailer obtains information about what a person likes. (Erevelles, 2016). This information should still be protected and can be very sensitive.

Finally, all the other industries observed in our sample: Web, Technology, Utility and Energy do not have a significant impact on the CAR. This could mean that investors do not pay a lot of attention or attribute less severity from a breach in those industries. Another reason could be that those industries have sharpened their cyber protection and this makes the investors less worried about what could happen. The results can be found in table 9.

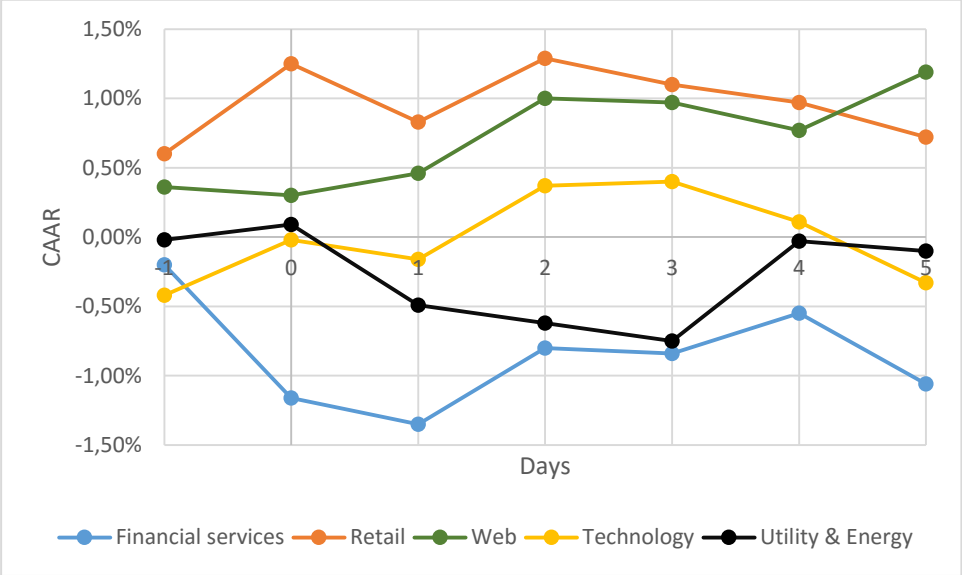
Table 9. Industry

This table presents the effect that the announcement of data breaches has on firm value in different industries. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
Financial services	37				
(0)		-0.96	12:25	-3.664***	-2.519***
(0, 1)		-1.14	16:21	-2.845***	-1.897**
(-1, 1)		-1.34	15:22	-2.856***	-1.863**
Retail	33				
(0)		0.65	18:15	2.246**	0.56
(0, 1)		0.22	18:15	1.272	0.234
(-1, 1)		0.82	18:15	1.302*	0.453
Web	16				
(0)		-0.06	7:9	-0.03	-0.014
(0, 1)		0.09	8:8	-0.192	-0.069
(-1, 1)		0.45	7:9	0.431	0.121
Technology	15				
(0)		0.40	7:8	0.784	0.204
(0, 1)		0.27	9:6	0.329	-0.085
(-1, 1)		-0.15	6:9	-0.496	-1.007
Utility and Energy	27				
(0)		0.11	15:12	0.98	1.358*
(0, 1)		-0.47	13:14	0.211	0.238
(-1, 1)		-0.50	14:13	-0.589	-0.422

The evolution of the CAAR per industry, in figure 8, confirms the statement made by table 9 and depicts the real effect of the announcement of the data breaches on the different industries. The financial services are hit the hardest from all the industries and this is very well captured in the event window (-1, 1). Even though we can observe trends such as decreases for the utility and energy and increases for Web and Technology, those are not significant effects. The increase on the announcement day for the retail industry is significant.

Figure 8. Evolution of CAAR per industry



5.1.4 Year

Overall, we can see that the years do not lead to a specific impact on the CAAR. However, some fluctuations over time are observable. The impact of data breach started to have a significant negative effect starting from 2011 with a 10% significance. The impact of a data breach resulted in a loss of 1.26% of firm value within 3 days in 2011 and was even increased to a loss of 2.66% within 2 days in 2012. In real term, this corresponds to a loss of 15,305,476 US dollars in firm value on average after a data breach. This can be explained by the start of significant attacks and the fear of the investors.

However, the year 2013 turned everything around. During this year, the data breach announcement resulted in a positive effect on the CAAR the same day or within 2 days at a 1% significance level. This is probably due to the new legislation in the USA and in Europe. In the USA, the Omnibus HIPAA rulemaking was created. This rulemaking forced entities to provide notice of a breach (Thomson Reuters, 2016). The same type of regulation was announced in the EU the 24th of June 2013. This made the investors more confident that data breaches will be reported and allowed them to have a clear overview on the situation. 2013 has been a year with a huge increase of firm value, which lead to an average of 46,939,220 US dollars gain after a data breach.

Finally, the last year, the data breach announcement again resulted in a drop of the CAAR. Probably, this is because investors wanted the government to do more about it and implement more regulation to stimulate firms to protect themselves more. This had a huge effect on the market value, since on average there is a loss of 48,836,392 US dollars of market price after the announcement of the breach. The results can be found in table 10.

Table 10. Year

This table presents the effect that the announcement of data breaches has on market value during different years in the study period (2006-2015). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
2006	13				
(0)		-0.08	5:8	-0.08	-0.393
(0, 1)		-0.12	7:6	0.364	0.221
(-1, 1)		0.03	7:6	0.389	0.389
2007	12				
(0)		-0.38	4:8	-0.765	-0.959
(0, 1)		-0.19	5:7	0.12	-0.378
(-1, 1)		-0.07	6:6	0.113	-0.045
2008	12				
(0)		0.80	9:3	0.325	1.135
(0, 1)		-0.74	5:7	-1.218	-1.189
(-1, 1)		0.21	6:6	-0.563	-0.585
2009	10				
(0)		-2.09	1:9	-0.912	-1.853***
(0, 1)		-2.80	2:8	-0.761	-1.444**
(-1, 1)		-2.03	2:8	-0.784	-1.444**
2010	13				
(0)		0.07	8:5	-0.064	0.31
(0, 1)		0.02	8:5	0.116	0.474
(-1, 1)		-0.67	8:5	-0.8	-0.212
2011	12				
(0)		-0.16	6:6	-0.182	0.053
(0, 1)		-0.42	5:7	-0.59	-0.342
(-1, 1)		-1.26	4:8	-1.474*	-1.629*
2012	13				
(0)		-1.63	5:8	-2.79***	-0.617
(0, 1)		-2.66	6:7	-3.273***	-1.309*
(-1, 1)		-1.48	7:6	-2.135**	-0.552
2013	16				
(0)		2.19	9:7	4.18***	0.852
(0, 1)		2.25	9:7	3.075***	0.822
(-1, 1)		2.14	9:7	2.371***	0.449
2014	14				
(0)		0.42	9:5	0.421	0.711
(0, 1)		0.67	7:7	0.066	0.2
(-1, 1)		1.00	7:7	0.081	0.159
2015	13				
(0)		-0.72	5:8	-1.243	-1.269
(0, 1)		-0.80	7:6	-1.027	-1.005
(-1, 1)		-1.64	5:8	-1.575*	-1.918**

5.1.5 Economic state

Hypothesis six tests if different economic states influence the CAAR in different ways. I observe that in a good state of the world the CAAR is affected negatively with a significance level of 10% for the event periods (0) and (-1, 1), while for the event period of (0, 1) this is significant at a level of 5%. This corresponds on average to a loss of 35,158,458 US dollars of market capitalization in good economic states. Therefore, we can say that there is a significant negative effect on the firm value when it gets breached during a good economic state of the world. However, when we are in a bad economic state the CAAR is not significantly different from zero. Therefore there is no effect on the CAAR during this time. This results are available in table 11.

The reason for this might be that investors are punishing more in good economic times and are scrutiny observing the firms with a loupe in order to closely observe what is going wrong. In bad economic times they do not want to punish the firms too badly since they mostly are already in a bad shape.

Table 11. Economic state effect

This table presents the effect that the announcement of data breaches has on firm value during different economic states of the world. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
Good economic state	45				
(0)		-0.14	22:23	-1.499*	-0.406
(0, 1)		-0.66	20:25	-1.773**	-1.135
(-1, 1)		-0.31	23:22	-1.283*	-0.643
Bad economic state	83				
(0)		-0.01	38:45	0.916	-0.36
(0, 1)		-0.19	41:42	0.292	-0.485
(-1, 1)		-0.25	37:46	-0.594	-1.379*

Figure 9 shows that there is a clear negative effect on the firm value if it gets breached during good economic state. We can also observe that this negative effect is often offset already the second day after the breach. We observe almost the same effect in bad economic state however the amplitude of the effect is less strong and therefore makes it insignificant.

Figure 9. Evolution of CAAR per economic situation

5.1.6 Recovery time

Recovery time is a factor that is highly linked with the CAR and has an effect on it. This is especially clear in this study by looking at the two extremes, which are short (0 days to a week) recovery time and long recovery time (over 30 days). The shorter it takes for a firm to recover the data and to solve the issues within the firm, the better it will be rewarded by the investors. While on the other hand, the longer it takes, the more costly it will become for the firm.

From this study we can see that a short recovery time leads to a significant increase on the CAAR. This is significant for a 5% level and the effect ranges between 0.40% and 0.89% of the firm value over the different event periods. Investors gain confidence and trust in firms that can solve data breaches very quickly and therefore reward those firms.

However, in the long run this can become very costly for the firm and can have a huge impact. With a significance level of 1% and robust at a 1% level as well we observe that if the company recovers after 30 days, then it will lose on average between 2.09% and 3.92% of its firm value. On average, a firm with a long recovery time would lose 31,592,917 US dollars of firm value after a breach. This shows that the company is not prepared for those attacks and cannot react quickly to it. Therefore, this creates fear for investors and they lose trust. Data breaches is expected to be solved quickly nowadays, and the companies should be exposed the less possible to those attacks. If the companies do not react, this must be punished by the investors. In our data set this occurred only 19 times, so this proves also that it is a rare event. Fortunately, most of the time the firms react quickly to it (82 observations). This results are displayed in table 12.

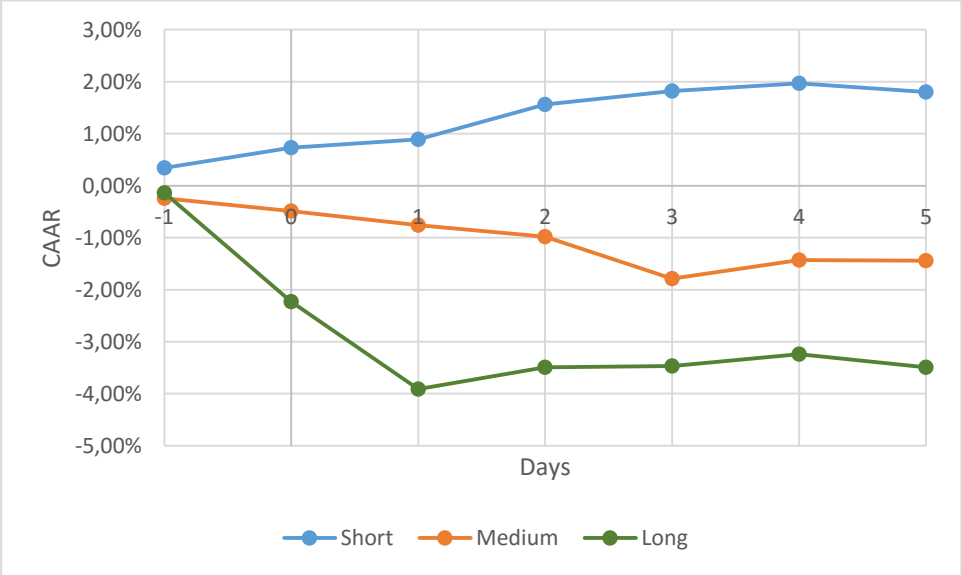
Table 12. Recovery time effect

This table presents the effect that the announcement of data breaches has on the market value of the firm with different recovery times after the breach. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
Short recovery time	82				
(0)		0.39	41:41	2.008**	0.552
(0, 1)		0.55	48:34	2.155**	1.029
(-1, 1)		0.89	46:36	1.811**	0.722
Medium recovery time	27				
(0)		-0.25	12:15	-0.112	-0.365
(0, 1)		-0.52	11:16	-0.804	-1.014
(-1, 1)		-0.75	8:19	-1.325*	-1.59*
Long recovery time	19				
(0)		-2.09	6:13	-4.637***	-2.86***
(0, 1)		-3.77	5:14	-5.279***	-3.861***
(-1, 1)		-3.92	4:15	-4.783***	-3.36***

The results from table 12 are illustrated in figure 10 and agree with those results. This graph accentuates even this effect by showing that this effect remains over the medium-term. However, statistical tests for the significance over the medium-term have not been performed.

Figure 10. Evolution of CAAR per recovery time



5.1.7 Continent

The 8th hypothesis is to test if being cyber attacked in different continents has an effect on the CAAR. As explained earlier, all the firms have their headquarters in the USA and have their stock prices quote on the American market. Therefore the attacks in Europe and in Asia are attacks on the subsidiaries. The first thing that is noticeable is that data breaches on home ground are punished by investors with a significance level of 1% and robustness of 5% for the event period of (-1, 1). If a firm is attacked in the USA and is quoted in the USA, this firm would lose on average 0.70% of its firm value. In this data set, this corresponds to an average loss of 12,992,312 US dollars in America. The reason for this might be that the defences in their home should be optimal and the strongest.

When we move to the results for Europe we observe a significant positive effect of about 3% firm value at a significance level of 1%. The reason for this might be that the sample for Europe is very small (13 observations). This could also be because this is not their core facility and not their home place so investors are less strict about the fact that a base abroad gets hacked. But this does not happen frequently in our sample results. The results can be found in table 13.

Table 13. Continent effect

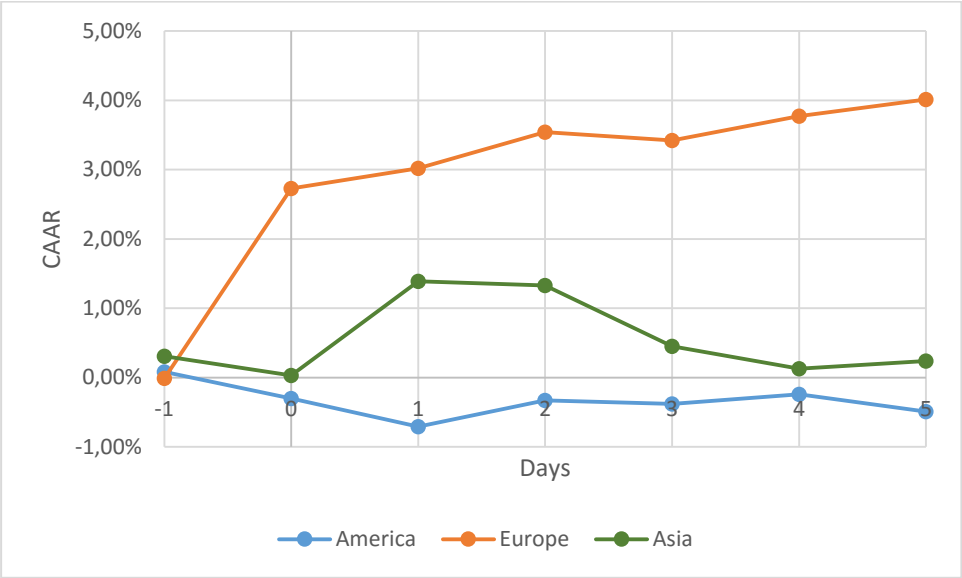
This table presents the effect that the announcement of data breaches has on firm value in different continent in the world. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank
America	112				
(0)		-0.38	51:61	-1.62*	-0.946
(0, 1)		-0.78	49:63	-2.197**	-1.643*
(-1, 1)		-0.70	54:58	-2.388***	-1.661**
Europe	13				
(0)		2.74	8:5	4.544***	1.288*
(0, 1)		3.03	9:4	3.35***	1.034
(-1, 1)		3.02	4:9	2.577***	-0.139
Asia	3				
(0)		-0.28	1:2	-0.55	-0.478
(0, 1)		1.08	3:0	1.119	1.111
(-1, 1)		1.39	2:1	1.13	1.076

Figure 11 presents a graphic version of the evolution of the CAAR per continent. The conclusion about Europe should be taken with caution since the number of observations is very small. However, we observe a significant increase in firm value after a breach in Europe. This increase is very sharp on the announcement day and then stays almost constant on the medium term. On the other hand, in

America we observe that the firm value declines slowly from the day before the breach until the day after. This effects stays steady over time.

Figure 11. Evolution of CAAR per continent



5.1.8 Number of times hit

The last hypothesis observes if the impact of the first time being breached is the same as being breached more than once. Most of the firms in this data set have been breached only once (95 observations). The result from this study shows that there is a significant negative impact when the company gets breached for the first time. This corresponds to a decrease of 0.30% of the market value of the firm within 3 days around the breach and is significant at a 10% level. On the other hand, there is no significant impact on the firm value if the firm gets breached more than once. Two reasons can be given for this. First, once a company has already been hacked it is more alert and has a faster respond time to the breach. Since the firm has experienced this before and knows the procedures, getting hacked a second time leaves less scratches. The second reason could be that the investors anticipate the fact that this firm could be hacked again and therefore account for it already in the stock price. The results from this test are exposed in table 14.

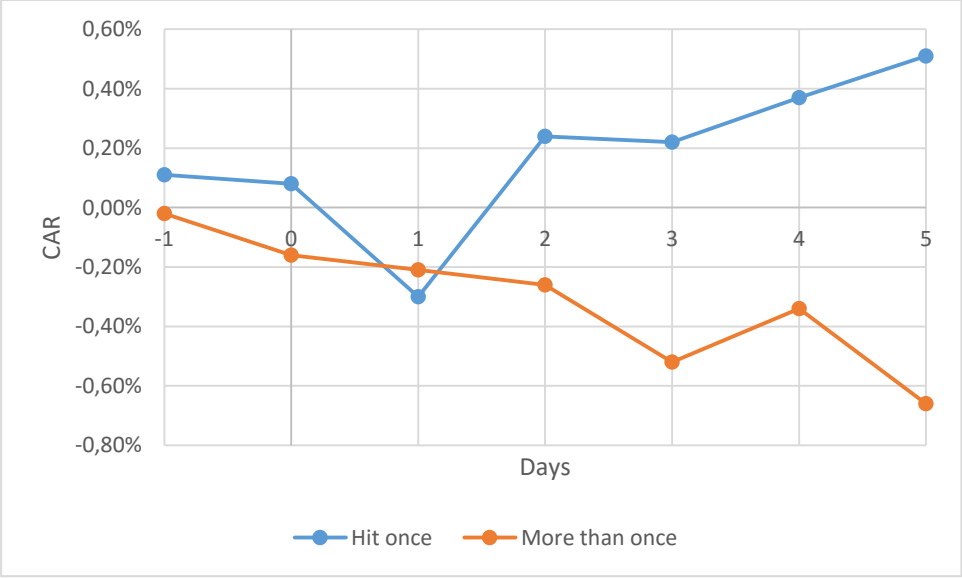
Table 14. Number of times hit

This table presents the effect that the announcement of data breaches has on the market value of the firm the first time it gets breached and when it has already been breached before. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement and then presents the results of the CAAR (0), (0, 1) and (-1, 1). The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively.

Event window	N	Mean CAR (%)	Positive:Negative	Patell Z	Corrado Rank test
One time	95				
(0)		-0.03	44:51	-0.158	-0.453
(0, 1)		-0.41	43:52	-1.065	-1.284*
(-1, 1)		-0.30	46:49	-1.530*	-1.446*
More than once	33				
(0)		-0.14	16:17	-0.03	-0.247
(0, 1)		-0.19	18:15	0.2	0.100
(-1, 1)		-0.21	14:19	0.156	-0.464

The illustration for the last hypothesis shows that the first time a company gets breached, it will encounter a sharp decline mostly the day after the announcement. However, this effect is then reversed the day after. On the other hand, if the company gets hit more than once we do not observe any significant effect, but it can be clearly seen that there is an impact over time. Significance test to proof this has not been executed however it would interesting to see if this would lead to a loss of market value on a longer time period.

Figure 12. Evolution of CAAR per times hit by cybercrime



The results from the detailed overview of the effect on the CAAR show that the firms that are mostly punished by investors are small financial companies that encounter a technological failure especially during the years when no new legislation about data breaches are passed. Adding to this good economic times is not in favor of companies. Finally, the longer the recovery time, being affected in

the home country and being hit for the first time, also affects significantly the CAAR. This does not particularly mean that investors do not take it into account. It might also mean that they do not want to take it into account and will not accept those particular events to happen.

The other types of firms are more or less immune against data breach effects since investors already account for it in the stock price or judge that a data breach would not play a major role on them.

5.2 Cross-sectional results

The cross-sectional results will provide a bigger picture on all the variables analyzed previously. We are going to observe if all variables observed previously have globally a significant impact on the CAR. The event period chosen for the CAR is the (-1, 1) since this has been proven to be the best estimator based on previous literature and this has also been shown statistically in part 4.3.

Since we have previously seen that the CAAR is not significantly different from 0, we should not expect a lot of significant result.

Table 15 shows that only one variable has a significant effect on the CAR (-1, 1). The recovery time is highly significant and lead to a loss on average of 2.2% at a 1% confidence level. This tells us that most of the variables are taken into account by investors, especially those stated in previous literature. However, recovery time is a new variable added to this regression and is proven to have a significant effect. This shows that investors should take this variable into account when investing in stock, since it can also affect the firm value of a firm.

Finally, a noticeable observation is that the adjusted R square is very low which proves that there are still more variables that affect the CAR after a data breach event.

Table 15. Cross-sectional regression

This table presents an OLS regression with dependent variable CAR (-1, 1). Independent variables include Attack type, which corresponds to a dummy: 0 if the company faced a technological failure and 1 if the company faced a human mistake. Firm size is also a dummy variable, where a small firm corresponds to 0 and a large firm to 1. The industry variable is an ordinal variable. The following numbers have been assigned per industry: 1 corresponds to the financial industry, 2 the technological industry, 3 the retail industry, 4 the web industry and 5 the utility and energy industry. The variable year has been assigned in the same way, chronologically, where 2006 corresponds to 1 and 2015 to 10. The economic state of the world has been defined as a dummy variable, 0 if the economic state is good and 1 if there is an economic depression. The recovery time corresponds to the time the company takes to solve the problem within the company, 1 is a short recovery time (shorter than a week), 2 is a medium recovery time (between a week and a month) and 3 is a long recovery time (longer than a month). The 3 continents have been assigned the following values: 1 for America, 2 for Europe and 3 for Asia. Finally, the number of times hit is also a dummy variable, where 0 is the dummy for being hit only once and 1 if the company got hit more than once. The market model has been used in order to estimate the abnormal returns over 120 days before the announcement. The signs ***, ** and * denote statistical significance at 1%, 5% and 10% respectively. Since there is heteroscedasticity, robust standard errors are used.

	Coefficient	Std. Error	T statistic
Intercept	0,019	0,022	0,8911
Attack type	0,002	0,013	0,1758
Firm size	-0,004	0,012	-0,3736
Industry	0,002	0,003	0,6012
Year	-0,003	0,003	-0,8927
Economic state	-0,002	0,015	-0,1333
Recovery time	-0,022	0,006	-3,415***
Continent	0,022	0,014	1,540
Times hit	0,007	0,011	0,614
Observations	128		
R square	0,126		
Adjusted R Square	0,067		

To conclude, we can say that overall data breaches do not have any impact on the stock prices of a firm anymore. Therefore we can accept the H1A hypothesis and affirm that investors take data breaches into account. We can also assume that the significant effects we observed in this study could be based on the fact that investors do not want to take this into account and still want to punish the firms that do not put the data security at the maximum priority. However, recovery time is an important variable that has to be taken into account nowadays. This can be estimated through peers from the same industry and from historical data in order to have an overview on how long it takes for a firm to recover from a breach.

CHAPTER 6 Conclusion

This paper presents the effects of data breaches on the stock prices of 107 public firms and 128 events between 2006 and 2015. Since data breaches is an increasing phenomenon nowadays I thought it would be interesting to find out if investors do take it into account when investing in public stocks and if they do, what do they especially look at. In order to come to a conclusion I have analysed different variables they could have taken into account and that might affect the CAAR: (i) attack type, (ii) firm size, (iii) industry, (iv) year, (v) economic state of the world, (vi) recovery time, (vii) continent and (viii) numbers of times hit by data breaches. Two main methods have been used in order to observe if those factors play a role in their decision making. First the event study methodology to give a detailed view within each variable and then a cross-sectional regression to give a bigger picture on those variables.

The main result from the event study methodology is that overall the investors already account for the data breaches in the stock price since the CAAR is not significantly different from 0. However, investors do not take into account some specific characteristics. First of all, there is a negative impact if the firm has been breached because of a technological failure. Investors want the firms to be well protected and therefore they punish the firms if they get breached because they are not enough protected on a technological level. The important firm characteristics to look at are the size and the industry. We have seen that small and financial firms are evolving in a sensitive area. The small firms, because they mostly do not improve their cyber protection at the same speed as they grow and financial firms since they control the most important resource: money and could lead to a contagion on other companies. The investors also do not take into account those characteristics which lead to a significant negative effect on the market value if they get breached. It has also been seen that there might be a link between the years and the release of new law related to data breaches. In years where new data breach laws pass we can observe a positive cumulative abnormal return, while this effect is negative when nothing happened on that side. Data breach regulations give the investors more trust in the companies and also give them hope that data breaches are an important matter today. This study has also shown that investors punish companies more harshly if the world is in a good economic situation. Concerning the recovery time, we can clearly say that investors do not take into account long recovery times. Data breaches must be solved quickly after the breach and the best outcome would be to solve it in a short time frame, which corresponds to a week. On the continent side, we have seen that investors do punish firms if they get breached in America. However, the effect for Europe and Asia is less clear due to the sample size. Finally, we have seen that if the firms faces a data breach for the first time, it will have to deal with a negative impact on firm value, because this was not often expected by investors. If the company gets hit more than once, we can see that the investors already anticipate the possibility of an attack.

The second method, which is the cross-sectional regression, gives us an overview on all the variables selected and helps us to observe which variable the investors do not take into account. From this study, we observed that most variables were accounted for. The only variable which has a significant negative impact after the announcement of the breach is the recovery time. One way to anticipate how long it would take for a firm to solve the data breach problem is by looking at their peers from the same industry and observe how much time on average they needed to solve the data breach.

Overall, this study has shown that investors do take most data breaches into account. However, it shows as well that in some specific areas negative cumulative abnormal returns occur. This is not only due to the fact that the investors do not take it into account, but also because simply anticipating data breaches would not underline the danger of it. This danger has to be reminded at each breach even more in industries with contain highly sensitive information.

The result that the data breaches does not have an impact on the CAAR is not in line with previous findings. However the overall negative impact that data breaches has on the CAAR has declined over time. Cavusolgu et al. (2004) found that the negative effect was 2.1% while Morse et al. (2011) found that this effect was 0.84%. Similarities with other literature have been found on the side of the company size, where Cavusoglu et al. (2004) also found that small size companies have a significant negative impact on the CAAR. The results that financial companies have a significant negative impact on the CAAR is also in line with the finding from Morse et al. (2011).

As all studies, this paper also contains limitations. First of all the number of observations used for this study is very limited due to the fact that all the data has been collected manually and due to time constraints. This has led to ambiguous results for Europe and Asia since the sample size is very small. However, in percentages this corresponds to reality. This does not mean that Europe and Asia are getting less attacked than America, but they are less open in disclosing information about breaches. Another limitation is that for the firm size variable a lot of attacks are against self-employed people (company of 1 person) and those attacks are not reported, which leads also to a bias in the sample.

Further research could be done on the medium-term effect of data breaches. This study only covered the very short time effect of data breaches, however, from the graphs we could see that there also might be a medium-term impact for the size of the firms and the number of times a company gets breached. It would also be interesting to observe what influences firms to disclose information about data breaches since a lot of companies still do not disclose breaches. This information is really important in order to avoid a Domino effect between the firms.

REFERENCES

Andobaidoo, F., Amoakogyampah, K., & Osei-Bryson, K. (2010). How internet security breaches harm market value. *IEEE Security & Privacy, 8*(1), 36-42.

Boehmer, E., Masumeci, J., & Poulsen, A. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics, 30*(2), 253-272.

Brounen, D., & Derwall, J. (2010). The impact of terrorist attacks on international stock markets. *European Financial Management, 16*(4), 585-598.

Campbell, C., & Wasley, C. (1993). Measuring Security Price Performance Using Daily Nasdaq returns. *Journal of Financial Economics, 33*(1), 73-92.

Campbell, K., Gordon, L., Loeb, M., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: empirical evidence from stock market. *Journal of Computer Security, 11*(3), 431-448.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach on market value: Capital Market Reaction for breached firms and internet security developers. *International Journal of Electronic Commerce, 9*(1), 69-104.

Chantler, A. N., & Broadhurst, R. G. (2006). Social Engineering and Crime prevention in Cyberspace. *Queensland University of Technology*.

Corrado, C. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics, 23*(2), 385-395.

D'Amico, A. (2000). What does a computer security breach really cost?. *The Sans Institute*.

Das, S., Mukhopadhyay, A., & Anand, M. (2012). Stock market response to information security breach: A study using firm and attack characteristics. *Journal of Information Privacy & Security, 8*(4), 27-54.

Erevelles, S., Fukawa, N., Swayne, L (2016). Big Data analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897-904.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review, 10*(1), 1-21.

Fama, E. F., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance, 47*(2), 427–465.

Feldman, M., & Santangelo, G. (2008). New perspectives in International Business Research.

Frisby, D. (2014). Bitcoin: The future of money?.

Garnitz, J., Nerb, G., Wohlrabe, K., Boumans, D., Oesingmann, K, & Nikolka, T. (2016). CESifo World Economic Survey May 2016. *CESifo World Economic Survey, 16*(2).

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

Goe, S., & Shawky, H. (2009). Estimating the market impact of security breach announcement on firm values. *Information & Management, 46*(7), 404-410.

Hadnagy, C. (2010). Social Engineering – The Art of human hacking.

Hoffer, J. A., & Straub, D. W. (1989). The 9 to 5 Underground: Are your policing computer crime?. *Sloan Management Review, 30*(4), 35-44.

Hovav, A., & D'Arcy, J. (2004). The impact of Denial-of-Service Announcements on the market value of firms. *Risk Management and Insurance Review, 6*(2), 97-127.

Hovav, A., & D'Arcy, J. (2004). The impact of virus attack announcements on the market value of firms. *Information Systems Security, 13*(3), 32-40.

Howard, J. D. (1997). An Analysis of Security Incidents on the Internet 1989-1995. *PHD Thesis, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA.*

Huq, N. (2015). Follow the Data: Analysing breaches by Industry. *Trend Micro Analysis of Privacy Rights Clearinghouse 2005-2015 Data breach records.*

Kannan, K., Rees, J., & Sridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce, 12*(1), 69-91.

Kelly, B. J. (1999). Preserve, Protect, and Defend. *Journal of Business Strategy, 20*(5), 22-26.

Lyon, J. D., & Barber, B. M. (1999). Improved methods for tests of long-run abnormal stock returns. *Journal of Finance, 54*(1), 165-202.

MacKinley, A. (1997). Event studies in economics and finance. *Journal of Economic Literature, 35*(1), 13-39.

Manworren, N., Letwat, J., & Daily, O. (2016). Why you should care about the Target data breach. *Business Horizons, 59*(3), 257-266.

McMillan, R. (2011). After delay, Hacker to show flaws in Siemens industrial gear.

Morse, E., Raval, V., & Wingender, J. (2011). Market Price effects of data security breaches. *Information Security Journal: A Global Perspective, 20*(6), 263-273.

Patell, J. (1976). Corporate forecasts of Earnings per share and Stock Price Behavior: Empirical Test. *Journal of Accounting Research, 14*(2), 246-276.

Power, R. (2001). 2001 CSI / FBI Computer Crime and Security Survey. *Computer Security Issues and Trends, 2*(1), 1-18.

Power, R. (2002). 2002 CSI /FBI Computer Crime and Security Survey. *Computer Security Issues and Trends, 3*(1), 1-28.

Savoie, M. J. (2012). Building Successful information systems: Five best practices to ensure organizational. *Harvard Business Review, 5*(2), 57-64.

Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics, 5*(3), 309-327.

Shackelford, S. J. (2012). Should your firm invest in cyber risk insurance?. *Business horizons*, *55*(4), 349-356.

Sprecher, R., & Pertl, M. (1988). Intra-Industry Effects of the MGM Grand Fire. *Quarterly Journal of Business and Economics*, *27*(2), 26-31.

Van Dongen, Y. (2014). Are financial firms well protected against cybercrime?. *Bachelor Thesis ECE Lyon.*

APPENDIX

APPENDIX A Other type of references

European Union's European Commission and the U.S. small Business Administration, 2010, available at http://smbresearch.net/sizing-up-smb/

Gemalto, 2015, available at http://breachlevelindex.com/assets/Breach-Level-Index-Infographic-2015-1500.jpg

Kapersky Lab, 2013, available at http://support.kaspersky.com/viruses/general/2313

Lyne J., 2013, available at https://www.youtube.com/watch?v=fSErHToV8IU

Merriam-Webster Online, 2016, available at http://www.merriam-webster.com/

Norton, 2016, available at http://us.norton.com/security_response/malware.jsp

Statista, 2015, available at http://www.statista.com/statistics/193436/average-annual-costs-caused-by-cyber-crime-in-the-us/

Thomson Reuteurs, 2016, available at http://uk.practicallaw.com/6-502-0467#a762707

Verizon, 2012, available at http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf

APPENDIX B Company names and countries used for the research

Company names / Country in which the data breach took place			
Aaron's	Canada	Fidelity National Information Services	USA
ACORN	USA	Flanigan's	USA
Adobe	USA	Gap Inc	USA
Allstate	USA	General cable	USA
American Express	USA	General Electrics	USA
AOL	USA	General Motors	USA
AON	USA	Genesco	USA
Apple	USA & Germany	Genworth	USA
Applied micro circuits	USA	Global Payments	Mexico
Aramark	USA	Google	Hong-Kong
AT&T	USA	Grainger	USA
Avis budget	USA	Groupon	USA
Bank of America	USA & Canada	Gymboree	USA
Barclays	USA	Harley-Davidson	USA
Barnes & Noble	USA	Heartland	USA
Bebe	USA	Hewlett Packard	USA
Bisys Group	USA	Home depot	USA
Blackbaud	USA	Honda motors	USA
Blockbuster	Canada	Humana	USA
Boeing	USA	IBM	USA
Broadridge	USA	Invitrogen	USA
Charter communication	USA	Kelly Services Inc	Mexico
Chase	USA	Kinetic concepts	USA
Cheesecake Factory	USA	Korn Ferry International	USA
Citigroup	USA & Canada	Kraft food	USA
Comcast	USA	LinkedIn	Canada
Compucredit	USA	Luxottica	USA
Countrywide Financial Corp	USA	Macy's	USA
Domino's Pizza	France	Massmutual	USA
Dun & Bradstreet	Canada	Matson	USA
Ebay	UK	McDonald	USA
Equifax	USA	Meetme	Canada
Expedia	USA	Merrill Lynch	USA
Exponent	USA	Metlife	Mexico
Facebook	USA	Microsoft	USA
Fairpoint communication	USA	Monster.com	USA

APPENDIX B Company names and countries used for the research

Company names / Country in which the data breach took place	
Morgan Stanley	USA
NASDAQ	USA
Netflix	Argentina
New York Times	USA
Nokia	Finland
Nordstrom	Canada
Nvidia	USA
Pfizer	USA
Prudential financial	USA
RBS Worldpay	USA
Seachange	USA
Sears	USA
Sony	Germany & Brazil
Sourcefire	USA
Sprint	USA
Staples	USA
Starbucks	USA
Target	USA
TD Ameritrade	USA
Tesco	UK
Textron	USA
Thomson Reuter	USA
Time Warner Cable	Canada
Toyota	Japan
Transcend	Brazil
UPS	USA
Verizon	USA
Vodafone	Germany & UK
Wachovia bank	USA
Wal-Mart	USA
Washington Post	USA
Wellpoint	USA
Wells Fargo	USA
Wendy's	Canada
Yahoo	Japan, Canada & Germany