Norwegian School of Economics
Bergen, Fall 2020

What the Hack?

An Empirical Analysis of the Stock Market Reactions to Hacking Announcements

Embla Kleiv Jansen & Stine-Mari Stavik

Supervisors: Steffen Juranek & Carsten Bienz

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This master thesis is a part of our Master of Science in Economics and Business Administration at the Norwegian School of Economics. We are both majoring in Finance and have an interest in technology.

Cyber security is a relevant subject in 2020, and we wanted to increase the awareness of the threat it imposes. Thus, when our supervisors suggested quantifying the effect of hacking announcements on the stock exchange, it sparked our interest. We believe that our thesis will contribute to IT managers' and top executives' awareness of financial consequences of hacks.

In the process of writing the thesis, we have enhanced our knowledge about cyber security, the drivers behind movement in security prices and statistical methods. Working with the thesis has been both challenging, engaging, and educational as we have implemented what we have learned throughout our studies. In addition, we have learned to code R-Studio and LaTeX, which will be valuable in our future employment.

At last, but not least, we want to express our gratitude towards our supervisors Steffen Juranek and Carsten Bienz who have contributed with useful and professional feedback in the process. We are especially grateful for their availability to answer questions and engage in discussions at short notice. We also want to thank our family and friends for all the support they have given us through our years at NHH.

Norwegian School of Economics

Bergen, December 2020

Embla Kleiv Jansen

Stine-Mari Stavik

Abstract

To raise awareness of the financial consequences for companies that do not safeguard personal data, this thesis investigates the stock market reaction following hacks. Furthermore, it investigates the role consumers and regulatory agencies play in inflicting financial consequences on companies that are hacked. While previous studies have focused on data breaches in general, this thesis focuses on hacks, because hacking is the most dominant form of data breaches and is increasing in frequency. The thesis contributes to existing literature by examining 42 of the world's largest hacks announced between 2007 and 2020. The research questions are answered by using event study methodology as described by MacKinlay (1997).

We find an average negative stock market reaction of 1.7% on the first trading day following the announcement of the hacks. Moreover, we find that the stock prices do not fully recover within the following ten days, indicating that shareholder value is at risk. When investigating the role of consumers, we find that when many client's records are exposed in the hack, the stock market reaction is stronger. This may be because investors expect that the consumers will use their market power to punish the companies that have been hacked, and that this will decrease the net value of the company. More surprisingly, we find no statistically significant impact when the data exposed in the hack is sensitive to the customers. Finally, we explore the stock market reaction to hacks prior to and after the implementation of the GDPR in 2018, with a subsample of 33 events. The GDPR has raised the maximum fines for companies that are hacked, however, we do not find evidence of stronger stock market reactions after it was put into effect in our data sample.

Our findings suggest that IT managers and top executives should be concerned with protecting the personal data that the company stores, because there exists a trade-off between investing in cyber security and carrying the costs of being hacked.

Keywords – Hack, Data breaches, Cyber security, Regulatory agencies, IT managers, GDPR, Event study, Consumers

Contents

List of Figures

List of Tables

1 Introduction

Cyber-attacks were identified as one of the top five global risks to economic growth by The World Economic Forum in 2019 (The World Economic Forum, 2019). As the world is getting more digitized the amount of personal data that is stored is increasing. At the same time, organizations had a 29.6% chance of getting breached within two years in 2019. This makes up an increase of 7% from 22,6% in 2014 (IBM Security, 2019). Hence, there is a growing risk that private information is leaked and abused. What if someone had access to all the emails you have ever written, or someone used your credit card? These issues have already affected millions of people across the globe, and individuals carry one part of the cost.

However, once the personal data is online, it is the organizations and companies that store the data that have the power to protect it from being exposed and abused. The question is whether companies are aware of the risk of getting hacked, and the expected financial consequences of not safeguarding personal data. There is probably a trade-off between investing in cyber security and paying the price of being hacked. The primary objective of this thesis is to shed light on the financial consequences for companies that are hacked to contribute decision relevant information about cyber security investments.

There are mainly two groups of corporate stakeholders that can inflict financial consequences on companies that do not protect the personal data that they collect from their customers. The first stakeholder is consumers who can vote with their wallets, and with their personal data, for the companies that meet their requirements and expectations. The second stakeholder is regulatory agencies that can pass laws which empowers the authorities to impose sanctions on companies for not safeguarding personal data. This way the external cost of a hack can be internalized by the companies. The secondary objective of this thesis is to explore the role of these stakeholders in inflicting financial consequences on companies that are hacked.

1.1 Background

Data Breaches and Hacks

A data breach is when a company experiences an intentional or inadvertent exposure of confidential information to unauthorized parties (Cheng, Liu, & Yao, 2017). Corbet and Gurdgiev (2019) divide data breaches into four groups depending on the cause of the data breach: (1) external data breach or hack, (2) employee release, (3) lost, stolen, or discarded internal data devices and (4) unintentional disclosure. Hacking (1) has grown substantially, while the other forms of cyber-attacks have remained constant (Corbet & Gurdgiev, 2019).

Hacking is convenient for criminals as there is no geographic barrier and cheap to perform. In addition, it is hard to identify the criminals behind the attack. Hence, it is expected that the number of hacker attacks will continue to increase and that the hackers will become more sophisticated in their methods (Jang-Jaccard & Nepal, 2014).

The Financial Consequences of Data Breaches

There are several possible ways data security breaches can impact the financials of a company. The most apparent are the tangible costs. First, the loss of revenue, as companies often must shut down their services to stop the attack from evolving. Second, employees must spend time on the aftermath of the attack, which slows down the productivity of the company. Third, there are direct costs related to repairing and replacing software and hardware that have been damaged in the data breach. Fourth, fees, compensations and fines related to legal prosecution from authorities and consumers that are affected by the data breach is a potential cost (IBM, 2019; The AME group, 2020; Yayla & Hu, 2011).

The intangible costs are harder to measure but can be just as devastating to a company's financials as the tangible costs. For example, damaged brand reputation and loss of customer trust and loyalty (Drinkwater, 2016). Customers may worry that the company is not trustworthy after the loss of private information and can be hesitant to use their products and services in the future. Bad reputation regarding data privacy can be hard to restore and hence affect the company's performance for years. This effect can be further strengthened if customers switch to competitors, which will change the competitive

landscape. Finally, there is a risk that investors are more hesitant to invest in the company, which will increase cost of capital (IBM, 2019; The AME group, 2020; Yayla & Hu, 2011).

The tangible and intangible costs impact the stock price of a company through the traditional financial valuation models, as explained in Appendix 1.

Data Protection Laws

As a reaction to the growing trend of cyber security events, there is an increasing regulatory recognition of the threat security breaches impose. In 2018 the General Data Protection Law (GDPR) was implemented in the European Union (EU). According to the EU (2020) it is the toughest privacy and security law in the world. The GDPR enables authorities to sanction companies that do not safeguard personal data of citizens in the EU. The EU informs that the maximum penalty for companies and organizations is €20 million or 4% of global revenue, whichever is higher. Data protection authorities can also issue sanctions, such as bans on data processing or public reprimands, that can lead to indirect cost such as damage of the brand or limitations of their operations (The European Union, 2020). Fines and sanctions will reduce the dividend payments for investors.

1.2 Research Question Development

In this thesis, we attempt to quantify the stock market reaction for companies that are hacked, to make companies that store personal data aware of the potential financial consequences of being hacked, and to facilitate adequate decisions about investments in cyber security. Furthermore, we focus on hacks, as opposed to unintentional exposure of data. This is because the risk of unintentional exposure of data can be mitigated by implementing internal routines and physical barriers. However, protection against hacking attacks demands more sophisticated and complex protective measures. In addition, hacking is the most dominant form of data breaches and has grown substantially in the 2000's (Corbet & Gurdgiev, 2019). This makes hacking highly relevant.

Most research papers on the topic focus on small and medium sized hacks and data breaches and some find significant stock market reactions while others do not. In this thesis we investigate the stock market reaction following the announcement of mega hacks, due to limited data availability of smaller hacks. However, we believe this is relevant to

decision-making on investments in proactive measures to avoid successful hacks. According to IBM Securities (2019), the average size of a data breach was 25 575 records lost in 2019. In this thesis, the average number of records lost is 36.7 million. The data sample consists of 42 large hacks from 2007 till 2020, where personal data was exposed.

Hypothesis 1:
H1: A firm-specific hack will influence the stock value negatively following the announcement of the hack to the public.

Our secondary objective is to investigate the determinants of the stock market reaction, to understand the role customers and regulatory agencies play in inflicting financial consequences on companies that are hacked. The number of individual records lost is used as a proxy to the number of customers that are affected by the hack. This is because at the announcement of a hack the number that is reported as the number of records lost is usually the same as the number of accounts or clients affected. The hypothesis is that when many customers are affected, it is more likely that the company experience a decrease in revenue because the customers chose to not buy the product or service again. Thus, the tangible and intangible costs to the company increase, and the stock market reaction is stronger. To measure the severity of the consequences to the customers who are victims in the hack, we use the data sensitivity of the data extracted. It is reasonable to think that when sensitive information is extracted, the potential damage it can do to the victims is higher and that the customers will punish the company accordingly.

Hypothesis 2:
H1: A firm-specific hack will have a stronger negative impact on the stock value following the announcement of the hack if it is expected to have great impact on the customers of the firm. The impact on the customers is measured as the number of records lost and the data sensitivity of the data exposed in the hack.

To explore the effect of regulation on the financial consequences to companies that are hacked, the stock market reaction to hacks that occurred before the GDPR was implemented is compared to hacks that happened after. The GDPR enables authorities to sanction companies that do not safeguard personal data of citizens in the EU. The GDPR builds on the 1995 Data Protection Law that is also meant to make businesses protect

data. However, the fines were miniscule and had little deterrent effect according to the EU commissioner Viviane Reding (2014). One will therefore expect that the companies breached after the GDPR was put into effect will experience a stronger market reaction than hacks before the GDPR. This because the potential sanctions increase the expected costs to the companies that are hacked. This result is expected regardless of where the company is registered as long as there are EU citizens in the customer base, because the company is then subject to the law.

> *Hypothesis 3:*
> *H1: A firm-specific hack will have a stronger negative impact on the stock value following the announcement of the hack if the firm is subject to regulations that empower authorities to sanction the specific firm for not safeguarding the personal data.*

The above hypotheses will be investigated using event study methodology and cross-sectional analysis.

1.3 Structure

This thesis consists of ten sections. In the first section we have now introduced the topic of hacks, defined the research question and presented some background information. The second section accounts for existing literature and the contribution of this thesis. The third section describes the market efficiency theory, followed by a thorough explanation of the event study methodology in section four. In section five the selection criteria for the sample are presented, followed by descriptive statistics of the final sample. Section six investigates the research question and presents the results from the analysis. In the seventh section the analysis is discussed and compared to the findings of similar research papers. Finally, in section eight and nine robustness tests are conducted and the analysis is assessed critically, before the thesis is concluded in section ten.

2 Relevant Literature

In this section event studies and reports about the cost of data breaches are presented and discussed. The literature regarding large data breaches and hacks is somewhat restricted in amount. This is potentially because the number of publicly announced mega hacks is limited. In addition, research is quickly outdated due to the rapid development on this field. Lastly, we elaborate on the thesis' contribution to existing literature.

2.1 Studies on the Cost of Data Breaches to Companies

IBM Security and the Ponemon Institute (2019) have published a report on the cost of data breaches, that is based on interviews of 507 small and medium sized companies subject to data breaches between July 2018 and April 2019. The report does not account for mega breaches such as the breaches of Equifax and Facebook. According to IBM Security and the Ponemon Institute, the last five years the average total cost of data breaches has grown by 12% to $3.92 million per company. Additionally, the life cycle of each data breach is longer than before, and the data breaches impact the organizations for years. In the report it is stated that the health sector has the highest average industry cost when breached of $6.45 million, which is likely due to their access to personal data. The conclusion is that organizations need to account for the risk of data breaches. This report offers valuable insight to the company perspective of data breaches, however one should be aware that interviews as a research method can be subject to biases. Additionally, it is uncertain if they are able to quantify the intangible costs, such as loss of revenue due to reputational damage. Hence, the cost could be much greater.

Shaen Corbet and Constantin Gurdgiev (2019) study 819 cyber security events that occurred between 2005 and 2015. Among these, 230 were severe hacks. They find that severe hacks are punished by significantly reduced abnormal returns. They also find that small data breaches are not punished at the stock exchange at all. Moreover, they state that the stock market volatility is strongly positively correlated with the size of the company and the number of records lost. Another relevant finding is that the frequency of cyber security events has increased over time, especially for hacks.

A study conducted by Yayla and Hu (2011) show that security breaches impact the abnormal return of breached companies. The data sample consists of 130 companies breached between 1994 and 2006. They also find that security events after 2001 had no statistically significant impact on the stock exchange. They suggest that this is an effect of investors being less sensitive to the announcement of a security event. Lastly, the analysis shows that there is a long-term effect on the stock price. Hence, top executives and IT managers should pay attention to cyber security.

Morse, Raval and Wingender Jr. (2011) studies the effect of data breaches on the behaviour of the stock markets using event study methodology. Their sample consists of 306 publicly traded companies that were breached between January 2000 and February 2010, and 34 of the breaches were hacks. In general, they find a negative stock market reaction where the effects are not temporary. The data sample is divided into three based on the key source of the data breach: hacking, fraudulent access, and stolen laptop. They find that hacking attacks do not draw any market effects. However, when analysing data breaches where a stolen laptop or fraudulent access is the key source, they find a negative stock market reaction. It is argued that hacking attacks are beyond the company's control, hence, the company management cannot be blamed for the data breach and the investors will not punish them. However, we find this argument questionable as further discussed in section 7.

A study conducted by Campbell, Gordon, Loeb, and Zhou (2003) use event study methodology to examine the economic cost of publicly announced information security breaches on publicly traded US corporations. The data sample consists of 43 events of security breaches in the period between January 1995 and December 2000. A subsample consisting of 11 events is used to investigate breaches with confidential information such as credit card data. They find limited evidence of a negative market reaction following the announcement of a breach. However, when announced that confidential information is extracted, they find a statistically significant negative market reaction. In conclusion, the findings suggest that investors value the affected firm's differently depending on the confidentiality of the information in the breach.

There are several older research papers that investigate the stock market reaction following the announcement of security breaches using event study methodology. Garg, Curtis and Halper (2003) studies 22 companies between 1999 and 2002, Cavusoglu, Mishra and Raghunathan (2004) use a sample of 66 observations from 1996 till 2001, and Kannan, Rees and Sridhar (2007) study 72 companies breached before 2001. They all conclude that there is a statistically significant negative stock market reaction following the announcement of data breaches.

2.2 Our Contribution to Existing Literature

This thesis contributes to existing literature by focusing on mega hacks that other researchers have avoided in their data sample, because these incidents are regarded as outliers. By studying the worst-case scenario, we believe that we can contribute to decision making in large corporations who run the risk of being hacked. Moreover, we offer a new perspective to the investigation of the determinants of the stock market reaction by exploring the role of consumers and regulatory institutions in inflicting financial consequences on companies that are hacked. To our knowledge, the topic of regulation has not yet been investigated using event study methodology and cross-sectional analysis.

Our data sample consists of recent hacks, which is a strength because of the rapid development in this area. In addition, it allows us to investigate whether the stock market reaction for hacks has changed after the GDPR was put into effect. The data sample also includes companies from all over the world, whereas other studies focus on companies from the US. In addition, the sample consists of hacking events, not data breaches in general, to provide information about the trade-off between investing in cyber security and the cost of being hacked.

Due to limited data availability and the time constraint, we had to make a trade-off between collecting a large data sample and investing time in the analysis. The final data sample consists of 42 events. The fact that our data sample contains 42 events, in contrast to other research which have larger data samples, offers some limitations but also some strengths. We have a clean data sample with carefully investigated event dates and little influence from confounding events.

3 Theory

This section accounts for the market efficiency hypothesis, which is a central assumption in the event study methodology.

Market Efficiency

The market efficiency hypothesis, introduced by Fama (1970), is the hypothesis that "security prices fully reflect all available information", as opposed to the hypothesis that security prices follows a random walk. Fama (1970) defines three strengths of market efficiency that defines subsets of available information that are fully reflected in security prices: weak form, semi-strong form, and strong form. In the weak form efficiency the security prices reflect the past stock prices. The semi-strong form includes the weak form, as well as all obviously publicly available information is reflected in the stock price. While in the strong form efficiency all information, both publicly available and inside information, is reflected in the price.

According to Fama (1991) it is generally accepted that the market is roughly semi-strong. If the market is semi-strong efficient it will quickly and fully reflect new information so an investor cannot use this information to generate extraordinary returns. Under this assumption, positive and financially relevant news about a company should lead to an immediate increase in the company's stock price. Consequently, information that suggests that the company will perform worse than previously expected should lead to a decrease in the company's stock price. The traditional stock valuation models are explained more extensively in Appendix 1.

4 Empirical Methods

In this thesis the event study methodology and cross-sectional regression analysis are implemented as described by MacKinlay (1997). Event studies are used to measure the effect of an economic event on the value of firms. By subtracting the estimated normal return from the actual return of the company following the announcement of a hack, we can approximate the financial consequences of the event. Cross-sectional analysis is used to investigate the link between the abnormal return and certain determinants of the stock market reaction.

In this section the event study methodology is explained. Furthermore, models to estimate normal performance are elaborated on. In the third part, the equations for the computation and aggregation of abnormal return are derived. At last, the cross-sectional test and the cross-sectional regression model are explained.

4.1 Event Study Methodology

There are four underlying assumptions to the event study methodology. The first is that markets are efficient, as elaborated on in section 3. Second, one assumes that the players in the market are rational. The third assumption is that the event is unanticipated, meaning that there must be new and unexpected information revealed at the event date. Fourth, there must be no confounding events, so that the impact on the stock market can be contributed to the event (McWilliams & Siegel, 1997).

When conducting an event study, the initial task is to select the events to analyse based on the objective of the study and the general selection criteria for event studies. Second, the event date of interest must be defined, which is often challenging. For example, to identify the event date one can investigate when newspapers first reported on the event. However, a common challenge is to decide with certainty whether the event is known to the market before it is reported in the news. Hence, to make it less probable to miss the event, the event window is often expanded to permit examination of periods surrounding the events (MacKinlay, 1997).

After choosing the event window, the estimation window must be defined. The estimation window will be used to calculate the normal performance of the stock before the event (MacKinlay, 1997). There is no correct answer when choosing the length of the estimation window. However, the interval should be long enough to minimize the variance of the daily returns and short enough to include only the latest price movements, thus, avoiding changes in systematic risk (Strong, 1992). Typically, the event window and estimation window do not overlap to prevent the normal return model from being impacted by the return in the event period (MacKinlay, 1997).

Figure 4.1: Timeline Event Study

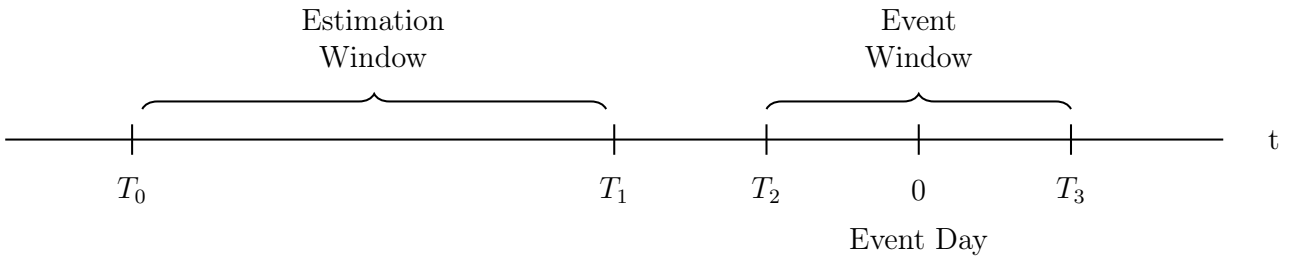


Figure 4.1 is a modification of the event study timeline presented by MacKinlay (1997). T_0 marks the starting point of the estimation window, while T_1 marks the ending point. T_2 marks the beginning of the event window, while T_3 marks the end. The period between T_1 and T_2 is the holdout window which is added to ensure that the estimation window and event window do not overlap. The announcement day is marked with "0".

In this thesis, we use daily stock return. T_0 is equal to -220 days, while T_1 is -20 days. Consequently, the estimation window is equal to 200 trading days. In addition, a holdout window of nine days is chosen, which gives an event window of [-10, 10]. This implies 21 trading days between T_2 and T_3.

The next step of the event study is to estimate the expected return of the security during the event window, conditioned on the event not taking place. Normal performance models used to calculate normal return are discussed in subsection 4.2. Once the expected return is calculated for the event window, the abnormal return is estimated by subtracting the actual ex-post return of the security. Formally, the abnormal return is derived by equation 4.1.

$$AR_{i\tau} = R_{i\tau} + E(R_{i\tau}|X_\tau) \tag{4.1}$$

Where AR is the abnormal return for firm i on event date τ. While $R_{i\tau}$ is the actual return and $E(R_{i\tau}|X_\tau)$ is the expected return for time period τ. The expected return is calculated conditioned on the event not taking place, expressed by the X_τ component.

4.2 Estimating Normal Performance

Normal performance models can be loosely categorised as either statistical or economic. Statistical models, such as the Market Model and the Constant Mean Return Model, are based on statistical assumptions about the behaviour of asset returns. Economic models, such as the Arbitrage Pricing Theory (APT) and the Capital Asset Pricing Model (CAPM), rely on economic assumptions concerning investors' behaviour in addition to the statistical assumptions. However, deviations from the CAPM have been discovered and in the APT model, the most important factor behaves like the market factor. Thus, the gains from using the economic models are relatively small when compared to the Market Model. Consequently, in event studies statistical models dominate (MacKinlay, 1997). Thus, only the statistical models are discussed in more detail below.

The Constant Mean Return Model

The Constant Mean Return Model assumes that the average return for a given security is constant over time (MacKinlay, 1997). Thus, a constant return parameter and a disturbance term is used to define the normal return. Although the Constant Mean Return Model is a simple model, Brown and Warner (1980) find that it often yields similar results as more sophisticated models. Formally, the Constant Mean Return Model is given by equation 4.2.

$$R_{i\tau} = \mu_i + \varepsilon_{i\tau} \tag{4.2}$$

$$\hat{\mu}_i = \frac{1}{L_1} \sum_{\tau=T_0+1}^{T_1} R_{i\tau} \qquad E(\varepsilon_{i\tau} = 0) \qquad var(\varepsilon_{i\tau}) = \sigma^2_{\varepsilon i}$$

In equation 4.2, the predicted normal performance for security i at time τ is represented by $R_{i\tau}$. $\varepsilon_{i\tau}$ is the error term, with an expected value of zero and a variance of $\sigma^2_{\varepsilon i}$. The average return of event i over the estimation period is expressed by $\hat{\mu}_i$. Furthermore, the estimation window is represented by *L1*.

The Market Model

The Market Model is based on the assumptions that there is a stable linear relationship between the return of a market portfolio and the security return. The model's linear specification follows from the assumed joint normality of asset returns (MacKinlay, 1997). In equation 4.3 below, the Market Model is defined for any security i.

$$R_{i\tau} = \alpha_i + \beta_i R_{m\tau} + \varepsilon_{i\tau} \tag{4.3}$$

$$E(\varepsilon_{i\tau} = 0) \qquad\qquad var(\varepsilon_{i\tau}) = \sigma^2_{\varepsilon\, i}$$

From the equation, $R_{i\tau}$ is the predicted normal return for security i at time τ, while $R_{m\tau}$ is the return on the market portfolio at time τ. $\varepsilon_{i\tau}$ is the error term, with an expected value of zero and a variance of $\sigma^2_{\varepsilon\, i}$. The parameters α_i and β_i are estimated by using OLS, based on the observations in the estimation window. A broad-based stock index is used for the market portfolio, such as the S&P500 Index, the CRSP Value Weighted Index or the MSCI World Index.

The Market Model is often preferred over the Constant Mean Return Model (MacKinlay, 1997). The Market Model assumes a linear relation between the stock return and the market return. Thus, by removing the portion of return that is tied to the market's return, the variation of the abnormal return is reduced. Consequently, the possibility of detecting event effects increases (MacKinlay, 1997).

4.3 Computing and Aggregating Abnormal Return

The normal return models that are described above are used to calculate the abnormal return of a security during the event window. The equations presented in this subsection are based on the Market Model. However, the analysis using the Constant Mean Return Model as the normal performance model is virtually identical (MacKinlay, 1997). The equations are used to calculate the thesis results, which will be presented in section 6.

When the parameters α_i and β_i in the Market Model are estimated, the abnormal return of the security during the event window can be predicted by the model, as expressed by equation 4.4. The abnormal return is the disturbance term of the Market Model calculated

on an out of sample basis (MacKinlay, 1997).

$$AR_{i\tau} = R_{i\tau} - (\hat{\alpha}_i + \hat{\beta}_i R_{m\tau}) \tag{4.4}$$

MacKinlay (1997) states that under the null hypothesis (AR = 0), the abnormal return will be jointly normally distributed conditional on the market returns of the event window with a zero conditional mean and conditional variance $\sigma^2(AR_{i\tau})$ where:

$$\sigma^2(AR_{i\tau}) = \hat{\sigma}^2_{\varepsilon i} + \frac{1}{L_1}[1 + \frac{(R_{m\tau} - \hat{\mu}_m)^2}{\hat{\sigma}^2_m}] \tag{4.5}$$

From equation 4.5 we have the same notation as when explaining the Market Model in subsection 4.2.2. The $\hat{\mu}_m$ expresses the estimated average return of the market in the estimation window. When the estimation window, *L1*, becomes large, the second term in equation 4.5 approaches zero, as the sampling error of the parameters α_i and β_i disappears (MacKinlay, 1997). As a result, the variance will be $\hat{\sigma}^2_{\varepsilon i}$.

In order to draw overall inference for the event of interest, the abnormal return observations must be aggregated (MacKinlay, 1997). The aggregation can be conducted both through time and across securities. Aggregating the abnormal return for a security across time, yields the cumulative abnormal return (CAR). Formally, the CAR is derived by equation 4.6.

$$CAR_i(\tau_1, \tau_2) = \sum_{\tau=\tau_1}^{\tau_2} AR_{i\tau} \tag{4.6}$$

The CAR from τ_1 to τ_2 is the sum of the abnormal return for security i (MacKinlay, 1997), as expressed by the equation above.

Furthermore, the CAR for each security can be averaged to find the $\overline{CAR}$ for all securities in the event pool (MacKinlay, 1997). The $\overline{CAR}$ is calculated by aggregating CAR for all firms and divide by the number of events in the sample. For the aggregation it is assumed that the event window of the N events does not overlap, and that there is not any clustering (MacKinlay, 1997). Formally, the $\overline{CAR}$ can be derived by equation 4.7.

$$\overline{CAR}(\tau_1, \tau_2) = \frac{1}{N} \sum_{i=1}^{N} CAR_i(\tau_1, \tau_2) \tag{4.7}$$

The variation of the $\overline{CAR}$ is expressed by equation 4.8, and is used to calculate the $\overline{CAR}$'s significance level.

$$var(\overline{CAR}(\tau_1, \tau_2)) = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2(\tau_1, \tau_2) \tag{4.8}$$

4.4 Cross-sectional Test

In order to examine if hacking events influence the sampled company's stock price, we use a cross-sectional test to investigate whether the $\overline{CAR}$ is significantly different from zero. This is a modified version of the Student's t-test and a parametric test, hence the different security's CAR should be normally distributed (MacKinlay, 1997). The t-statistic is calculated by dividing the $\overline{CAR}$ on its corresponding standard error. The cross-sectional test is derived formally in equation 4.9.

$$t_{\overline{CAR}(\tau_1,\tau_2)} = \frac{\overline{CAR}(\tau_1, \tau_2)}{var(\overline{CAR}(\tau_1, \tau_2))^{\frac{1}{2}}} \sim N(0, 1) \tag{4.9}$$

Because σ_ε^2 is unknown an estimator must be used to calculate the variance of the abnormal return. The estimator is the sample variance measure of $\hat{\sigma}_\varepsilon^2 i$ from the Market Model regression in the estimation window (MacKinlay, 1997). The distributional result is asymptotic with respect to the length of the estimation window and the number of securities. MacKinlay (1997) states that the ARs must be uncorrelated in the cross-section for the estimator of variance to be consistent. For this to hold, there must not be any clustering in the event window of the included securities. In equation 4.9, the variance of $\overline{CAR}$ is derived by equation 4.10.

$$var(\overline{CAR}(\tau_1, \tau_2)) = \frac{1}{N^2} \sum_{i=1}^{N} (CAR(\tau_1, \tau_2) - \overline{CAR}(\tau_1, \tau_2))^2 \tag{4.10}$$

Due to issues with heteroscedasticity in the error terms, robust standard errors are employed in the significance tests. When calculating each security's variance individually the standard errors are robust (MacKinlay, 1997).

4.5 Cross-sectional Regression Analysis

Cross-sectional regression analysis is used to examine the determinants of the stock market reaction (MacKinlay, 1997). The model can be derived by the following equation.

$$CAR_j = \delta_0 + \delta_1 x_{1j} + ... + \delta_M x_{Mj} + \eta_j \tag{4.11}$$

$$E(\eta_j = 0) \qquad var(\eta_j) = \sigma^2_{\eta_j}$$

From equation 4.11, CAR represents the cumulative abnormal return for the j^{th} event observation, while x_{1j} indicates firm specific characteristics 1, for the j^{th} event observation. η_j is the zero mean disturbance term, which is uncorrelated with the $\delta's$. Its expected value is zero, and the variance is $\sigma^2_{\eta_j}$.

According to MacKinlay (1997) interpretation issues can arise when executing the cross-sectional regression. The abnormal return will often be related to firm specific characteristics through both the valuation effects of the event and anticipated effects due to investors forecasting the likelihood of an event. Observed valuation effects may be different form their true value in this case (MacKinlay, 1997).

For simplicity, $\overline{CAR}$ and $\overline{AR}$ [1] will be referred to as CAR and AR in the remainder of the thesis.

[1] $\overline{AR}$ expresses the average abnormal return for all securities in the event pool.

5 Data and Sample Description

The data sample studied in this thesis consists of 42 companies that were hacked between 2007 and 2020. The events are selected from a database based on the general selection criteria for event studies. Information about the number of records lost and the data sensitivity of the leaked information is provided in the database. The data is analysed in R Studio (R Core Team, 2020).

The criteria for selection of the event study pool and the collection of financial data is elaborated in this section. In the final part, we provide descriptive statistics of the data sample.

5.1 Event Data Sample Selection

In this thesis the hacks studied are selected from a database that lists the *World's Biggest Data Breaches & Hacks* (McCandless & Evans, 2020). By using a database, the risk of selection bias is mitigated. The database consists of 354 events of data breaches and hacks that were announced between 2007 and 2020. It was downloaded the 22^{nd} of September 2020. There are three sources quoted for the database: (1) The Identity Theft Resource Center, a US non-profit organization which supports victims of identity theft, (2) DataBreaches.net, a website created by an anonymous individual with special interest in data security, (3) and news articles. The database includes the name of the entity that is hacked, the number of records lost, the year of the hack, the method used by the hackers and the data sensitivity.

To answer the hypotheses of this thesis, the events where the method is described as "Hacked" are selected to the event pool. Hence, all events where poor security, lost device or inside job are listed as the reason for the data breach are filtered out (354 to 214 cases).

Events with apparent confounding events close to the event window are eliminated from the event pool. According to McWilliams and Siegel (1997) it is difficult to isolate the impact of the studied event if other financially relevant events occurred during the event window. Preferably, all events effecting the stock price that are not related to the announcement of the hack should be excluded. However, to exclude price effects from confounding events

manually is comprehensive and almost impossible. Large companies are often written about in the media and consequently it is challenging to identify the news that classify as confounding events. We excluded companies where we find apparent confounding events within the event window, based on our subjective opinion.

The remaining events are screened based on the following criteria:

- The entity is publicly listed on a stock exchange.
- The company has not been acquired, merged, or delisted.
- The event is unanticipated.

When applying the selection criteria as laid out above, we are left with a dataset of 42 events. The complete data sample is provided in Appendix 2.

The Data Sample Used to Analyse the Effect of the GDPR

To analyse the stock market reaction before and after the implementation of the GDPR, naturally the sample must be limited to the companies that are subject to the GDPR. The GDPR applies to all companies that collect, store, transmit or analyse data of citizens in the EU (The European Union, 2020). We investigated the customer base of each company in the data sample and eliminated those that did not fit the criteria. Hence, the data sample for the analysis of GDPR consists of 33 events.

5.2 Data Sources

Daily stock price information is downloaded from Yahoo Finance to calculate the abnormal return of the companies in the event pool. To ensure the quality of the data, the closing price reported at Yahoo Finance was compared to the prices listed on Bloomberg for a selection of companies. We found that the prices were identical. The closing price is adjusted for splits and dividend distributions. The formulas used to adjust the closing prices vary from the different providers and Yahoo Finance uses the Center for Research in Security Prices (CRSP) standards (Yahoo! Finance, 2020).

Additionally, the market value of equity is extracted from Bloomberg to calculate the numerical changes in net value for the companies in the event pool following the

announcement of a hack. Bloomberg reports yearly measures of equity and many of the periods end and start the 31^{st} of March, while for other companies it is stated for the first day of the year. The equity values used in the thesis are from the same twelve-month period as the hack was announced.[2]

In the application of the Market Model, the broad based MSCI World Index (Bloomberg, 2020) is used as the market portfolio. The MSCI World Index is a market capitalization weighted index of 1603 companies across the world. It is chosen to reflect the wide variety of companies in the sample, that are listed on different stock exchanges across the world, such as the Tokyo and Osaka Stock Exchange in Japan, NASDAQ, the New York Stock Exchange, the Hong Kong Stock Exchange, the Australian Securities Exchange and the London Stock Exchange.

The event dates are not reported in the database. Therefore, they were established by investigating when the first articles about the specific hacks were published, or when the companies first announced publicly that they had been hacked. It is crucial that the hack was known for the stock market to react. In this thesis, the event date is the first trading day that investors could possibly trade on information about the hack. For example, if the hack was announced on Friday afternoon before easter, the event date is the first trading day after easter on the stock exchange that the stock was listed. In effect, this thesis only studies the influence of what is assumed to have been revealed in the first announcement. As news reports do not always include all information, subjective judgement was applied.

5.3 Descriptive Statistics

This subsection describes the sample, which includes the market value of equity of the companies in the sample, the number of records lost, the data sensitivity and the timing relative to the GDPR.

Table 5.1 shows the median, mean, minimum value and the maximum value of the marked value of equity and the determinants in our analysis. Additionally, the number of observations and grouping of the different determinants are included.

[2]Ideally, we would use the market value of the equity one week prior to the event, to avoid the effect of the event in the valuation of the equity. However, this information was not available to us.

Table 5.1: Descriptive Statistics

	Equity $M (1)	Records Lost (2)	Records Lost (3)	Data Sensitivity (4)	GDPR (5)
Median	33 847	6 700 000	0.5	1	0
Mean	101 685	37 634 640	0.5	0.69	0.35
Min	237	1 025	0	0	0
Max	904 128	383 000 000	1	1	1
Standard deviation	179 823	68 394 505	0.5	0.47	0.48
Number of observations					
Dummy variable: 0			21	13	22
Dummy variable: 1			21	29	11
Total	42	42	42	42	33

Note: Market value of equity is in million USD. Records lost (3), Data sensitivity (4) and GDPR (5) are dummy variables. Records lost is divided based on the median value, while data sensitivity is grouped based on the data extracted being of a sensitive character (1) or not (0). Additionally, GDPR is divided into groups based on whether the company was hacked before (0) or after (1) the GDPR was put into effect.

From the first column one can observe that the median of the market value of equity is about one third of the mean. This implies that the distribution is skewed. Hence, there are some companies with a high market valuation of equity which increases the average. However, all companies in the sample are large in terms of market value of equity, as all companies have equity above $237 million. This indicates that the sample consists of well-established companies.

The second column shows the number of records lost. One can see that the number of records lost vary from 1025 in the smallest hack (Wendy's restaurant in 2016) to 383 million in the largest hack (Marriott International in 2018). The mean is 37.6 million which is almost six times greater than the median that is 6.7 million. This indicates a skewness in the distribution because some large observations increase the mean.

To investigate the impact of records lost on the stock market reaction, the data sample is divided into two groups based on the number of records lost relative to the median. The statistical properties of the dummy for records lost are presented in column three. Naturally, there are 21 observations in each group. The dummy takes the value 1 if the number of records lost is above 6.7 million, and 0 otherwise.

The fourth column presents the statistical properties of the dummy variable for sensitivity of the data extracted in the hack. Low data sensitivity is defined as online information such as email addresses, information collected in loyalty programs, purchase history and search history. This data is assumed to have little impact on the victim's life when it is extracted in a hack. There are 13 observations in this group. In contrast, high data sensitivity is defined as personal information such as social security number, credit card details and health records. We assume that the loss of such information is more likely to cause negative consequences for the victims of the hack, such as financial loss or identity theft. There are 29 observations in this group.

The last column in table 5.1 shows the dummy variable for timing relative to the GDPR. There are 33 events that are relevant for investigating the impact of the GDPR on the stock market reaction following the announcement of a hack. The 33 events are divided into two groups depending on the timing of the hack relative to the implementation of the GDPR the 25^{th} of May 2018. The dummy takes the value 0 if the hack occurred prior to the implementation of the GDPR. There are 22 events in this group. Also, there are 11 hacks that occurred after the implementation of the GDPR.

Our sample contains hacks in a period ranging from 1^{st} of January 2007 until today. The distribution of hacked companies is shown in figure 5.1.

Figure 5.1: Number of Hacks per Year

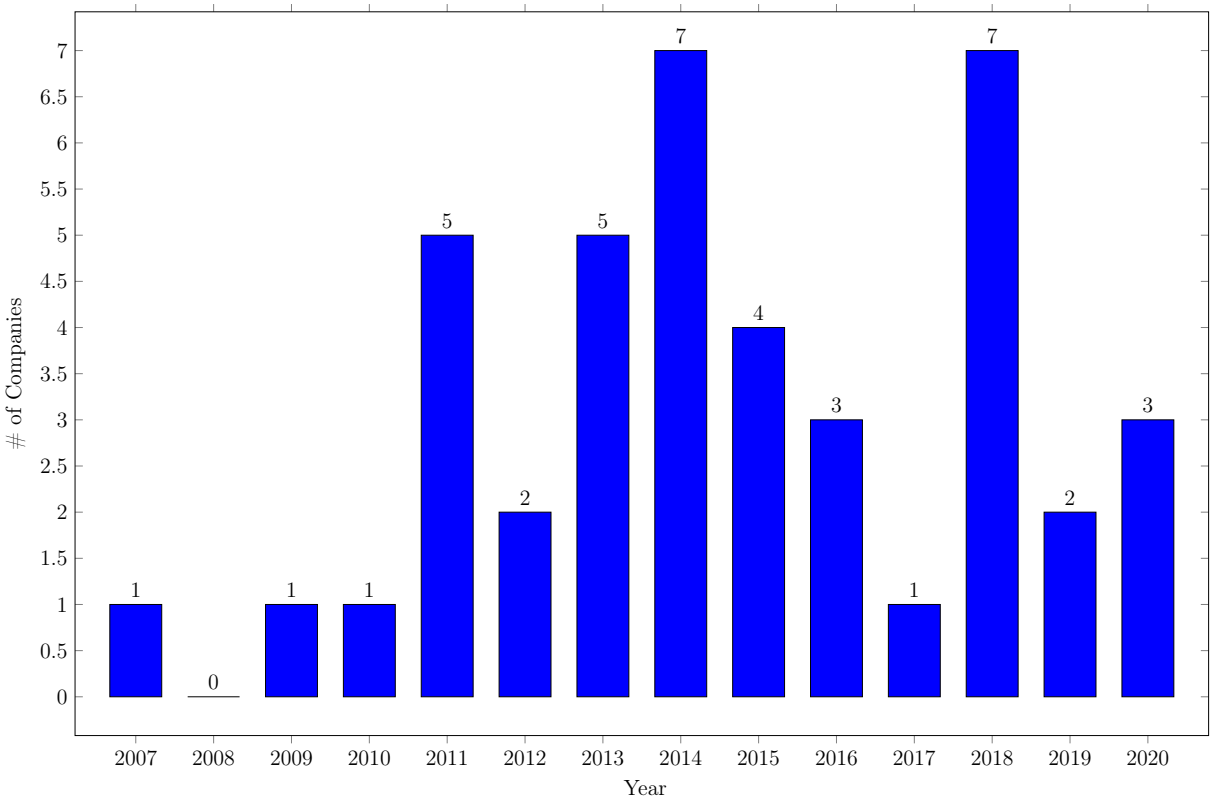


Figure 5.1 shows the distribution of the hacks in the data sample over time. The distribution is uneven, and the number of hacks per year varies from zero to seven. Most of the hacks in our data sample occurred after 2010 and the median year is 2014. The distribution over time in our data sample reflects the general trend that the frequency of hacks is increasing (IBM Security, 2019).

In the cross-sectional regression model, we include a dummy variable for time. The statistical properties of the dummy are equal to those of the dummy variable "Records lost" (3) in table 5.1, because the dummy is derived based on the median date.

6 Analysis

The research questions presented in subsection 1.2 are investigated by using the event study methodology. The AR and CAR[3] are estimated using equations 4.4, 4.6 and 4.7. Different event windows are presented to investigate the longevity of the effects. To reject the null hypotheses the CAR measured must be negative and statistically significant. The hypothesis tests for the CAR are applied as described by MacKinlay (1997). Furthermore, the t-statistic for the difference between CAR follows a Student's t-distribution. Heteroscedastic robust standard errors are employed in the significance tests, as the variance is calculated for each individual company.

In the tables presented in this section, the numbers in the squared brackets indicate the days relative to the event date that have been summarized to calculate the CAR. The significance levels of the statistical tests are indicated with stars. The y-axis of the graphs presented in the analysis are adjusted to fit the data that is presented. Hence, the graphs cannot be compared without taking the scale of the Y-axis into account.

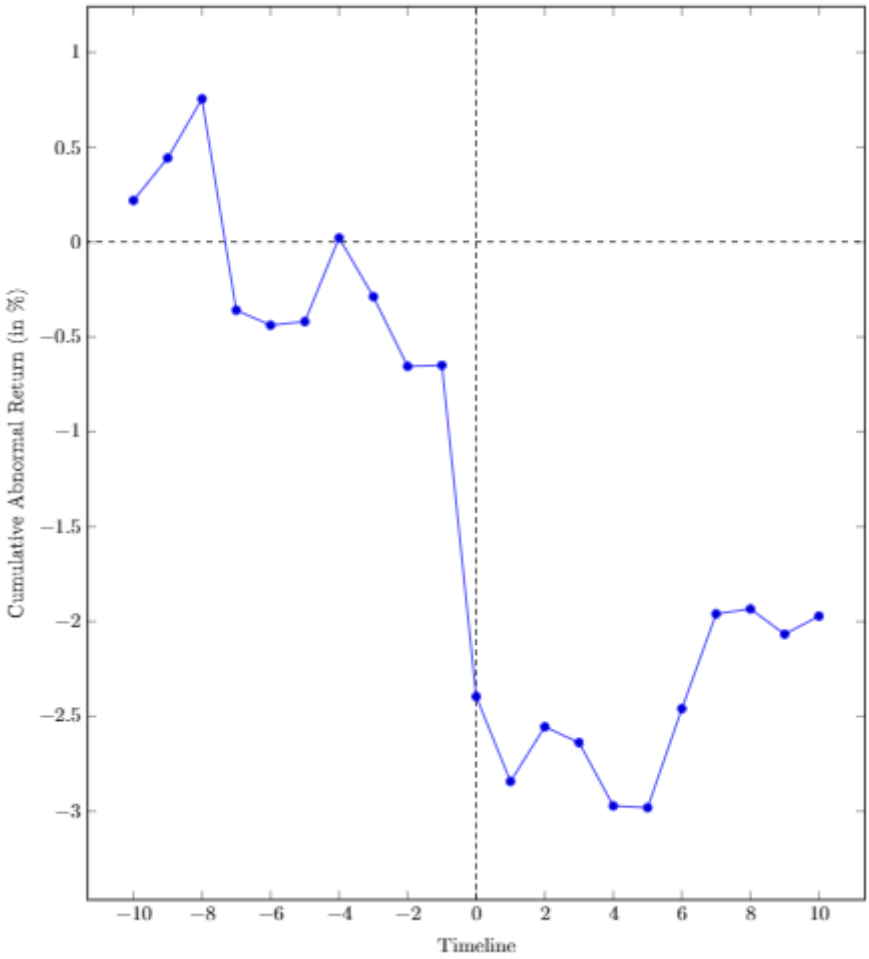
The analysis consists of several parts. First, the CAR is analysed for the 42 events to investigate the hypothesis regarding the stock market reaction following the announcement of a hack. To highlight the practical relevance of the stock market reaction, the CAR is translated into numerical values for each individual company. An illustration of the development in CAR for the individual events is also presented. From the illustration, four outliers are identified and discussed. Second, the events are split into groups depending on the number of records lost and the sensitivity of the data extracted in the hacks. Then the CAR for the two groups is compared to investigate the impact on the stock market reaction. Third, a smaller sample is used to explore the effect of the implementation of the GDPR on the stock market reaction. Finally, a cross-sectional regression analysis is conducted to investigate the combination of the three determinants. The goal is to explore the role consumers and regulatory agencies play in inflicting financial consequences on companies that are hacked and consequently expose private client records.

[3]Average abnormal return and cumulative average abnormal return

6.1 The Stock Market's Reaction to Announcements of Hacks

Figure 6.1 illustrates the development of the CAR for the 42 events, from 10 days before the event date till 10 days after. The CAR is calculated by using the Market Model estimated in relation to the MCSI World Index. The figure shows that on the day of the announcement, the CAR decreases.

Figure 6.1: CAR - Market Model



In the pre-event date window, the CAR fluctuates around zero. However, there is a negative trend four days before the event date. In the post-event date window, the CAR remains low for some days before it starts to recover. However, the figure does not show full recovery within the first ten days after the event date.

Table 6.1 shows the results of the cross-sectional tests for the CAR in the event window for the full sample of 42 events.

Table 6.1: CAR Estimated with the Market Model

Timeline	Market Model
[0]	-0.017***
	(5.60)
[-1, 1]	-0.022**
	(3.51)
[-5, 5]	-0.025**
	(2.46)
[-10, 10]	-0.020
	(1.45)
Post-Event Day Windows	
[1, 5]	-0.006
	(0.79)
[1, 10]	0.004
	(0.31)
Observations:	42

Note: One-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01.

As expected, the AR is different from zero at a 1% significance level at the event day. The table shows that on average the companies in the event pool experienced a negative AR of 1.7% at the event day, which indicates that the market reacts instantaneously to the news.

The most extreme stock market reaction is measured for the time interval of five days prior to, and five days after the event day, where the CAR is negative 2.5%. In the post event windows, the CAR is close to zero and not statistically significant, which indicates that there is no recovery of the stock price. In summary, the average market reaction of the hacks in the data sample is negative and instantaneous, with no statistically significant recovery in the post-event day window.

6.2 The Cost of Being Hacked

To show the practical relevance of the negative stock market reaction, we estimate the numerical change in market value of equity for the individual companies following the announcement of the hack. The cost is calculated by multiplying CAR by the market value of equity the same twelve-month period as the hack occurred. This gives an

approximate valuation of the financial losses of the company following the hack. On average, a time interval of two days seems to capture the stock market reaction most adequately. However, to make more exact estimations, each event should be studied in detail to define the most representative time interval. In the following table the cost is calculated for all the events and sorted by the cost.

Table 6.2: Approximations to the Cost of Being Hacked

Name	Date of the hack	Equity $M	AR [0]	AR [1]	CAR [0, 1]	Cost $M	T-value	P-value
Facebook	19.03.2018	512 793	-0.05	-0.03	-0.08	-40 511	3.27	4 %
Facebook	28.09.2018	561 779	-0.02	-0.02	-0.04	-20 224	0.93	22 %
Microsoft	15.04.2019	904 128	0.00	-0.01	-0.01	-5 425	0.21	43 %
Apple	19.07.2013	360 225	-0.02	0.00	-0.02	-5 403	0.42	36 %
Equifax	08.09.2017	16 545	-0.14	-0.09	-0.23	-3 756	11.63	0 %
Marriott International	30.11.2018	45 154	-0.06	0.03	-0.03	-1 535	1.27	17 %
Cathay Pacific Airways	25.10.2018	43 823	-0.02	-0.02	-0.04	-1 534	0.84	24 %
Home Depot	03.09.2014	107 344	-0.03	0.01	-0.02	-1 503	0.91	23 %
HSBC Turkey	12.11.2014	182 235	-0.01	0.00	-0.01	-1 458	0.49	33 %
Zoom	02.04.2020	21 267	-0.11	0.05	-0.06	-1 276	0.94	22 %
Ebay	21.05.2014	69 989	-0.01	-0.01	-0.02	-1 260	0.69	28 %
Gmail (Oracle)	10.09.2014	184 310	0.00	-0.01	-0.01	-922	0.25	41 %
Target	19.12.2013	40 824	-0.02	0.00	-0.02	-816	1.12	19 %
TalkTalk	23.10.2015	4 545	-0.05	-0.12	-0.17	-786	5.87	1 %
Anthem	05.02.2015	33 693	-0.01	-0.01	-0.02	-741	0.91	23 %
Sony PSN	21.04.2011	51 045	0.01	-0.02	-0.01	-715	0.45	35 %
Dell	29.11.2018	34 000	-0.01	-0.01	-0.02	-680	0.64	29 %
UPS	22.08.2014	93 831	0.00	-0.01	-0.01	-657	0.50	33 %
Honda Canada	27.05.2011	67 997	-0.02	0.01	-0.01	-612	0.35	38 %
Experian	02.10.2015	16 145	-0.04	0.00	-0.04	-597	1.44	14 %
Sony Online Entertainment	27.05.2011	32 278	-0.02	0.01	-0.01	-452	0.46	35 %
Tesco Clubcard	04.05.2020	28 557	-0.01	0.00	-0.01	-343	0.41	36 %
T-Mobile	24.08.2018	50 622	0.00	-0.01	-0.01	-304	0.21	43 %
Sega	20.06.2011	4 394	-0.02	-0.02	-0.04	-163	0.92	23 %
KT Corp.	30.07.2012	8 134	-0.03	0.01	-0.02	-122	0.50	33 %
Nintendo	08.07.2013	13 721	-0.01	0.01	0.00	-55	0.09	47 %
Wendy's	07.07.2016	2 528	-0.01	0.00	-0.01	-35	0.43	36 %
Heartland	21.01.2009	335	-0.03	-0.06	-0.09	-31	1.10	19 %
VTech	27.11.2015	3 583	-0.01	0.00	-0.01	-29	0.34	38 %
Global Payments	04.04.2012	4 055	0.00	-0.01	-0.01	-20	0.14	45 %
UbiSoft	03.07.2013	1 038	0.00	-0.01	-0.01	-7	0.13	45 %
Quest Diagnostics	12.12.2016	11 764	0.01	-0.01	0.00	0	0.00	50 %
Interpark	26.07.2016	237	0.00	0.00	0.00	1	0.09	53 %
Dominios Pizzas (France)	16.06.2014	1 263	0.00	0.02	0.01	16	0.33	62 %
Adobe	03.10.2013	22 958	-0.01	0.01	0.00	23	0.04	51 %
Dixons Carphone	13.06.2018	3 256	-0.03	0.04	0.01	46	0.33	61 %
TD Ameritrade	14.08.2007	11 922	0.01	-0.01	0.01	60	0.14	55 %
AT&T	09.06.2010	152 689	-0.01	0.01	0.00	611	0.22	58 %
Nintendo	09.06.2020	45 989	0.01	0.01	0.02	1 012	0.64	71 %
Toyota	29.05.2019	165 503	0.01	0.00	0.01	1 821	0.59	69 %
JP Morgan Chase	02.10.2014	225 188	0.00	0.02	0.02	4 279	1.03	80 %
Citigroup	09.06.2011	129 093	0.02	0.02	0.04	5 293	1.51	86 %
Mean		101 685	-0.02	0.00	-0.02	-1 876		
Median		33 847	-0.01	0.00	-0.01	-397		
Standard deviation		179 823	0.03	0.03	0.05	7 060		

Note: Numbers in the brackets represent days relative to the event day. Two-tailed t-test.
Cost is calculated as equity multiplied by CAR.Equity and cost is in million US dollars. The events are sorted by cost.

The CAR per company is statistically significant at a 5% significance level for Facebook (March 2018), Equifax and TalkTalk. The median of the estimated cost over the two days is $1 876 bn and the mean cost is $0.397 bn. There are large variations in the estimated cost, as one can see from the standard deviation which is $7 060 bn.

The cost calculations show that a relatively small decrease in stock price might reflect substantial numerical changes to the net value of the company. We find that there are eleven companies in the data sample with an estimated cost above one billion dollars related to the announcement of a successful hacker attack. This highlights the importance of the topic of cyber security to companies that store personal information, and gives a benchmark for the budget for cyber security investments.

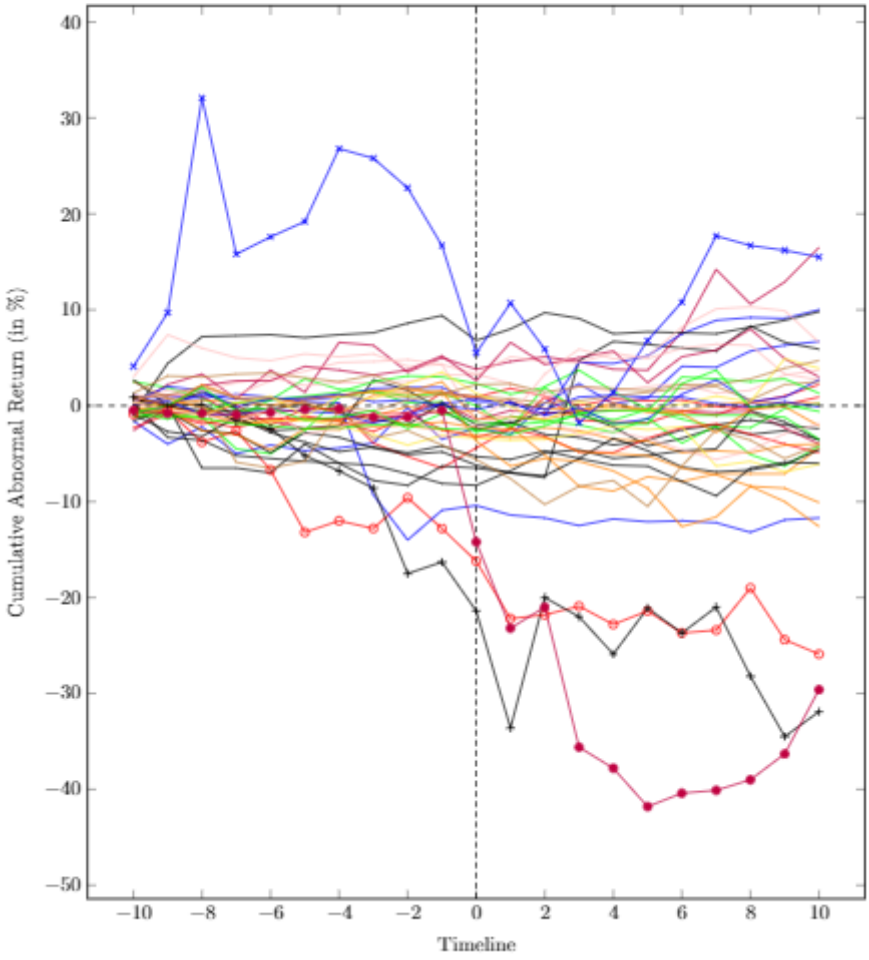
Table 6.2 also shows that the CAR varies across the events. There are ten companies which experienced a positive CAR around the event date. However, this is due to the selected event date. By shifting the time interval of two days one day forward or backward, we find negative CAR for all events in our event pool.

Some of the companies experienced a relatively large decrease in stock price at the day of the announcement of the hack, and the following day. These observations seem to have an unproportional influence on the results of the analysis. The extreme observations are explored further in the following subsection.

6.3 Illustration of the Individual Event Studies

Figure 6.2 illustrates the development of CAR for each of the individual companies in the event pool. From the figure one can see a tendency of negative AR at the event day, when investigating each line separately. Also, most of the lines representing individual companies are below zero around the event date, as expected. The illustration shows that the results of our analysis may be somewhat driven by selected companies. These companies are identified as Zoom (blue line with stars), Heartland (red line with circles), Talktalk (dark green line with stars) and Equifax (purple line with dots). In the robustness analysis in chapter 8, the analysis of the general stock market reaction is repeated without the outliers.

Figure 6.2: CAR - All Events in the Thesis



We provide a table with descriptive statistics of the outliers, so that the reader can be aware of the influence these events have on the result in the further analyses.

Table 6.3: CAR Descriptive Statistics of the Outliers

Name	Event date	Subject to GDPR	Number of Records Lost	Data sensitivity	CAR [-10, 10]
Heartland	21.01.2009	No	130 000 000	High	-25.9 %
TalkTalk	23.10.2015	Yes	157 000	High	-31.9 %
Equifax	08.09.2017	No	143 000 000	High	-29.6 %
Zoom	02.04.2020	Yes	500 000	Low	15.5 %

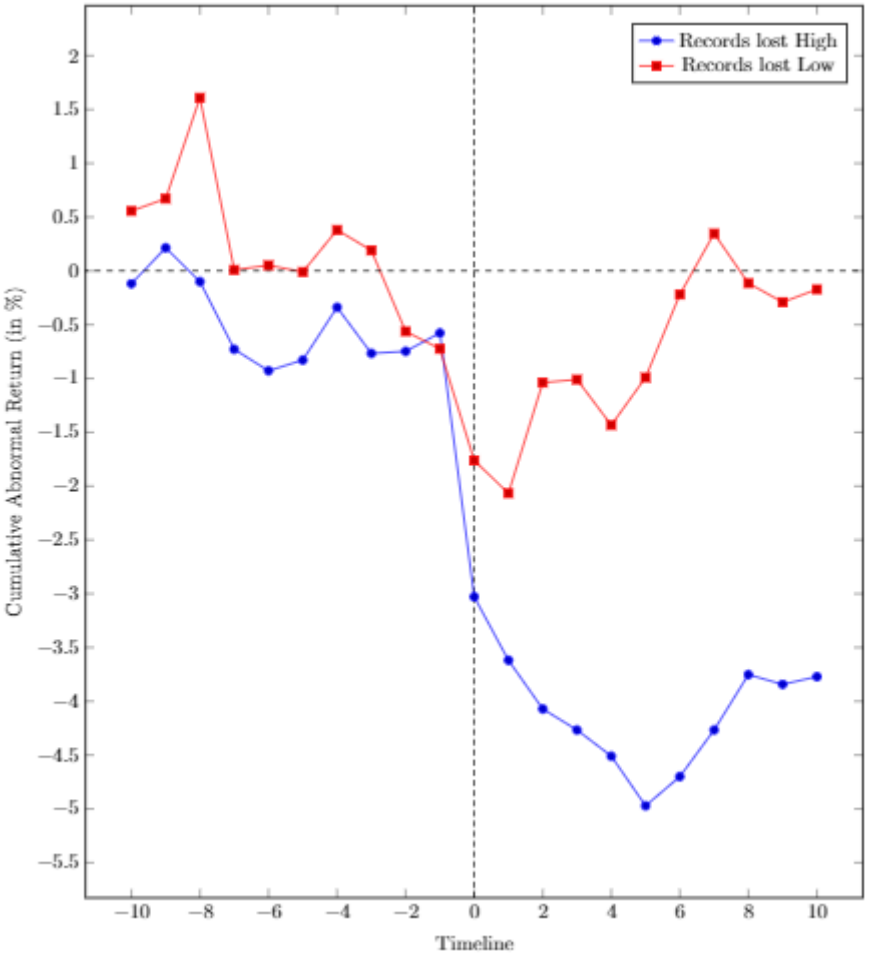
From table 6.3 we find that Heartland, TalkTalk and Equifax have a negative CAR during the event window at approximently 30%. Zoom on the other hand, has a positive CAR of 15% during the event window, even though the measured CAR at the event day is negative. Moreover, Heartland and Equifax will be in the group of "high number of records lost"

and "high data sensitivity". However, these two companies are not subject to the GDPR and will hence not be in the data sample when analysing the effect of the regulation. The hack of Talktalk involved a relatively low number of records lost, but the data sensitivity is regarded as high. Zoom has a low number of records lost and the data sensitivity in the hack is low. Since Talktalk and Zoom have customers in the EU, these companies are subject to the GDPR.

6.4 The Effect of the Amount of Records Lost

Figure 6.3 shows the development of CAR when the event pool is split into two groups depending on the amount of records that are lost in the hack.

Figure 6.3: CAR - Number of Records Lost



Note: The number of records lost is high if more than 6.7 million records were extracted, which is the median of the sample. In contrast, records lost is considered low if less than 6.7 million records were extracted.

From figure 6.3 one can observe that the group of events with relatively many records lost, somewhat follows the expected pattern. At the event day there is a sharp decline in AR. Moreover, prior to the event date, CAR fluctuates around zero and somewhat below. After the event date, CAR continues to decrease, and does not recover within the ten first trading days after the event date. The second group does not follow the expected pattern. There is not a sharp decline in CAR on the event date, as the decrease starts four days prior to the event. One can also observe that the CAR starts to recover the second day after the event date, and almost recovers completely within ten days.

To further analyse the trends that one can observe in figure 6.3, we present significance tests of CAR for different time intervals. In table 6.4 the group of events with a relatively high number of records lost is compared to the group with a lower number of records lost. In the third column the results of a two-tailed Student's t-test of the difference between the groups are presented.

Table 6.4: CAR for the Number of Records Lost

Timeline	Records lost high (1)	Records lost low (2)	Difference (3)
[0]	-0.025***	-0.010**	-0.014
	(5.59)	(2.34)	(1.56)
[-1, 1]	-0.029***	-0.015	-0.014
	(3.27)	(1.70)	(0.87)
[-5, 5]	-0.040**	-0.010	-0.030
	(2.78)	(0.71)	(1.18)
[-10, 10]	-0.038*	-0.002	-0.036
	(1.88)	(0.08)	(1.19)
Post-Event Day Windows			
[1, 5]	-0.019*	0.008	-0.027
	(1.79)	(0.72)	(1.67)
[1, 10]	-0.007	0.016	-0.023
	(0.37)	(0.78)	(1.28)
Observations:	21	21	42

Note: Numbers in the brackets represent days relative to the event day. T-statistic in parentheses.
Two-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01.High number of records is defined as more than 6.7m.

Table 6.4 shows that the companies in the sample that lost a high amount of records in the hack, experienced a statistically significant AR of negative 2.5% on average. When expanding the event window symmetrically around the event day, the measured CAR

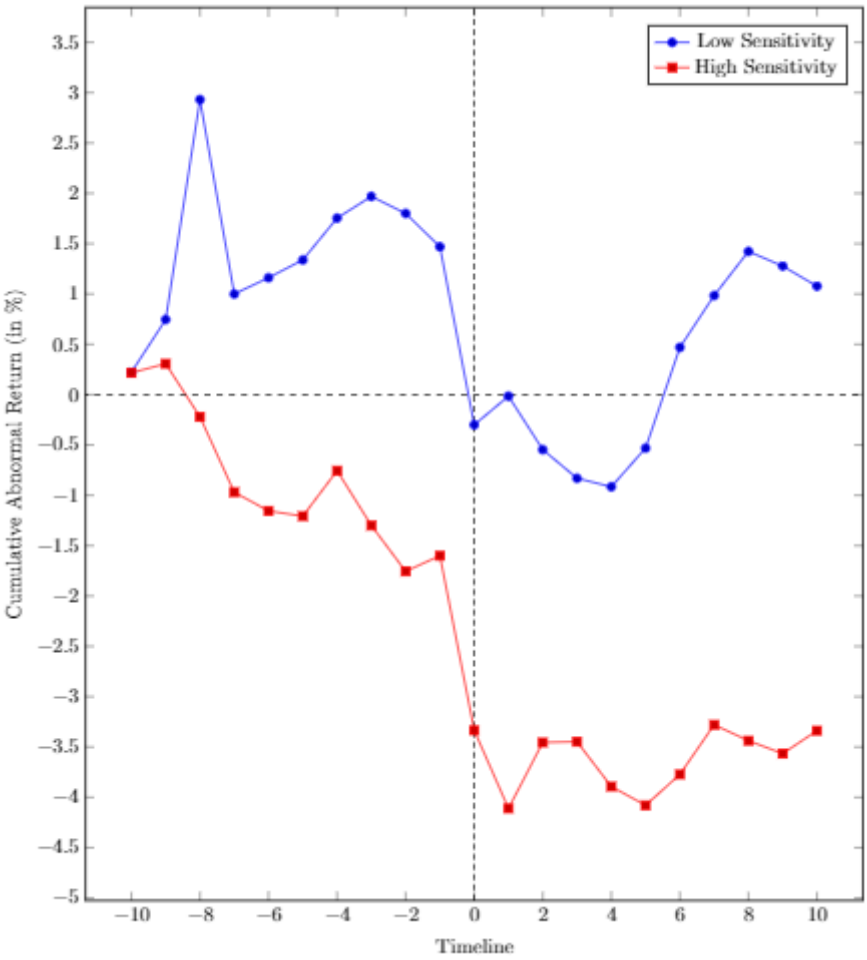
increase. At the most extreme, the measured CAR is negative 4%. On the other hand, the group of fewer records lost, only shows a significantly negative AR at the event date and the decline is at negative 1% on average, and this result is significant at a 5% significance level.

The most important finding in this analysis is that the difference between the groups is close to significant for the measured AR on the event day and in the post event day window. This indicates that there is a difference between the groups, and that the stock market reaction following the announcement of a hack is impacted by the number of records lost.

6.5 The Effect of Data Sensitivity

From the data sample, we have derived two groups based on the data sensitivity of the hacks. Figure 6.4 illustrates the development of CAR for the hacks with sensitive data, compared to the group with less sensitive data. At first sight, it seems that there is a large difference between the two groups. However, when investigating the cross-sectional tests and the Student's t-tests in table 6.5, a different conclusion is drawn.

Figure 6.4: CAR - Data of Low and High Sensitivity



Note: Low sensitivity: online information such as email addresses, information collected in loyalty programs, purchase history and search history. High sensitivity: personal information such as social security number, credit card details and health records.

Figure 6.4 shows negative AR for the events in the sample of low and high sensitivity at the event day. While the CAR for hacks of low sensitivity is positive across almost the entire event window, one can see that it is mostly negative for the group with high data sensitivity. The pattern of the group with low sensitivity is similar to the pattern of Zoom from the graph in subsection 6.3, where the individual events are illustrated. Since it seems that Zoom affects the results, a graph without Zoom is provided in Appendix 3. The CAR for the group with high sensitivity is mostly decreasing during the event window, with a steep decrease at the event date. Additionally, one can observe that the CAR for sensitive data does not recover from the effect of the hack during the ten days after the event date.

In table 6.5 the results of the hypothesis tests are presented. The third column shows the results from a two tailed t-test of the difference between the CAR for hacks with high and low data sensitivity.

Table 6.5: CAR for Data Sensitivity

Timeline	Low data sensitivity (1)	High data sensitivity (2)	Difference (3)
[0]	-0.018***	-0.017**	0.000
	(4.76)	(3.04)	(0.03)
[-1, 1]	-0.018**	-0.024**	0.005
	(2.44)	(2.06)	(0.36)
[-5, 5]	-0.017	-0.029	0.012
	(1.37)	(1.54)	(0.56)
[-10, 10]	0.011	-0.033	0.044
	(0.63)	(1.28)	(1.59)
Post-Event Day Windows			
[1, 5]	-0.002	-0.007	0.005
	(0.09)	(0.44)	(0.35)
[1, 10]	0.014	0.000	0.014
	(1.12)	(0.00)	(0.72)
Observations:	13	29	42

Note: Numbers in the brackets represent days relative to the event day. T-statistic in parentheses. Two-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01.

From table 6.5 one can see that the difference in CAR between the two groups is not statistically significant for any time interval. This indicates that there is no difference in stock market reaction at the event date depending on the sensitivity of the data in the hack.

The AR for data extracted with high sensitivity is statistically significant at the event day, and when expanding the interval with two days. The same result is measured for the group with low data sensitivity. When examining the five days after the day of the event, both sensitive and less sensitive data have negative CAR. These results are not significant, indicating that the abnormal return is zero. Hence, there is no recovery. The difference between the two groups is close to significant for the entire event window of 21

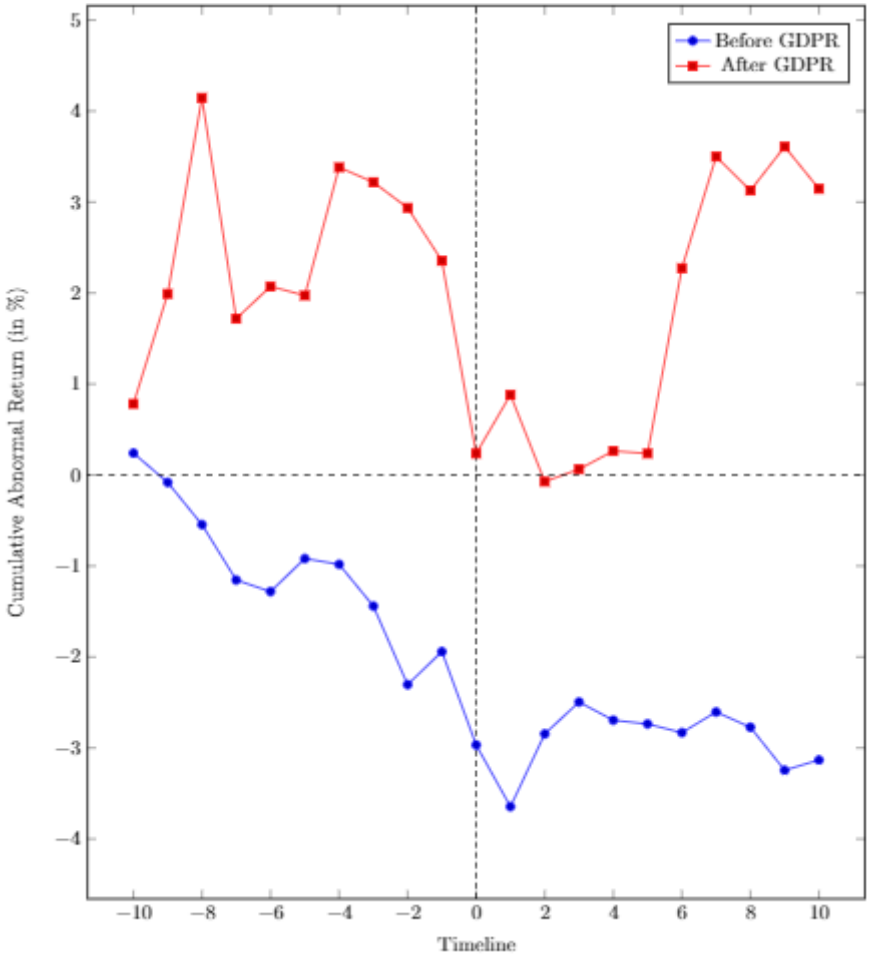
days. However, this is probably due to events in the pre-event window, which do not tell us anything about the reaction to the announcement of hacks.

In summary, there is no statistically significant differences between the groups in the time intervals close to the event date. Hence, the first impression from looking at the graph is misleading, and we find no support for the hypothesis that the sensitivity of the data in the hack has an impact on the stock market reaction.

6.6 The Effect of the GDPR

Figure 6.5 shows the difference in CAR for companies hacked prior to and after the GDPR was put into effect. Note that the sample contains the 33 events that are subject to the GDPR. At first sight, it seems that there is a large difference between the two groups. However, when investigating the hypothesis test in table 6.6 around the event date, there are no significant differences.

Figure 6.5: CAR - Before and After the Implementation of the GDPR



Note: The 33 events that are subject to sanctions are divided depending on the hack happing before or after the GDPR was put into effect the 25^{th} of May 2018.

From figure 6.5 one can observe that the decline in AR on the event day differs between the groups. For the companies which announced that they had been hacked after the GDPR was put into effect the decline is relatively stronger. Nevertheless, one can observe that the CAR for these companies recovers shortly after the event date. Also, note that Zoom, which is identified as an outlier, is in this group. Zoom had a strongly positive CAR prior to the event, which can explain some of the volatility prior to the event date. A graph without Zoom is provided in Appendix 3. The companies that were hacked before the introduction of GDPR, experienced a gradually decreasing CAR over the event window. The outlier TalkTalk is in this group.

Table 6.6 shows how CAR differs for hacks that happened before and after the GDPR was put into effect for different time intervals. The third column in the table shows the difference in the CAR between the two groups, and the results from the t-tests.

Table 6.6: CAR for Time Relative to the Implementation of GDPR

Timeline	Before GDPR (1)	After GDPR (2)	Difference (3)
[0]	-0.010**	-0.021***	0.011
	(2.43)	(3.34)	(0.86)
[-1, 1]	-0.013	-0.021	0.007
	(1.58)	(1.62)	(0.07)
[-5, 5]	-0.015	-0.018	0.004
	(1.04)	(0.87)	(0.60)
[-10, 10]	-0.031	0.031	-0.063**
	(1.61)	(1.08)	(2.56)
Post-Event Day Windows			
[1, 5]	0.002	0.000	0.002
	(0.22)	(0.00)	(0.19)
[1, 10]	-0.002	0.029	-0.031*
	(0.08)	(2.07)	(1.85)
Observations:	22	11	33

Note: Numbers in the brackets represent days relative to the event day. T-statistic in parentheses.
Two-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01. GDPR was implemented the 25^{th} of May 2018.

The table shows that the difference between the stock market reaction for the two groups is not statistically significant in the time intervals close to the event date. For the group with hacks that ocurred after the GDPR was implemented, we find a significant average decrease of 2.1% in AR. Similarly, AR is significantly negative for the companies hacked before GDPR with a decrease of 1% at the event day. Furthermore, when expanding the event window the CAR is still negative, but not statistically significant.

There are two time intervals where the difference between the two groups is statistically significant. The first is when investigating the entire event window of 21 days. However, this can probably be attributed to the abnormal return prior to the event for volatile companies such as Zoom, which is in the group of companies that were hacked after the GDPR was implemented. The second time interval with statistically significant differences between the two groups is the ten days after the event date.

In summary, figure 6.5 gives the impression that there are large differences between the groups. However, when studying table 6.6, one can see that for the time intervals close to the event date and at the event date, there is no statistically significant differences. Hence, we find no support for the hypothesis that GDPR has impacted the stock market reaction.

6.7 Cross-Sectional Regression Analysis of Hack Announcements

As a final stage of our analysis, we provide a cross-sectional regression analysis of hack announcements, with AR and CAR as the dependent variable and number of records lost, data sensitivity and time as the explanatory variables. The explanatory variables are coded as dummy variables and the dummy for GDPR is replaced by a time dummy. This is because we suspect that the analysis of GDPR captures the effect of time, not the effect of a new regulation. This hypothesis is discussed in section 7. Thus, to create the dummy variable for time, the data sample is divided into two groups with 21 events in each group. Since October 2014 is the median date of the data sample, the time dummy takes the value 0 if the hack occurred before October 2014 and 1 otherwise.

The assumptions for OLS are investigated using plots and formal tests. The residual errors are not normally distributed, which impacts the t-statistics. However, this is due to the outliers in the data sample. When Heartland, Zoom, TalkTalk and Equifax are excluded, the OLS assumptions hold. The cross-sectional regression analysis without outliers is presented in the robustness analysis in subsection 8.3.

In table 6.7 the results of the regression analysis is presented using the Stargazer package (Hlavac, 2018).

Table 6.7: Cross-Sectional Analysis of Hack Announcements

	CAR			
	[0]	[0, 2]	[0, 5]	[0, 10]
Number of Records Lost	-0.016*	-0.035**	-0.044*	-0.043
	(1.69)	(2.64)	(1.89)	(1.62)
Data Sensitivity	0.006	0.014	0.011	-0.004
	(0.62)	(0.94)	(0.42)	(0.12)
Time	-0.016*	-0.025*	-0.036	-0.025
	(1.81)	(1.90)	(1.61)	(0.98)
Intercept	-0.005	0.002	0.010	0.025
	(0.57)	(0.14)	(0.43)	(0.92)
Observations	42	42	42	42
Degrees of Freedom	38	38	38	38
R^2	0.137	0.214	0.137	0.093
Adjusted R^2	0.068	0.152	0.069	0.022

Note: *p<0.1; **p<0.05; ***p<0.01. T-stat in parenthesis. The numbers in the brackets indicate the number of days beyond the event date. All explanatory variables are dummies. Number of records lost: 0 when the number of records lost is below the median, 1 otherwise. Data Sensitivity: 0 if the data sensitivity is low, 1 otherwise. Time: 0 if the hack occurred before October 2014, 1 otherwise.

The variables in the regression analysis account for some of the variation in CAR. Moreover, the results are consistent with the results in the analysis from subsection 6.1-6.6. The number of records lost have a statistically significant effect on the CAR of the companies that are hacked, but the effect is not significant when expanding the time interval to including day six till ten. Furthermore, data sensitivity does not show a statistically significant effect on CAR.

There is new insight from the time variable in the analysis, as it has a statistically significant effect on the CAR. This result indicates that time may be the effect that is captured in the analysis of the GDPR. However, the time effect diminishes over the longer time intervals. It shows that for our data sample, hacks announced after October 2014 have a significant impact on the negative stock market reaction.

6.8 Summary of the Analysis

To summarize the results from the analysis we provide a table with the three hypothesis of the thesis and some concluding remarks. We show the results for two time intervals and our conclusions based on the results from the analysis. "Supported" meaning statistically significant negative CAR. "Weakly supported" meaning indications of difference between the groups that are tested, but no significant finding. "Not supported" meaning no indication of differences. In addition, we provide a short comment for each hypothesis.

Table 6.8: Summary of the Results in the Analysis

Hypotheses	Event Windows [0, 5]	Event Windows [0,10]	Comments
H1: A firm-specific hack will influence the stock value negatively upon the announcement of the hack to the public.	Supported	Supported	The CAR is significantly negative and only partly recovers over the post event window.
H2: A firm-specific hack will have a stronger negative impact on the stock value at the announcement of the hack if it is expected to have great impact on the customers of the firm. The influence of the customers is measured as:			
H2a: The number of records lost	Supported	Weakly supported	When considering the cross-sectional regression analysis we find support for the hypothesis. In the robustness analysis we find support for the hypothesis in the time interval of 11 days.
H2b: the data sensitivity of the hack	Not supported	Not supported	We find no statisticaly significant difference between hacks with high and low data sensitivity
H3: A firm-specific hack will have a stronger negative impact on the stock value at the announcement of the hack if the firm is subject to regulations that empower authorities to sanction the specific firm for not safeguarding the personal data.	Not supported	Not supported	The group with hacks that were announced after GDPR, recovers significantly slower than hacks before GDPR was implemented. Hence, the hypothesis is not supported as the effect is opposite of what was expected. In the cross-sectional regression analysis we find a statistically significant short term effect on time.

7 Discussion

In this section we discuss the findings in the analysis and compare the results to previous research on the topic. First, we discuss the CAR following the announcement of hacks. Second, we explore the hypothesis that the consequences to the customers impact the stock market reaction. Finally, we discuss the hypothesis regarding the implementation of regulations.

7.1 The Stock Market Reaction Following Hacking Announcements

We find support for the hypothesis that there is a negative stock market reaction following the announcement of a hack. In general, the negative stock market reaction that is measured in this thesis can be explained by the expectation of increased future costs and revenue losses when a company has been hacked. Under the assumption that markets are efficient, investors price the expected future tangible and intangible costs into the stock price instantaneously.

Our result is consistent with previous research on cyber security breaches in general, as described in the literature review. When considering cyber security breaches in the form of hacking, research is inconclusive. Morse et. al. (2011) study 34 hacks that occurred between 2000 and 2010 and find that hacking, as opposed to other types of breaches, do not draw any market effects. They argue that this is because attacks by hackers can be perceived as out of the company's control. Any company can be hacked independent of the data security measures in place. However, one can argue that the fault of the company is not relevant for the net value of the company. There are still tangible and intangible costs related to the hack that impact the net value of the company, which in turn decrease the dividend payout and the stock price.

On the other hand, the findings of Corbet and Gurdgiev (2019) is consistent with our result. For companies exposed to cybercrime in the form of hacking, they find significantly large volatility effects. It is argued that the financial markets are becoming more aware of the consequences for companies of hacking attacks and are hence punishing the companies

that have been breached accordingly.

The market value of equity is expected to be somewhat volatile, thus there must be a long-lasting effect for management to be influenced by the stock price. Hence, the stock market reaction the days following the announcement must be further investigated to provide decision relevant information to IT managers and top executives (Yayla & Hu, 2011).

7.2 The Longevity of the Negative Stock Market Reaction

Our analysis shows that the stock does not recover the first five days after the announcement of a hack. However, from day five till ten after the announcement, there is some recovery of the stock prices on average. Nevertheless, the recovery is not statistically significant, which indicates that the stock market reaction is not transitory. The findings are consistent with the research of Corbet and Gurdgiev (2019) who study a prolonged event window including 30 days after the announcement of the hacks. Their result show that the stock price starts to recover 20 days after the event date, but that it is not fully recovered at day 30.

According to the market efficiency theory, for the abnormal return to be positive, there should be new financially relevant information that increases the net value of the company. New information may be related to the hack, such as information that the attack has been stopped. However, signs of recovery could also be related to confounding events. This makes it hard to tell whether the recovery implies that the effect of the hack on the stock price is short-term or long-term. Since the recovery measured in our data sample is not statistically significant, there are strong indications that the effect of the hack is long-term. In conclusion, our findings point towards that the shareholder value is at stake. Thus, it suggests that IT managers and top executives should pay attention to preventing hacking attacks.

7.3 The Customers Role in Affecting the Stock Market Reaction

We find some support for the second hypothesis that investigates the stock market reaction in relation to the consequences for the customers that are affected by the hack. The

cross-sectional regression analysis shows that the number of records lost has statistically significant negative impact on the stock market reaction. In terms of data sensitivity, there are no indications of differences between the groups.

A possible reason why the stock market reaction is stronger when many records are lost is that more people are affected by the hack, thus the reputational damage is possibly greater. Also, there are more victims who can take legal action, which is likely to cause increased legal costs. The result is in line with the findings of Corbet and Gurdgiev (2019) who states that the volatility in the stock market is highly positively correlated with the number of client's records stolen during a hack and the size of the company. Naturally, the number of records lost, and the size of the company are positively correlated, hence part of the effect that we observe can possibly be attributed to the size of the company.

For data sensitivity, we hypothesised that the loss of data of high sensitivity led to relatively higher costs to the company because of the severe consequences for the customers. Examples of such costs are legal fees, more forensic investigations, damage of the brand, and following higher investments in improved cyber security. Our results do not support this hypothesis: we find that the groups draw equal market effects. However, our result is not consistent with previous literature. According to Campbell et al. (2003) investors have a significantly negative reaction to security breaches when confidential information is extracted. IBM Security (2019) investigated the differences in cost related to data breaches for different sectors and found that the health sector, which holds highly sensitive information, had higher costs than other sectors. To our knowledge, there is no previous research on data sensitivity and security breaches in the form of hacks.

The deviation from previous research might be related to two different aspects of our research. First, we have a small data sample which makes it more difficult to find statistical significance. Second, our definition of high and low sensitivity of data: Online information is defined as low sensitivity, however online data can be sensitive, such as contact information. It is also possible that the damage caused by the breach is not related to the sensitivity of the information. Data that is of low sensitivity can be abused to achieve great damage. For instance, email addresses and passwords can be used to access more sensitive information, since many people use the same email addresses and password for several accounts.

The number of records lost, and the sensitivity of the data could be related, which also might offer some explanation for our findings. Companies that store many records about customers often collects information of low sensitivity, such as e-mail. While companies that collect data of high sensitivity often hold fewer records, such as health institutions. This argument points toward that the number of records lost is a stronger driver of stock market reaction than the sensitivity of the data in this sample.

In summary, the management of corporations that store personal data seems to have converging interests with their customers. When the company stores many records, our results point towards larger and more prolonged damage of the company which is reflected in the stock price. In conclusion, IT managers and top executives should invest more in cyber security if they store large amounts of customer data.

7.4 The Stock Market Reaction in Relation to the Implementation of the GDPR

We hypothesised that the companies that were hacked after the GDPR was put into effect would experience a stronger market reaction than hacks before the GDPR. This because the potential sanctions increase the expected costs to the companies that are hacked. The maximum fine is €20 million or 4% of global revenue, whichever is higher. Our analysis is inconclusive on the effect the GDPR has on our event pool, probably due to the sample size. There are only 11 companies in the data sample that were announced hacked after the 25^{th} of Mai 2018, when the law was implemented. This increases the risk of conducting a type 2 error.

The market reaction on the day of the event is stronger after the GDPR was implemented, but the effect is not significantly different from the other group. Even so, the result points towards stronger market reactions after the GDPR. This may be explained by investors pricing the potential fine into the valuation of the hacked company immediately.

In terms of recovery there is a statistically significant difference between the two groups, however the effect is opposite of what was expected. For companies that were hacked after the GDPR was implemented, the stock price is fully recovered after ten days. The companies that were hacked before the implementation of the GDPR does not recover

within ten days and the difference is significant at a 10% significance level. However, it is possible that the determinant that is measured is not the implementation of the GDPR, but other factors such as time.

To further investigate if the observed effect is time, a time dummy is included in the cross-sectional regression analysis. We find that more recent hacks draw a stronger market reactions than earlier hacks. The recent study of Corbet and Gurdgiev (2019) of the development of stock market reactions over time is consistent with our findings. A potential explanation of the development is that more information is stored online than before, and thus, more damage can be done when a company is hacked. Also, one might speculate that criminal hackers have become more talented and sophisticated and thus are able to extract more information than before. Another possible explanation is that consumers have become more aware of the potential consequences to hacks and cyber security breaches in general, and thus use their market power to punish companies that are hacked.

On the other hand, Yayla and Hu (2011) studied hacks between 1994 and 2006 and find that there is no negative stock market reaction in more recent years. This indicates that the stock market reaction has been diminishing over time, which is the opposite of what we find. They provide two possible explanations for their observation. First, leakages of information about the hack before the public announcement. The use of Internet is increasing, which facilitates leakage. When there are leakages of information the stock market reaction is spread over several days, and it is harder to measure the effect due to confounding events. We recognize this as a potential issue for some of the events in our data sample. For example, the hack of Zoom was first discovered when personal data from hacks were posted for sale on the dark web. Other hacks were carried out years before the company publicly admitted that they had been hacked. Thus, there is a risk of leakages. Second, Yayla and Hu (2011) suggest that investors have become less sensitive to the announcement of security events because there are more frequently news stories about security breaches today than before.

In conclusion, there are reasonable explanations for both increasing and a decreasing stock market reaction over time. When investigating the data sample based on the timing relative to the GDPR, our findings are consistent with those of Yayla and Hu (2011).

However, when investigating time in general, our findings are in line with those of Corbet and Gurgiev (2019). Due to our small sample size, it is not possible to conclude. There are no previous studies on the effect of GDPR that we are aware of, presumably because it was recently put into effect.

7.5 Summary of the Discussion

In summary, our findings regarding the negative stock market reaction are consistent with previous research. This indicates that IT managers and top executives have incentives to invest in cyber security to prevent successful hacking attacks in general. Furthermore, the number of records lost seem to impact the stock market reaction. However, we do not find any evidence of impact from the data sensitivity. This is not consistent with previous literature and can potentially be explained by our definitions of low and high data sensitivity. Thus, the conclusion of the second hypothesis is that the incentives of managers to protect personal data is somewhat intertwined with the interest of their customers. Finally, we do not find support for the hypothesis that the GDPR has had an impact on the stock market reaction following the announcement of a hack. However, this is probably due to the small sample and that there are other determinants of the hack. In the cross-sectional regression analysis, we investigate the effect of time in general and find that there is an increase in negative stock market reaction in our data sample. The result regarding time is consistent with previous research.

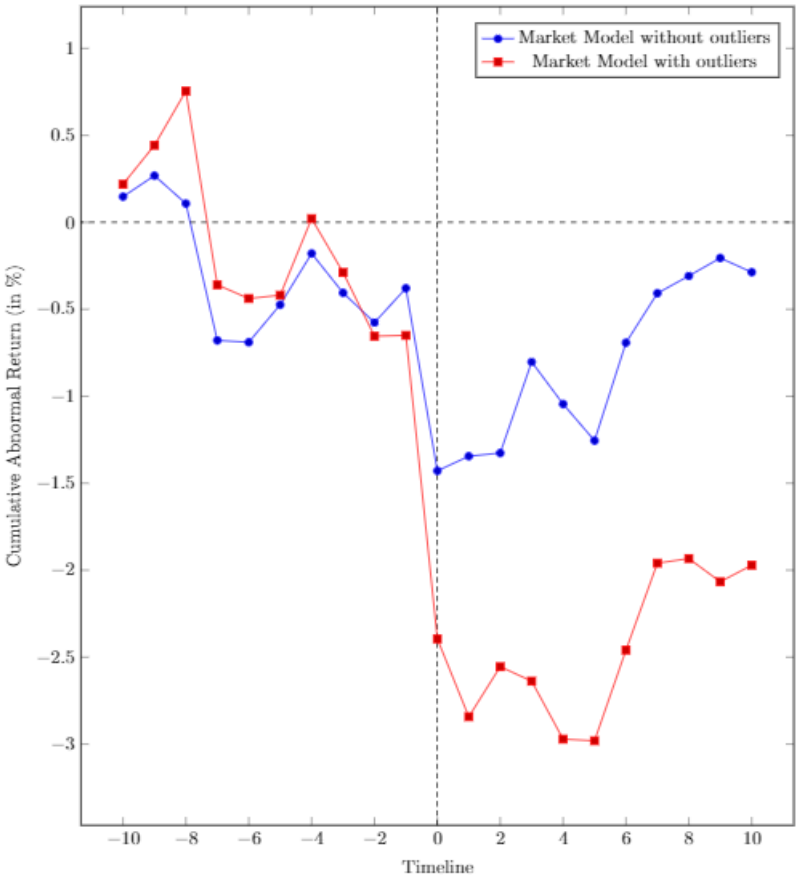
8 Robustness Analysis

In this section we test the impact of some of the research design choices that are made, to show how these choices influence the result. First, we test the impact of the outliers on our main hypothesis. Second, we test the impact of the choice of normal performance model and investigate the differences in estimated abnormal return between the Market Model and the Constant Mean Return Model. At last, the cross-sectional regression analysis is provided without outliers.

8.1 Omitting Outliers

The analysis of the main hypothesis is revisited because there are four outliers that might influence the result unproportionally. The outliers identified in subsection 6.3 are the hacks of Equifax, Zoom, Talktalk and Heartland. In figure 8.1 the development of the CAR for the Market Model with and without the outliers are illustrated.

Figure 8.1: CAR - Market Model with and without Outliers



The illustration shows that the outliers impact the AR at the event day. For the group with outliers there is a steeper decrease in the stock price on average when compared to the group without outliers. In terms of recovery, the group without outliers shows signs of strong recovery already four days after the announcement, and the negative stock market reaction is fully recovered at day eight.

Statistical tests of AR and CAR for different time intervals are provided in the table below to give further insight into the impact of the outliers. The CAR for the time interval of one till ten days after the event date is 1.1%, which is equal to the AR at the event date. However, the recovery is not statistically significant. Hence, there are indications that the negative market reaction is more transitory, but this cannot be confirmed by the statistical tests.

Table 8.1: CAR with and without Outliers

Timeline	With outliers (1)	Without outliers (2)	Difference (3)
[0]	-0.017***	-0.011***	-0.007
	(5.60)	(3.77)	(1.30)
[-1, 1]	-0.022***	-0.008*	-0.014*
	(3.51)	(1.64)	(1.64)
[-5, 5]	-0.025**	-0.006	-0.020*
	(2.46)	(0.61)	(1.36)
[-10, 10]	-0.020	-0.003	-0.017
	(1.45)	(0.22)	(0.93)
Post-Event Day Windows			
[1, 5]	-0.006	0.002	-0.008
	(0.77)	(0.25)	(0.77)
[1, 10]	0.004	0.011	-0.007
	(0.41)	(1.24)	(0.60)
Observations:	42	38	

Note: Numbers in the brackets represent days relative to the event day. T-statistic in parentheses. One-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01.

The results presented in table 8.1 show that the four companies with the largest stock market reaction influence the results. However, the overall conclusion does not change: There is a negative stock market reaction on average at the day of the announcement, and

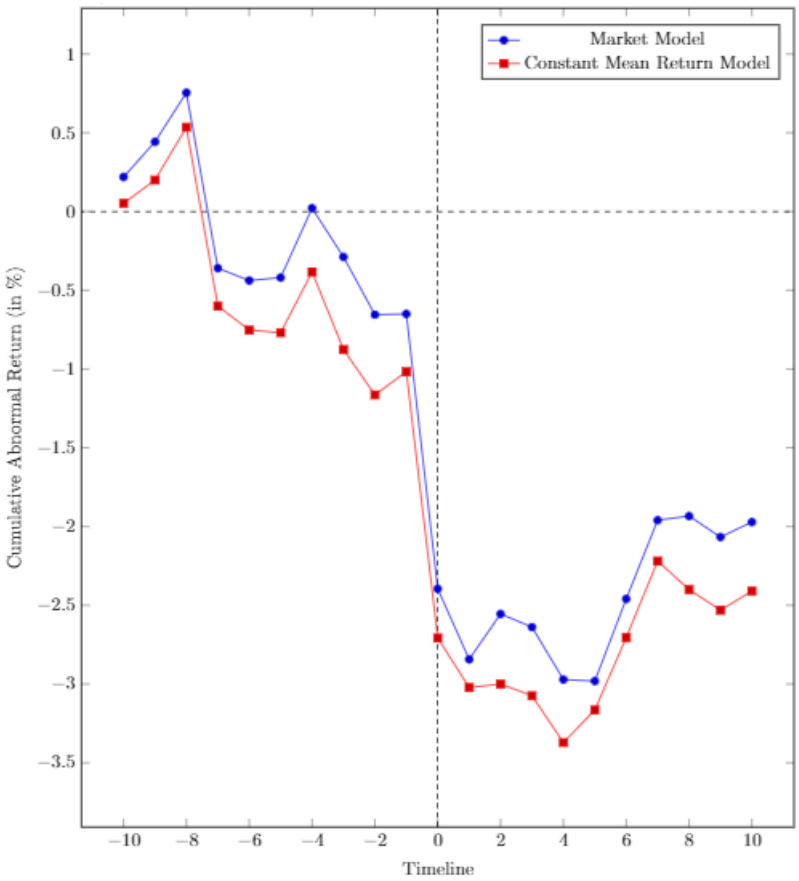
that there is limited recovery. As expected, the stock market reaction is weaker without the outliers.

New insight from the analysis of the outliers is that the effect seems more transitory, as there is stronger stock price recovery within the first ten days. However, the recovery is not statistically significant.

8.2 Alternative Normal Performance Models

The abnormal return is calculated using normal performance models. In this thesis we use the Market Model because of its favourable qualities and because it is the most common model for conducting event studies (MacKinlay, 1997). However, the choice of normal performance models influences the results. Thus, we provide a comparison of the Market Model and the Constant Mean Return Model estimation of CAR below. The Constant Mean Return Model and the Market Model are estimated as described in equations 4.2-4.4.

Figure 8.2: CAR - Normal Performance Models



From figure 8.2 one can observe that there are small differences in the estimation of CAR for the two models. To further explore the size of the differences, we provide a table with hypothesis tests below.

Table 8.2: CAR for Normal Performance Models

Timeline	Market Model (1)	Constant Mean Return (2)	Difference (3)
[0]	-0.017***	-0.017***	-0.001
	(5.60)	(5.59)	(0.08)
[-1, 1]	-0.022**	-0.019**	-0.003
	(3.51)	(3.07)	(0.30)
[-5, 5]	-0.025**	-0.024**	-0.001
	(2.46)	(2.40)	(0.07)
[-10, 10]	-0.020	-0.024*	0.004
	(1.45)	(1.82)	(0.20)
Post-Event Day Windows			
[1, 5]	-0.006	-0.005	-0.001
	(0.79)	(0.60)	(0.11)
[1, 10]	0.004	0.003	0.001
	(0.31)	(0.22)	(0.09)
Observations:	42	42	42

Note: Numbers in the brackets represent days relative to the event day. T-statistic in parentheses. Two-tailed t-test. * p <0.10, ** p <0.05, *** p <0.01.

Table 8.2 shows that the models provide similar results. The difference between the models is tested using a two-tailed studentized t-test. From column 3, we can see that there are no statistically significant differences between the estimated abnormal return for any time intervals. In addition, the most prominent differences are for the entire event window, where the numerical difference is 0.4%, and for the event day where the numerical difference is 0.1%. In conclusion, the results are independent of the choice of normal performance model.

8.3 The Cross-Sectional Regression without Outliers

As previously mentioned, the residual errors are not normally distributed in the cross-sectional regression analysis. This affects the t-statistics. When excluding the outliers from the data sample used in the regression, the OLS assumptions hold. The analysis without outliers is provided below.

Table 8.3: Cross-Sectional Analysis of Hack Announcements without Outliers

	CAR			
	[0]	[0, 2]	[0, 5]	[0, 10]
Number of Records Lost	-0.016***	-0.029***	-0.030**	-0.033*
	(3.11)	(3.44)	(2.27)	(1.80)
Data Sensitivity	0.003	0.014	0.016	0.018
	(0.62)	(1.49)	(1.14)	(0.90)
Time	-0.005	-0.014	-0.017	-0.010
	(0.95)	(1.68)	(1.32)	(0.53)
Intercept	-0.003	0.003	0.003	0.012
	(0.52)	(0.33)	(0.26)	(0.67)
Observations	38	38	38	38
Degrees of Freedom	34	34	34	34
R^2	0.241	0.311	0.177	0.100
Adjusted R^2	0.174	0.250	0.104	0.021
Note:		*p<0.1; **p<0.05; ***p<0.01		
		T-stat in parenthesis.		

From table 8.3, one can observe that the time variable is not statistically significant when the outliers are excluded. However, for number of records lost, the statistical significance is higher. This further strengthens the conclusion that we find support of the hypothesis that the number of records lost has an impact on the stock market reaction. However, our conclusion regarding time is dependent on including the outliers.

9 Critical Assessment

In this section we assess the data sample critically by discussing the sample size and the uncertainty regarding the event date. Furthermore, the limitations of the event study methodology are outlined by discussing the assumptions for the methodology and our choice of research design.

9.1 Limitations of the sample

Sample size

As a result of limited data availability and the strict selection criteria for event studies, we have a small sample. With 42 observations we were able to investigate each hack to understand its context. Also, we were able invest time in finding the exact time and date for the public announcement of the hack to identify the first trading opportunity after the announcement. However, the size of the data sample imposes some limitations on our thesis.

Skewness is often higher for small samples because outliers could have a larger impact on the results (Wooldridge, 2013). To control for the impact of outliers, the main analysis was repeated in section 8, without the four observations that draw the strongest market effects.

When using small samples there is also an increased risk of not rejecting a false null hypothesis (type 2 error). Our data sample consists of 42 observations and for the analysis of the impact of GDPR the data sample contains 33 observations. In addition, when analysing the determinants, the observations are split into two groups, resulting in even smaller samples. Consequently, the statistical power of the analysis is reduced and there is a risk of type 2 errors.

Uncertainty of Event Date

The event dates are established by examining when the first articles regarding the hacks were published. However, this way of identifying the event day may not capture the correct date, as the market may have been aware of the event before the article was published.

Thus, there is uncertainty tied to the speed of information to market participants and leakages of information prior to official announcements. Additionally, the information about a hack is likely to be announced in stages, as the severity of the hack often is not known the day the hack is discovered. To account for the possibility of imprecise dates of the announcement of hacks, the event window contains ten days before and ten days after the event date. This permits us to analyse the CAR of different time intervals within the event window.

9.2 Inherent Limitations of the Methodology

The Assumptions Behind the Method

Critics of the event study methodology will stress that there are several general limitations and weaknesses of the methodology. This is because it builds on assumptions that are disputed. For instance, the market efficiency hypothesis is widely researched, but the financial literature has not reached consensus on whether it holds. Also, the assumption that players in the market are rational does not necessarily hold. In addition, Kothari and Warner (2004), argue that predictions of normal return based on expected return models such as the Constant Mean Return Model and the Market Model are imprecise. Finally, the assumptions behind the statistical hypothesis tests that are conducted to conclude on whether the CAR is different from zero does not necessarily hold, especially not for small samples. According to Brown and Warner (1985), the cross-sectional test is prone to event-induced volatility, which lowers the statistical power. However, despite the limitations and weaknesses, the event study methodology is widely accepted.

The Choices of Research Design

The research design of an event study impacts the results. Examples of research design choices are the length of the estimation window, the length of the event window, the choice of normal performance model and the input to the normal performance model.

The estimated parameters in the normal performance models may differ with the length of the estimation window. In this thesis we use 200-days to minimize the variance of the daily return. One could also use a shorter estimation window to better reflect the most recent stock movements. The drivers behind the β are the capital structure of the

company and how cyclical the industry is. These factors may change; thus the estimation window should not be too long so that the movements in the estimation window reflects the stock movement in the event window.

The Market Model is sensitive to the choice of market portfolio. In this thesis we use the MSCI World Stock Index because the data sample consists of companies from all over the world. The results would differ with another market portfolio.

10 Conclusion

The objective of this thesis was to increase the knowledge about the stock market reaction at the announcement of hacks for companies that store personal data. The intention was to raise awareness of the trade-off between investing in cyber security and carrying the cost of being hacked. The 42 greatest hacks in the world between 2007 and 2020 are studied by using the event study methodology. We find a statistically significant CAR of negative 1.7% on the event date for the companies in the event pool. Moreover, the effect is not transitory which indicates a long-term effect on the stock price. As a benchmark for cyber security investments, we calculate the value of the negative CAR and find that eleven companies in the event pool have an estimated cost of above one billion dollars related to the announcement of hacks. The conclusion of our first hypothesis is consistent with most previous research. When excluding the four most extreme observations, the overall conclusion is unchanged. Hence, our findings provide IT managers and top executives an incentive to invest in preventive measures to protect personal data, as shareholder value is probably at risk.

Furthermore, we explore the role of consumers and regulatory agencies in inflicting financial consequences on the companies that are hacked. First, we hypothesise that when there are many victims in a hack, and the data extracted is of sensitive character, the stock market reaction is relatively stronger. We find weak support for this hypothesis. The number of records lost in the hack seems to influence the strength of the stock market reaction, which indicates that there is some incentive alignment between top executives and customers. The sensitivity of the data on the other hand does not.

Second, we hypothesise that data protection laws that enables authorities to impose fines and sanctions on companies that are hacked, influence the market reaction by increasing the expected tangible and intangible costs to the company. We find that on the event day the average stock market reaction is stronger after the GDPR was implemented, however the difference between the two groups is not statistically significant and the companies that were announced hacked after the GDPR was put into effect recovered relatively faster. However, the observed determinant may be time. In the cross-sectional regression analysis, we find that the negative stock market reaction increased over time in our data

sample. The results are consistent with previous research.

Every day we upload more personal data online, and every day more successful hacks are carried out. Consequently, our personal data is continuously at risk of being exposed and abused. In this thesis we find that top executives should be concerned with preventing hacks because shareholder value is probably at risk. We also find that the negative stock market reaction is possibly stronger when the number of records lost is high. Maybe this effect originates from the expectation that consumers use their market power to punish companies that have been hacked. Consequently, consumers should vote with their wallets and personal data for the companies that protect their data from cyber-attacks. We do not find support for the hypothesis that the GDPR gives a stronger stock market reaction at the announcement of a hack. However, regulation internalizes the cost of hacks for the companies, thus, we believe regulation contributes to solving the issue of cyber-attacks. With this thesis we want to raise awareness of the financial consequences of being hacked for companies that store personal data. We believe that companies play an important role in ensuring that cyber-attacks are no longer one of the mayor risks to economic growth.

Suggestions for Further Research

For further research to increase the awareness of the financial consequences of cyber security events, we suggest exploring the impact of the GDPR with a larger data sample, as more hacks are disclosed over time. Moreover, the focus of this thesis is personal data, but it would be interesting to analyse the market reaction when data related to the performance of a company is extracted, such as blueprints or crucial information for the company's operations. In addition, we were not able to access data about cyber security events in Norway. However, it would be interesting to investigate the stock market reaction of cyber security events on companies listed on Oslo Stock Exchange.

References

Berk, J., & DeMarzo, P. (2014). Corporate finance. In (3rd ed., pp. 271 – 285). Pearson.

Bloomberg. (2020). *MSCI Stock Index [index].* (Last accessed October 21, 2020)

Brown, S. J., & Warner, J. B. (1980). Measuring Security Price Performance. *Journal of Financial Economics*, *8*(3), 205–258.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*, 31.

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

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers. *International Journal of Electronic Commerce*, *9*(1), 70–104. doi: 10.1080/10864415.2004.11044320

Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: Causes, challenges, prevention, and future direction. *WIREs Data Mining and Knowledge Discovery*, *7*(5). doi: 10.1002/widm.1211

Corbet, S., & Gurdgiev, C. (2019). What the hack: Systematic risk contagion from cyber events. *International Review of Financial Analysis*. doi: 10.1016/J.IRFA.2019.101386

Drinkwater, D. (2016, January 7th). Does a data breach really affect your firm's reputation? *CSO*. `https://www.csoonline.com/article/3019283/does-a-data-breach-really-affect-your-firm-s-reputation.html`.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575–1617.

Garg, A., Curtis, J., & Halper, H. (2003). The Financial Impact of IT Security Breaches: What Do Investors Think? *Information Systems Security*, *12*(1), 22–33.

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables.* (R package version 5.2.2. ed.). `https://CRAN.R-project.org/package=stargazer`.

IBM Security. (2019). *Cost of a data breach report* (Tech. Rep.). Ponemon Institute. `https://www.ibm.com/downloads/cas/ZBZLY7KL`.

Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, *Volume 80*(5), 973-993. Retrieved from `https://www-sciencedirect-com.ezproxy.nhh.no/science/article/pii/S0022000014000178`

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. doi: 10.2753/JEC1086-4415120103

Kothari, S., & Warner, J. B. (2004). Econometrics of event studies. In B. E. Eckbo (Ed.), *The handbook of corporate finance: Empirical corporatefinance.*

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

McCandless, D., & Evans, T. (2020). *World's Biggest Data Breaches & Hacks.* `https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/`. (Last accessed September 22, 2020)

McWilliams, A., & Siegel, D. (1997). Event studies in management research: Theoretical

and empirical issues. *Academy of Management Journal*, *40*(3), 626–657.

Morse, E. A., Raval, V., & Wingender Jr., J. R. (2011). Market price effects of data security breaches. *Information Security Journal: A Global Perspective*, 263–273. doi: 10.1080/19393555.2011.611860

R Core Team. (2020). *R: A language and environment for statistical computing.* `https://help.yahoo.com/kb/SLN28256.html`. Vienna, Austria. (Last accessed October 6, 2020)

Reding, V. (2014, January 28th). `https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_14_62`. (Last accessed December 6, 2020)

Strong, N. (1992). Modelling abnormal returns: A review article. *Journal of Business, Finance and Accounting*, *19*(4), 533–553.

The AME group. (2020). *Data Security Breach: 5 Consequences for Your Business.* `https://www.theamegroup.com/security-breach/`. (Last accessed October 16, 2020)

The European Union. (2020). `https://gdpr.eu/faq/?cn-reloaded=1`. (Last accessed November 1, 2020)

The World Economic Forum. (2019, 15th of January). *The Global Risks Report 2019.* `http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf`. (Last accessed December 6, 2020)

Wooldridge, J. M. (2013). Introductory econometrics: A modern approach. In (5th ed., pp. 327–334). Michigan State University: South-Western Cengage Learning.

Yahoo! Finance. (2020). *What is the adjusted close?* `https://help.yahoo.com/kb/SLN28256.html`. (Last accessed September 23, 2020)

Yayla, A. A., & Hu, Q. (2011). The impact of information security events on the stock value of firms: The effect of contingency factors. *Journal of Information Technology*, *26*(1), 60–77. doi: 10.1057/jit.2010.4;

Appendix

A1 Asset Pricing Theory

Asset pricing theory is helpful to understand how the expectation of increased costs and decreased revenue affect the stock price of a company. According to Berk and DeMarzo (2014) investors compute the value of a company by using models such as the Dividend Discount Model and the Discounted Free Cash Flow Model. Hence, these two models are elaborated below.

Dividend-Discount Model

An expanded dividend-discount model with constant long-term growth is used to calculate the present value of dividends to find the value of the company's stock (Berk & DeMarzo, 2014). In the model, dividends are the cash flows paid to the shareholders. The fair value of the investment is the present value of all future dividends and the selling price of the stock. According to Berk and DeMarzo (2014) when the investors have the same beliefs, for any time horizon T the expanded model can be written as the following equation.

$$P_0 = \frac{DIV_1}{r_E - g} + \frac{DIV_2}{(r_E - g)^2} + ... + \frac{DIV_T}{(r_E - g)^T} + \frac{P_T}{(1 + r_E)^T}$$

In the equation, DIV_1 represents the dividend paid at time 1. The required rate of return on equity is expressed as rE and P_0 is the stock price at the time of the investment. T represents the final time period in which the stock is sold. When there are expected decreases in future dividends, the value of the stock decreases.

Discounted Free Cash Flow Model

The Discounted Free Cash Flow Model is used to estimate a company's enterprise value by discounting its future free cash flows (Berk & DeMarzo, 2014). This method allows the user to value companies which do not regularly pay out dividends. Generally, the cash flows for a time period is forecasted, and then the terminal value for the free cash flow beyond the time period is calculated. The present value is the company's enterprise value.

The discounted free cash flow model can be written as equation bellow, formally.

$$V_0 = \frac{FCF_1}{1 + r_{WACC}} + \frac{FCF_2}{(1 + r_{WACC})^2} + ... + \frac{FCF_T + V_T}{(1 + r_{WACC})^T}$$

$$V_T = \frac{FCF_{T+1}}{r_{WACC} - g_{FCF}}$$

In the equation, V_0 is the discounted free cash flow, while FCF_1 is the free cash flow at time 1. Furthermore, r_{WACC} is the weighted average cost of capital, and the terminal value of the free cash flow past time period T is expressed by V_T. The constant growth rate the free cash flows grow at beyond time T is equal to g_{FCF}. When costs are expected to increase and revenue is expected to decrease, the free cash flow is affected and the price of the stock decreases.

A2 Full Datasample

The table below provides an overview of the data sample that is analysed in this thesis. It also shows which groups the events belong to. The events are sorted by CAR.

Table A2.1: Information about all the Events in the Sample

				Dummies		
Name	Date of the hack	CAR[0, 2]	Records Lost	Records Lost	Data Sensitivity	GDPR
Equifax	08.09.2017	-0.205	143 000 000	High	High	X
Zoom	02.04.2020	-0.108	500 000	Low	Low	After
Heartland	21.01.2009	-0.09	130 000 000	High	High	X
Facebook	19.03.2018	-0.071	50 000 000	High	Low	Before
Cathay Pacific Airways	25.10.2018	-0.062	94 000 000	High	High	After
Marriott International	30.11.2018	-0.055	383 000 000	High	High	After
Facebook	28.09.2018	-0.051	29 000 000	High	High	After
Sony PSN	21.04.2011	-0.043	77 000 000	High	Low	Before
Experian	02.10.2015	-0.039	15 000 000	High	High	Before
TalkTalk	23.10.2015	-0.037	157 000	Low	High	Before
Dell	29.11.2018	-0.034	100 000	Low	Low	After
Target	19.12.2013	-0.033	70 000 000	High	High	X
Apple	19.07.2013	-0.029	275 000	Low	Low	Before
Anthem	05.02.2015	-0.025	80 000 000	High	High	Before
KT Corp.	30.07.2012	-0.02	8 700 000	High	High	X
UPS	22.08.2014	-0.014	4 000 000	Low	High	Before
Honda Canada	27.05.2011	-0.012	283 000	Low	High	X
UbiSoft	03.07.2013	-0.012	58 000 000	High	High	Before
Ebay	21.05.2014	-0.011	145 000 000	High	Low	Before
Interpark	26.07.2016	-0.01	10 000 000	High	High	X
T-Mobile	24.08.2018	-0.01	2 000 000	Low	Low	After
Dixons Carphone	13.06.2018	-0.009	10 000 000	High	Low	After
Tesco Clubcard	04.05.2020	-0.009	600 000	Low	Low	After
Global Payments	04.04.2012	-0.008	7 000 000	High	High	Before
VTech	27.11.2015	-0.008	6 400 000	Low	High	Before
Adobe	03.10.2013	-0.007	36 000 000	High	High	Before
HSBC Turkey	12.11.2014	-0.006	2 700 000	Low	High	Before
Quest Diagnostics	12.12.2016	-0.006	34 000	Low	High	Before
Wendy's	07.07.2016	-0.005	1 025	Low	High	X
AT&T	09.06.2010	-0.004	114 000	Low	Low	Before
Gmail (Oracle)	10.09.2014	-0.004	5 000 000	Low	Low	Before
Sony Online Entertainment	27.05.2011	-0.003	24 600 000	High	High	Before
Nintendo	08.07.2013	0.001	240 000	Low	High	Before
Toyota	29.05.2019	0.002	3 100 000	Low	High	After
Home Depot	03.09.2014	0.003	56 000 000	High	High	X
Microsoft	15.04.2019	0.003	44 000 000	High	High	After
Sega	20.06.2011	0.007	1 290 755	Low	High	Before
JP Morgan Chase	02.10.2014	0.014	76 000 000	High	High	Before
Dominios Pizzas (France)	16.06.2014	0.032	600 000	Low	Low	Before
TD Ameritrade	14.08.2007	0.038	6 300 000	Low	Low	X
Nintendo	09.06.2020	0.066	300 000	Low	High	After
Citigroup	09.06.2011	0.074	360 083	Low	High	Before

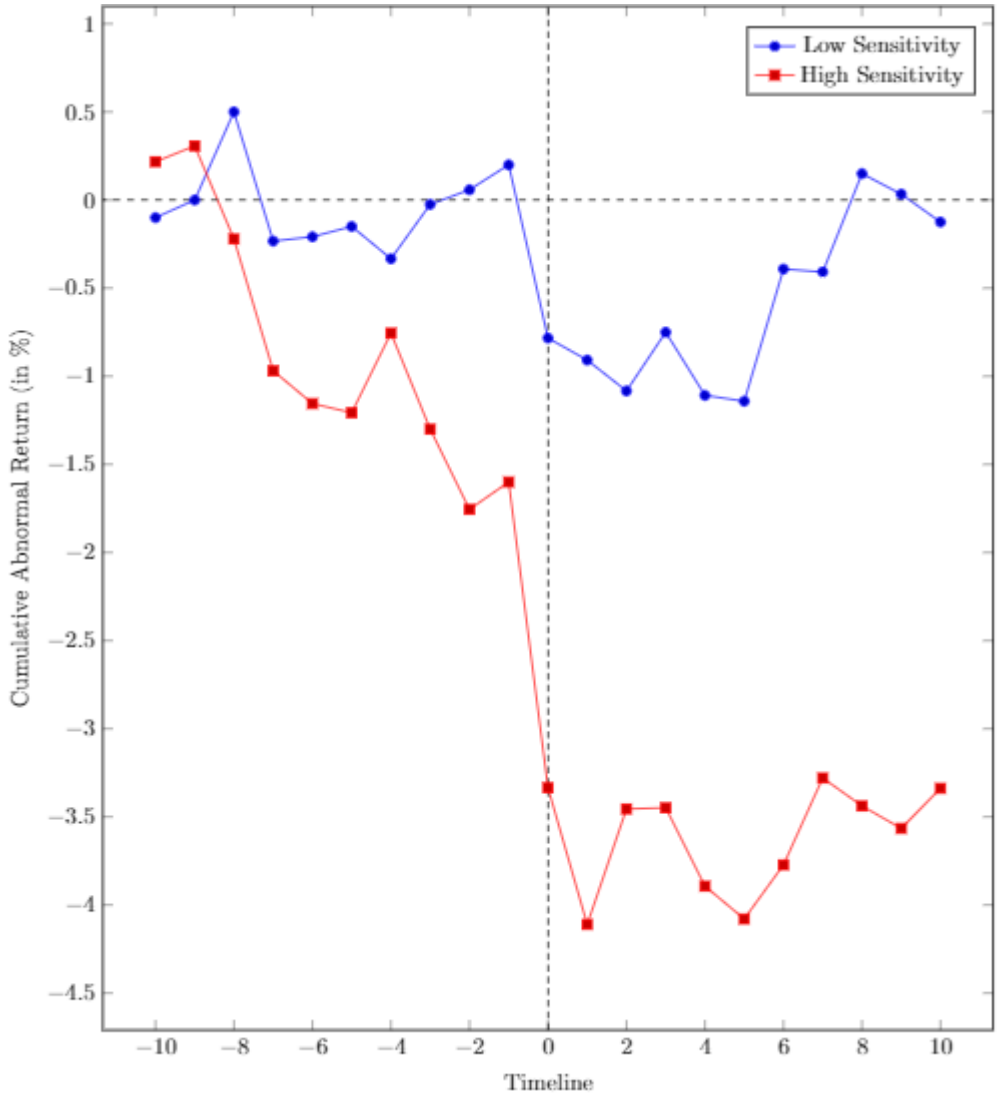
The events are sorted from smallest to highest CAR for the time interval from the event day till two days after the event. The table shows how the different events in the sample were sorted when the analysis was conducted.

A3 Figures without Zoom

Data Sensitivity in Analysis without Zoom

From the figures below, one can observe the effect of data sensitivity and the GDPR when Zoom is excluded from the sample. This is because Zoom increases the difference between the groups, due to high volatility during the event window. Zoom was originally in the group with hacks of low sensitivity and in the group after the GDPR was implemented. When excluding this outlier, one can see that the groups that included Zoom have shifted down and the difference between the two groups is smaller. The conclusions in the analysis above do not change.

Figure A3.1: Data Sensitivity without Zoom



GDPR in Analysis without Zoom

Figure A3.2: GDPR without Zoom