DATA BREACHES AND THE ROLE OF INVESTORS

Author: W.M. van Wieringen

Student number: 400631

Thesis supervisor: dr. R. Wang

Erasmus School of Economics

MSc. Economics and Business

Financial Economics

Abstract

In this thesis I provide an empirical analysis on the role of investors surrounding the announcement of data breaches for the period 2005-2018. A sample of 313 data breaches is used for 216 distinct companies that are located in the United States. Data breaches are obtained from the PRC (Privacy Rights Clearinghouse) database and I combine this data with firm and industry information from Compustat. I find evidence that a firm's visibility positively affects the likelihood of experiencing a data breach. Furthermore, I find evidence that the announcements of data breaches generate abnormal stock returns. Moreover, my findings indicate that data breaches that involve the loss of financial or valuable information have a greater impact on a company's stock price than data breaches that do not involve financial or valuable information. However, no strong evidence is found that the nature of a breach impacts the magnitude of the stock price reaction. Since data breaches significantly impact the stock price of its targets, investors could benefit financially by betting against the company's stock price. However, I find no evidence in both the equity market neither in the option market that informed trading takes place surrounding data breaches.

Table of contents

1. Introduction

In the first week of September 2017, Equifax, an Atlanta based, consumer credit reporting agency reported that hackers had gained access to personal information of 148 million customers for several months, including: social security numbers and credit card numbers. This data breach can be seen as one of the largest of the last years (Bernard, Hsu, Perlroth and Lieber, 2017). At the day of the announcement Equifax's stock price tumbled 13.8% and one week later the company already lost 30.3% of its total value, corresponding to $5.1 billion. As illustrated by the case of Equifax, the costs of data breaches are substantial and a real threat for companies operating in several industries (Kamiya, Kang, Kim, Milidonis and Stulz, 2018; Akey, Lewellen and Liskovich, 2018). According to a report of McAfee, cybercrime has increased from $500 billion in 2014 to $600 billion in 2018, which corresponds to 0.8% of the global GDP. By selling information on the black market and monetizing the gains through digital currencies it has become easier for cybercriminals to profit from stolen data.

The stolen information is sensitive and therefore can be valuable for different actors. One group that could be interested in this information is investors. Investors are constantly looking for valuable information and they could make potential investments based on the stolen information. In the days around the announcement of a data breach, the stock price of listed firms typically tumbles (Hilary, Segal and Zhang, 2016; Johnson, Kang, Lawson, 2017; Bianchi and Tosun, 2019). Investors could benefit financially by betting against the target's stock price. By taking a short position or buying put options, investors are able to profit in case the underlying stock decreases (Mitts and Talley, 2018).

Several months after the data breach of Equifax was disclosed, three employees including Mr. Ying, who was to become the company's chief information officer, were convicted of illegal trades. By selling his stocks Mr. Ying avoided losses of $117.000 (SEC, 2018a). Another employee bought put options on the company's underlying stock and sold these the day after the breach was announced. The software engineer earned $75.000 (SEC, 2018b).

Due to technological developments and more opportunities to gain illegal profits, regulation and enforcement have emerged. Although informed trading typically occurs around mergers and acquisitions (Keown and Pinkerton, 1981; Jarell and Poulsen, 1989), earnings announcements (Womack, 1996) and analyst recommendations (Christophe, Ferri and Hsieh, 2010; Blau and Wade, 2012), the announcement of data breaches could be an event in which privileged information results in financial benefits.

In this study, I will investigate the role of investors surrounding the announcement of different data breaches. First, I will analyze which firm- and industry-specific characteristics influence the likelihood of firms

experiencing a data breach. Secondly, I will perform an event study on stock returns to determine whether there are abnormal trading returns surrounding the announcement of data breaches. Thirdly, I will perform an OLS regression to investigate which firm- and industry-specific characteristics have an impact on the magnitude of the stock price reaction. I distinguish between the different types of data breaches and try to assess whether the nature of the breach as well as losing valuable information impacts the magnitude of the stock price reaction. Based on the current literature, I expect the stock price reaction of the complete sample to be significant and negative. Therefore, investors could potentially benefit financially by betting against the company's stock price. For that reason, I will analyze trading volumes in the equity market and search for any abnormal patterns in the run-up to data breach announcements. Lastly, I will seek for evidence of informed trading in the option market, as investors could shift away from the equity to the option market.

So far, there is only one paper that focuses on data breaches and informed trading. Mitts and Talley (2018) investigate the relation between 145 data breaches and informed trading. Since this is the only paper that focuses on data breaches and informed trading, there is a gap in the literature. Therefore, in this paper I will try to fill the gap by investigating the role of investors around different types of data breaches and examine attacks that were initiated by insiders as well as by outsiders.
More specifically, this paper will focus on the occurrence of informed trading around the announcement of data breaches. Therefore, the main research question of this paper is:

Research question: "Do investors benefit financially from informed trading surrounding the announcement of data breaches?"

The literature regarding data breaches is still limited and so far, only one study explicitly investigates and links informed trading to data breaches. Therefore, this study will contribute to the existing literature in several ways. Firstly, this study contributes to the existing literature by including more and recent data, thereby extending current research on the relation between data breaches and stock price reactions. Secondly, by distinguishing between the different types of breaches, this study tries to examine whether the nature of a breach and the loss of valuable information impacts the stock price reaction. Thirdly, this study incorporates more and a greater variety of firm- and industry-specific characteristics to find out which firms are more likely to experience a data breach. Lastly, I fill the gap in the literature by assessing the role of investors surrounding the announcement of data breaches. This work is not only relevant for academics, but could also have implications for regulators, as informed trading calls for regulatory action.

2. Literature review

This section discusses the literary motivation behind this paper. Section 2.1 briefly discusses several concepts and the main terminologies used in this paper. Section 2.2 debates which firms are more likely to become a victim of a data breach. Section 2.3 discusses the short- term effects of data breaches on a company's stock price. Section 2.4 summarizes the effects of data breaches on market activity and liquidity. Section 2.5 discusses the long-term effects of data breaches on the operational performance of targets and section 2.6 outlines the effects of data breaches on target policies. Section 2.7 evaluates the current literature on informed trading and finally section 2.8 formulates the hypotheses which are tested in the remainder of this paper. The key concepts in this section are data breaches, cyberattacks and informed trading. Furthermore, the hypotheses for this thesis will be developed based on evidence presented in previous studies.

2.1 Data breaches

The relevant literature related to data breaches is still limited but rapidly growing as it is becoming an important topic on the public agenda. In 2015 former US president Barack Obama declared a national security emergency as a result to the threats created by cyberattacks. The terms 'cyberattack', 'security breach' and 'data breach' are often used interchangeably. This paper refers to a data breach as: "accidental or unlawful destruction, loss or unauthorized disclosure of information that is stored in the cyber environment" **(**International Standards Organization, 2016, ISO/IEC 27050, 3.3). Most of the relevant literature has focused on an extensive group of events categorized as data breaches. Data breaches could take on different forms such as: external hacks, lost or stolen portable devices, payment card fraud, an insider who breaches information and physical theft of documents containing information. Although, often used as a synonym for data breach, a cyberattack refers to an attack initiated by an external party. The main difference between a data breach and a cyberattack is that a cyberattack is deliberately initiated by an outside party whereas a data breach could be the result of an outside as well as an inside party (Von Solms and Van Niekerk, 2012). This paper investigates the role of investors surrounding the announcement of data breaches, and distinguishes between external hacks, lost or stolen portable devices, payment card fraud, an insider who breaches information and physical theft of documents. In this study I use the words 'Cyberattack' or 'Hack or Malware' interchangeably.

Finally, the words 'cybersecurity' and 'information security' have been used interchangeably in the current literature. However, information security can be subdivided into the protection of the physical environment and the protection of information that is stored in the cyber environment, while cyber security solely includes the protection of information that is stored in the cyber. According to Von Solms and Van Niekerk (2012),

"cybersecurity is the collection of tools, policies and security concepts that can be used to protect the cyber environment of a company in which the customer's information is stored".

2.2 Firm-specific characteristics of targets

Kamiya et al. (2018) examine whether certain firm- and industry-specific characteristics are associated with the likelihood of experiencing a cyberattack[1]. They argue that firms that are more visible have a higher chance of becoming the victim than firms that are less visible. The results show that companies with more total assets, which are older, owned by a larger percentage of institutional block holders and included on the list of Fortune 500 companies are more likely to become the target of a cyberattack. The reason behind these findings could be that these firms have more valuable information stored in the cyber than smaller and less visible firms. Moreover, lower leverage, poorer past stock performance, higher profitability, higher growth opportunities, more intangible assets, less capital expenditures and less investments in R&D are positively related with the probability of becoming a target. These findings indicate that firms that rely more on customer information, have a bigger chance of becoming a target. Camp and Lewis (2006) suggest that bigger companies are more likely to become a target for hackers, since those companies have more money. They also suggest that larger companies invest more in the protection of the cyber environment than smaller companies, as the risk for large companies is much higher than for smaller companies. Contrary to these results, Akey et al. (2018) and Lending, Minnick and Schorno (2018) find that less visible firms have a higher chance of becoming a target. Companies with a lower market capitalization are more likely to become a victim of a data breach. Akey et al. (2018) also find that targets with a lower Market-to-book ratio (M/B) are more likely to become a target. Profitability does not seem to have an influence whether a firm experiences a data breach or not. The opposite results could be the effect of a sample selection difference. The studies of Akey et al. (2018) and Lending et al. (2018) include data breaches that take on different forms whereas, Kamiya et al. (2018) only focus on cyberattacks caused by hacking or malware that resulted in the loss of personal information.

2.3 Short-term effects of data breaches on shareholder wealth

Typically, studies that investigate data breaches examine the short-term impact on a target's stock price. The existing literature disagrees both on the significance of the stock market reaction and on the factors that have an impact on the magnitude of the reaction. This could potentially be the result of the different types of data breaches that are investigated in these papers as well as the small samples that have been used. Garg, Ashish, Curtis and Halper (2003b) investigate the short-term effect of data breaches on the target's market value. The

[1] Kamiya et al. (2018) included events that were initiated by an outside party and only included those events that lost personal information.

sample includes: web site defacements, Denial of Service (DoS) incidents[2], theft of customer information and theft of credit card information. All different data breaches had a significantly negative impact on the target's short-term value. By dividing data breaches into categories, the study shows that the loss of credit card information creates the most negative stock price reaction. Also, the stock price reaction to DoS incidents was more severely than for website defacements. In line with this paper, the study of Cavusoglu, Mishra and Raghunathan (2004) investigate the effect of data breaches on a firm's market value and the included different types of breaches. The paper shows that stock prices declined significantly. However, as opposed to the study of Garg et al. (2003b), their study shows that the nature of the breach has no impact on the magnitude of the stock movement. The results of the study of Gatzlaff and McCullough (2008) are in line with the paper of Cavusoglu et al. (2004) and find that the nature of a breach has no impact on the magnitude of the stock price movement. Nevertheless, they do find a significantly negative stock price reaction.

In addition to the significantly negative market reaction, the study of Gatzlaff and McCullough (2008) shows that the magnitude of the stock price reaction is significantly influenced by firms with higher growth opportunities, and a higher market capitalization mitigates the negative impact on a company's wealth. They also find that in case the data breach was detected in a subsidiary, the stock price reaction was less significant than in case the breach happened in the mother company. In line with these findings, the study of Cavusoglu et al. (2004) find that the market reaction is significantly higher for smaller firms than for larger companies.

As opposed to the papers mentioned earlier, the results of Campbell, Gordon, Loeb and Zhou (2003) and Hovav and D'Arcy (2003) do not show a significantly negative stock price reaction for the complete sample. The sample of Campbell et al. (2003) consists of confidential as well as non-confidential data breaches. The complete sample includes 43 events of which 25.6% of the breaches contained confidential information. Of the data breaches 32.6% represent virus attacks and 18.6% represent DoS incidents, which are not related to confidentiality. Including viruses in the sample could mitigate the effect on the overall sample, since viruses often have an impact on more than one company (Garg et al, 2003b). When solely looking at the breaches that involve unauthorized access to confidential information, the results of Campbell et al. (2003) and Hovav and D'Arcy (2003) are highly significant. These results are in line with the fact that cyber security protects different information sources and when the data breach involves confidential data or valuable information, the stock price reaction is more severe.

To further support the argument that data breaches significantly affect stock prices, Spanos and Angelis (2016) did a systematic review of the existing literature related to data breaches and the effects on stock prices. Based

[2] A Denial of Service attack is a cyberattack in which a machine or network is made unavailable

on 37 papers with 45 different studies they conclude that the majority (75.6%) of the studies report a negative significant reaction.

2.4 Short-term effects of data breaches on market activity and liquidity

In addition to shareholder wealth, Bianchi and Tosun (2018) focus on short-term market trading activity and liquidity. The study includes data breaches subdivided into stolen hardware, insider attacks, poor security and hacking. They find that for targeted firms the traded volume is positively and significantly affected at the day of the announcement, and liquidity, measured by an increase in the bid-ask spread decreased significantly. Traded volume and liquidity are not affected the day before or after the announcement. A difference-in-difference analysis finds that the increase in traded volume is the result of a selling pressure. This is in line with the expectation that a data breach could harm the firm's valuation and investors rather sell than buy the particular stock.

2.5 Long-term effects of data breaches on target operational performance

The announcement effect on a target's stock price seems to disappear after a few days (Acquisti, Friedman and Telang, 2006). However, data breaches could have a long-term effect on the company's performance and policy. Costs incurred by targets can be divided into short- and long-term costs. Short-term costs can be seen as tangible costs such as: litigation costs, replacement costs of the IT system and loss of projects due to problems within operations. Longer-term costs include the potential loss of current customers, costs of not being able to acquire new customers and the loss of business partners. Breached firms incur tangible and intangible costs which could impact the longer-term financial performance of companies (Tsiakis and Stephanides, 2005). The study of Ko and Dorantes (2006) examine the effect of data breaches on the organizational performance in the four quarters after the event. The overall performance, which includes sales and operating income, did not decrease in the next quarters after the breach. Only the return on assets decreased significantly. However, based on a between firm difference analysis, the performance of the control group outperformed the performance of the group that experienced a data breach. This indicates that the targets' long-term overall performances were affected by a data breach. Bianchi and Tosun (2018) make the assumption that data breaches have a negative effect on the targets' reputations. A data breach does not directly affect the operational activities of a company, but could affect products and services indirectly through the damaged reputation. Although they assume that data breaches damage the reputation, the study does not show a significant long-term impact on the targets' operating performance. Long-term sales growth and return on assets were not affected by the data breaches. However, the targets' Price-to-Earnings ratios (PE) fell significantly. As the PE ratio is an indicator for future

growth prospects, the findings suggest that reputational damage results in a decline of the future growth prospects.

2.6 Long-term effects of data breaches on target policies

Due to reputational damage and operational risk, target firms could be changing their long-term policy to overcome breaches in the future and mitigate the negative effect that the breach could have on the firms' financial performances.

Bianchi and Tosun (2018) investigate the effect of data breaches on the long-term policy of target firms. The study shows that the targets' long-term policies typically become tighter; dividends and R&D expenses are significantly reduced up to five years after the event. This could indicate that companies reduce these expenses to manage the increased risk.

Contrarily, total pay and incentive pay of chief executive officers (CEO) increases years after the breach compared to control firms. A potential reason could be that target firms invest more in their management teams to address structural problems and hereby motivating their management. One could also expect that target firms fire the responsible person such as the manager of the risk committee or the CEO. However, the study finds no effect on the turnover rate of managers and CEOs. As opposed to these findings, Lending et al. (2018) find a significant effect on the targets' CEO and chief technology officer (CTO) turnover rates. According to their study, targets are more likely to change the CEO and CTO compared to control firms. A possible explanation could be that the firms' boards of directors propose to hire a new CEO to send the companies into new directions or that the board would like to appoint a new CEO to win back the trust of the stakeholders (Larcker et al., 2017). Both actions that could be taken are in place to rehabilitate the reputation of the company. The study of Akey et al. (2018) show that CEOs of affected firms are less likely to leave the company after a data breach.

According to Akey et al.'s (2018) research, firms try to recover their reputation by increasing its investment in Corporate Social Responsibility (CSR). The study finds that firms that are affected by a data breach, experience cumulative abnormal returns in the 30 days following the disclosure of the breach, however this effect is smaller at firms who invested more in CSR ahead of the breach. The paper mentions that in case the negative reputation shock has a short-lived effect on the firms' stock price and financial performance, CEOs will not invest in expensive activities. Such as CSR, which takes long to materialize. In case the CEO expects the negative reputation shock to have a long-term effect on the firm, it might take costly actions to rebuild the company's reputation.

Akey et al.'s (2018) results find that data breaches have a long-lasting effect on a firm's valuation and profitability. They find a one-year change in the M/B ratio of -10% to -20% for target firms compared to control firms. Also, the return on equity (ROE) and the PE ratio show that both financial performance and forward-looking profitability are negatively affected until four years after the breach. In addition to the already mentioned factors, Kamiya et al. (2018) find that in the three years after a cyberattack, financial health of targets deteriorated. This was measured by a decrease in credit ratings and an increase in bankruptcy filings.

2.7 Informed trading

As data breaches have a negative short- as well as a long-term effect on a company, it could be profitable to trade in advance of the data breach announcement. Many studies investigate the impact of informed trading on financial markets. Informed trading reduces the trust in the financial sector and this could reduce investor participation in the financial market. Eventually this could lead to higher cost of capital (Easley and O'hara, 2004). Based on the definition of the Security and Exchange Commission (SEC), informed trading refers to "buying or selling securities on the basis of material, nonpublic information". Insider trading also includes 'tipping' information and trading by the person who received this non-public information[3].

Informed trading around the announcement of several corporate events have been investigated broadly. However, the literature on informed trading around the announcement of a data breach is limited. The private information that investors, who are involved in informed trading, receive can contain two signals. Firstly, it contains information about the timing of a disclosure. Secondly, it contains information about the direction of the announcement return as a result of the news. By receiving information in advance of a disclosure, investors are able take a position and gain substantial financial benefits (Augustin, Brenner and Subrahmanyam, 2018).

There is reason to believe that informed trading could also take place in the event of a data breach, since there are some predictable short-term effects of data breaches on company stock performance, as outlined earlier. Arcuri, Brogi and Gandolfi (2017) examine the effect of data breaches on financial and non-financial firms. They find that the average cumulative abnormal returns are significantly negative. However, the significance of the event windows differ. The event windows (-10,-1), (-5,-1) and (-3,-1) are negative (-1.08%, -0.87% and -0.90%) and significant at the 90% confidence level. These outcomes may imply that informed trading takes place. However, other studies do not show any link between pre-announcement returns and data breach announcements (Gatzlaff and McCullough, 2008; Johnson et al., 2017 and Kamiya et al., 2018).

[3] https://www.investor.gov/additional-resources/general-resources/glossary/insider-trading

2.7.1 Short selling

As data breaches typically have a negative impact on the company's short-term stock price, investors could benefit financially by betting against the company's stock price. Investors could profit by short-selling the target's stock before the breach is announced. According to the SEC a short sale is the sale of a stock that an investor does not own, but the transfer of a stock borrowed by the investor. The investor liquidates the position by returning back the borrowed stock to the lender in the future, usually by buying the security on the open market. Traders who sell short believe that the underlying stock price will drop, purchase the stock at the open market and try to benefit financially.

So far, there are no studies that investigate the relation between short-selling and data breaches. However, several studies investigate the relation between other negative corporate events and short selling. Christophe, Ferri and Hsieh (2010) investigate the relation between informed trading and analyst downgrades. The paper studies pre-announcement short selling of analyst downgrades for firms that are listed on the Nasdaq in the years 2000 and 2001. They find a significant increase in short-selling in the three days prior to the public announcement. They show that in the event window (-3,-1), the average daily short selling is four times higher than normally during the sample period. This could imply that investors had knowledge of the upcoming analyst downgrade before it was made public. Another study that investigates the relation between negative announcements and informed trading is the paper of Blau and Tew (2014). They study whether short selling activity surges during the pre-filing period of class-action lawsuits. This study was examined as a reaction to several complaints that filing law firms were leaking information to investors about the date of the filing. The results show an abnormal increase in short-selling activity in the days ahead of the class-action lawsuit filing. This indicates that investors received information prior to the upcoming event. Finally, the study of Christophe, Ferri and Angel (2004) examine short-sale transactions in the run-up to earnings announcements. They find a significantly negative relation between short-selling before the event and the stock price after the earnings announcement. This could signify that a meaningful portion of short-sellers are informed. The study also shows that the statistical results are stronger for firms without tradable put options. The reason could be that informed traders prefer to make use of put options rather than short-selling the stock, because buying put option is less expensive that short-selling a stock. Investors should only short-sell when the expected profits exceed the costs of the transaction. Therefore, investors can limit their risk by acquiring a put option, giving them the right, not the obligation, to sell the stock.

2.7.2 Option trading

As discussed above, a possible alternative for short-selling could be to buy a put option on the firm's underlying stock. Arguments for trading in the option market instead of the equity market include: high leverage and the downside protection in the option market (Charakvarty, 2004). The SEC defines an option as: "a contract to buy or sell an underlying asset at a fixed price on a specified date". A put option is: "a contract that gives the buyer the right to sell shares of an underlying stock at the strike price for a specified period of time" (SEC, 2010). A put option is a downside bet, because it is worth more if the underlying stock price declines.

Option markets may be relatively more appealing for traders who act on negative information. Option trading tends to increase with the liquidity of the option market. Therefore, illiquidity in the stock market could make the option market an attractive alternative (Easley and O'hara, 2004).

Mitts and Talley (2018) investigate the relation between informed trading and cybersecurity breaches. They mention that investors have a strong incentive to take a short position in firms that have been the target of a data breach. They empirically test this, by analyzing the open interest and trading volume in put options of listed companies that have been the target of a data breach. Open interest refers to the total number of outstanding options or future contracts on the stock of the underlying company. Volume refers to the amount of options or future contracts traded during a period of time. Mitts and Talley (2018) analyze 145 events that have occurred in the period of 2010-2016. The results of their cross-sectional analysis indicate that in the two months prior to the data breach the average open interest and the average trading volume in put options increased significantly for target firms compared to control firms. These results imply that investors have early notice of the upcoming announcement and are able to exploit the information they have received.

Although, this is the only study that investigates the relation between data breaches and option trading, more studies investigate the relation between negative events and the option market. For instance, Poteshman (2006) finds that in the days prior to the terrorist attack on 11 September, 2001 there was unusual option trading. The study shows a relation between put option trading in the run-up to the 9/11 attack. He uses volume ratios to measure whether there was abnormal trading in the days ahead of the terrorist attack. Poteshman (2006) looks into long and short call and put volumes. An abnormal long put volume was examined for the two airlines that crashed into the World Trade Center, American Airlines (AMR) and United Airlines (UAL), in the days prior to the terrorist attack. By buying put options on the underlying stock of the two airlines, insiders tried to benefit financially. Although this is one of the most striking cases, there are more studies about the relation between negative corporate events and option trading with inside information. For instance, Ge, Hu, Humphery-Jenner and Lin (2014) examine the relation between informed trading and the announcement of bankruptcy filings. As the option market can be an alternative for trading in the equity market, they investigate the effect of

bankruptcy on option trading. They study the relationship between the pre-filing ratio of options trade to stock trade (o/s) and the returns around the disclosure of the filing, and hypothesize that in the period ahead of a bankruptcy filing, option trading predicts stock returns. They find a significant negative relationship between filings announcement returns and pre-filing level of options trade to stocks trade.

2.8 Hypotheses

Since both the likelihood and the cost of data breaches have become such an important risk for companies, it is important to address which factors influence the likelihood of becoming a victim as well as the impact of those breaches. Since only one study addresses the occurrence of informed trading around the announcement of data breaches, I aim to assess whether investors are involved in informed trading around the announcement of data breaches and go into more detail by distinguishing between the different types of breaches. I strive to answer multiple hypotheses related to data breaches and informed trading.

2.8.1 Firm-specific characteristics of data breach victims

According to the study of Akey et al. (2018) and Lending et al. (2018) smaller firms are more likely to become the victim of a data breach. Firm size is measured as the natural logarithm of total assets. Akey et al. (2018) also find that targets with less growth opportunities, measured by M/B ratio are more likely to become a target.

In contrast with the above-mentioned studies, Camp and Lewis (2006) suggest that larger companies are more likely to become a target for hackers as those companies have more money. In line with this suggestion, Kamiya et al. (2018) find that more visible firms have a higher chance of experiencing a cyberattack than less visible firms; firms with more total assets, a higher percentage of shares held by institutional block owners and firms that are included on the list of Fortune 500 companies have a higher chance of becoming a target. But these findings need not be contradicting. In fact, it could be that for particular types of data breaches, i.e. cyberattacks, visibility is actually positively related to the probability of being a victim of such an attack. However, my study investigates different types of breaches and the results of prior studies indicate that for samples with different breaches, the likelihood of firms experiencing a data breach is higher for less visible firms. Assuming that firm size, percentage of shares held by institutional block owners and presence on the list of Fortune 500 companies are good proxies for firm visibility, I arrive at my first hypothesis:

H1: Data breaches are more likely to occur at firms that are less visible

I will also distinguish between the different types of breaches to find any differences in firm-specific and industry-specific characteristics between the targets.

2.8.2 Stock market impact of data breaches

Given the findings of (Hilary et al., 2016; Johnson et al., 2017; Bianchi and Tosun, 2019; Garg et al., 2003b and Cavusoglu et al., 2004) who suggest that in the short term the stock price is negatively affected by a data breach, I arrive at my second hypothesis:

H2: Data breaches have a short-term negative impact on the targets' stock price

Several studies find evidence that breaches that involve the loss of confidential data or valuable information impact the magnitude of the stock price more severe. The results of Campbell et al. (2003) and Hovav and D'Arcy (2003) show that when a data breach includes valuable or confidential information, the stock price reaction is significantly higher. This is also in line with a more recent study of Kamiya et al. (2018), who show that the magnitude of the stock price reaction differs between cyberattacks that include the loss of financial information and cyberattacks that do not. This informs my third hypothesis:

H3: The negative impact on a targets' stock price is larger for data breaches that involve the loss of financial information

As prior studies do agree that the loss of financial or valuable information impacts the magnitude of the stock price reaction, these studies not agree whether the nature of a breach has an impact on the magnitude of the stock price reaction. Garg, Ashish, Curtis and Halper (2003b) investigate the short-term effects of data breaches on the target's market value. The study investigates different types of breaches and all data breaches had a significantly negative impact on the target's short-term value. By dividing data breaches into categories, the study shows that the market reacted most heavily to credit card information theft. However as opposed to this study, Cavusoglu, Mishra and Raghunathan (2004) and Gatzlaff and McCullough (2008) find that the nature of the breach has no impact on the magnitude of the stock movement. Hereby arriving at my fourth hypothesis:

H4: The nature of a data breach has no impact on the magnitude of the stock price movement

To find any evidence on informed trading I will also test for abnormal trading volume. However, as most studies do not find any evidence of informed trading related to data breaches in the equity market, it could be of more interest to further investigate the option market.

2.8.3 Option market impact of data breaches

As the majority of the literature agrees on the impact of data breaches on the targets' short-term valuation, it is possible for investors to benefit financially by making a downside bet on the companies' underlying stock. While informed trading is often analyzed by looking at stock market activity, some findings have revealed that informed trading can also be present in the option market. Poteshman (2006) and Ge et al. (2014) find evidence

that in the run-up to negative events option trading increases. In line with these papers, Mitts and Talley (2018) find evidence that in the two months before the announcement of a data breach the open interest and volume of put options increases. This brings me to the fifth and last hypothesis:

H5: The open interest and volume of put options increases in the two month period ahead of the announcement of a data breach

3. Data

This section discusses the data used in this study together with its sources. Section 3.1 describes the sample construction and section 3.2 describes the descriptive statistics.

3.1 Sample selection

To construct the sample of data breaches, 9.041 data breaches from the Privacy Rights Clearinghouse database (PRC) are used as starting point. Governments, educational and non-profit organizations are removed from the sample, which leaves 6.558 data breaches within the United States, over the sample period 2005-2018. For data breaches to be incorporated in the sample, the data breach must be classified as: fraud involving payment cards (excludes the loss of information by hacking), hacking or malware, an insider intentionally breaches information, physical loss of information, lost portable devices and lost stationary devices. As a result, 5.235 data breaches remain. All privately held companies must be excluded, as the research will focus on stock price reactions. To find stock prices, company names are manually matched with names recorded in the Center for Research in Security Prices (CRSP). Furthermore, when firms have multiple breaches in one year only the first breach is kept in the sample. This leaves a final sample of 313 breaches for 216 different firms. Of these 313 breaches, 152 breaches (48%) were initiated by an external party by hack or malware. See Table I for a complete overview of the breaches. Table A in the Appendix presents an overview of the data breaches per industry based on the two-digit SIC codes.

Table I: Overview of data breaches

Type of data breach	Number of data breaches	Percentage of total
Hacking or Malware	152	49%
Credit card fraud	10	3%
Insider breaches information	53	17%
Physical loss of information	25	8%
Loss of portable devices	63	20%
Loss of stationary devices	10	3%
Total	**313**	**100%**

Secondly, daily information on options in the period 2005-2017 is obtained from Optionmetrics, as 2017 is the last year available in the Optionmetrics database. The sample of 313 data breaches, obtained from the event

study on stock prices, is used. However, due to missing data the final sample is reduced to 259 events for 174 distinct firms.

Thirdly, yearly firm-specific and industry-specific information is obtained from Compustat. Data is obtained from 2004-2017, as the yearly information is needed one year before the data breach.

3.2 Descriptive statistics

To investigate whether certain firm-specific and industry-specific characteristics have an influence on the likelihood of experiencing a data breach, several variables are incorporated.

The variables that measure a firm's visibility are inspired by those used in the research conducted by Kamiya et al. (2018). The natural logarithm of total assets is used as a proxy for firm size. The natural logarithm of total assets is chosen to reduce the impact of outliers. Since the firms in this sample are active in different industries, the absolute value of total assets could differ enormously. Prior studies do not agree whether visibility, when measured by firm size, has a positive or negative influence on the likelihood of experiencing a data breach. Therefore, institutional block ownership is also added as an indicator for firm visibility. This is measured as the total percentage of outstanding shares held by investors who own more than 5% of the shares. Lastly, a dummy variable is added that takes on the value of one for firms that are included on the list of Fortune 500 companies and zero if not.

In addition, several factors that could influence the relation between firm visibility and data breaches are controlled for. Return on assets is used to control for firm performance, which is in line with the study of Akey et al. (2018). Additionally, Sales growth is added as a control variable for firm performance.

The Tobin's q is incorporated to measure a firm's growth opportunities. Kamiya et al. (2018) find that firms with more growth opportunities are more likely to experience a data breach. Other studies, such as Lending et al. (2018) use the M/B ratio as a proxy for growth opportunities. They find contradicting results, and find a negative relation between the M/B ratio and the likelihood of experiencing a data breach. As Tobin's q and the M/B ratio are both proxies for growth opportunities, I solely include Tobin's q.

Furthermore, Leverage is included to measure whether the level of debt impacts the likelihood of experiencing a data breach. This is in line with the study of Lending et al. (2018), who find a positive relation between the debt level of companies and the likelihood of experiencing a data breach.

In addition, Research and development expenses, Capital expenses, the natural logarithm of a Firm's age and Asset intangibility are added as control variables to measure the impact of those variables on the likelihood of experiencing a data breach. According to the study of Kamiya et al. (2018), Asset intangibility and the likelihood of experiencing a data breach are positively related. This means that firms with fewer tangible assets have a

higher chance of experiencing a data breach. Additionally, they find that younger firms are more likely to experience a data breach, than older firms.

Furthermore, to measure the impact of certain industry-specific characteristics, three extra variables are added. The Industry Herfindahl index is added to capture industry competitiveness. The Industry Herfindahl index is measured as the sum of squared market shares based on the two digit SIC code. A dummy is added to account for industry uniqueness. The dummy takes the value of one if a firm's ratio of selling expenses divided by sales is in top quartile of the industry based on the two-digit SIC code, and zero otherwise. Lastly, the Industry Tobin's q is added to measure the impact of industry growth opportunities on firms experiencing a data breach. Industry Tobin's q is measured by the median Tobin's q of all firms in the same industry based on the two digit SIC code.

Year- and industry-fixed effects are included in both the probit and OLS regressions. The year-fixed effects are included to capture macro-economic trends over time. The industry-fixed effects are included to capture industry specific trends. See Table II for the descriptive statistics of all explanatory variables. Two robustness checks are performed to test for multicollinearity and heteroscedasticity of the variables.

Table B in the Appendix depicts the Pearson correlation matrix for all variables included in this study. Results of the Pearson correlation test that are higher than 0.8 are indicative of multicollinearity. However, as can be seen in the Table, there is no evidence of multicollinearity.

Table C in the Appendix reports the Breusch-Pagan / Cook Weisburg test for heteroscedasticity. As can be seen in the Table, there is evidence of heteroscedasticity. Therefore, to prevent the coefficients to be biased, the standard errors are adjusted for heteroscedasticity at the firm level. Consequently, the OLS regressions are performed with robust standard errors.

Table II: Descriptive statistics

This Table presents the descriptive statistics from the characteristics of target and control firms used in this study. The Table reports statistics for 313 firm-year observations that are the victim of a data breach in the following fiscal year. The remaining 56.836 firm-year observations do not experience a data breach in the following year. Data is comprised from Compustat in the period 2005-2017. Differences in means and medians between firms that experience a data breach and firms that do not are measured by t-tests and Wilcoxon signed-rank tests.

	Firm-years without data breach (N=56.836): A		Firm-years followed by data breach (N=313): B		Test of difference (A-B)	
	Mean	Median	Mean	Median	Mean	Median
Firm size	7.635	0.789	53.738	14.783	-46.103***	-13.994***
Firm age	19.503	15.000	31.387	27.000	-11.884***	-12.000***
Block ownership	0.131	0.066	0.143	0.113	-0.012	-0.047**
Fortune 500 (indicator)	0.147	0.042	0.655	1.000	-0.508***	-0.958***
Tobin's q	1.952	1.415	1.933	1.529	0.019	-0.114**
ROA	-0.018	0.024	0.050	0.042	-0.068***	-0.018***
Sales growth	0.158	0.072	0.091	0.056	0.067**	0.016
Leverage	0.213	0.159	0.251	0.221	-0.038***	-0.062***
R&D / assets	0.044	0.000	0.020	0.000	0.024***	0.000***
CAPX / assets	0.042	0.023	0.035	0.026	0.007***	-0.003
Asset intangibility	0.767	0.876	0.811	0.886	-0.044***	-0.010
Industry Herfindahl index	0.071	0.042	0.093	0.048	-0.022***	0.006***
Unique industry (indicator)	0.528	1.000	0.537	1.000	-0.009	0.000
Industry Tobin's q	1.572	1.499	1.503	1.485	0.069***	0.014**

*** p<0.01, ** p<0.05, * p<0.1

4. Methodology

This section discusses the empirical methods used in this study. Section 4.1 extensively explains the probit model used in this study. Section 4.2 discusses the event study used in this study to find any abnormal trading in the stock market, 4.3 debates the OLS regression that is used to find the determinants of the stock price reaction and section 4.4 discusses the event study used to find any abnormal trading in the option market.

4.1 Likelihood of becoming the victim of a data breach

Firstly, to search for firm- and industry-specific characteristics that influence the likelihood of firms becoming the victim of a data breach, targets will be compared with companies that did not experience a data breach. One cautionary note: it could be that firms included in the comparable group did experience a data breach, but did not disclose the breach. Non-targets are retrieved from Compustat and are listed on the New York Stock Exchange or the Nasdaq in the period 2004-2017. Firstly, to compare firm- and industry-specific characteristics of targets and non-targets two-sample t-tests are performed to compare the means of both groups. Secondly, Wilcoxon signed-rank tests are performed to compare the medians.

To further test the likelihood of firms becoming the victim of a data breach, a binary model will be used. In a linear model, the probability can become negative or larger than one, therefore a non-linear estimation method will be performed. Both a probit and a logit model are suitable for this analysis. The difference between the models lies in the assumption about the distribution of the error terms. A probit model assumes a normal distribution, whereas a logit model employs a logistic distribution. Preference for one model over the other varies per discipline (Horowitz and Savin, 2001). In line with the study of Kamiya et al. (2018), a probit model is used to test which firms are more likely to be targeted. In a probit model the dependent variable takes the value of one if a firm experiences a data breach and zero if not. I will analyze how multiple explanatory variables that are measured one year before the breach influence the probability that the dependent variable equals one. In the regression, firm- as well as industry-specific characteristics will be inserted as explanatory and control variables. Also, a control variable for year- and industry-fixed effects (based on two-digit SIC code) will be included.

In model 1, the indicators for firm visibility are added. Firm visibility is measured by Firm size, Institutional block ownership and presence on the list of Fortune 500 companies.

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \varepsilon i \quad (1)$$

In model 2, several firm-specific characteristics are added as control variables. In model 3, year-fixed effects are added to capture any macroeconomic variation that happens over time. In model 4, industry-fixed effects are added to capture any variation that occurs between industries.

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \beta_4 Tobin'sq + \beta_5 ROA + \beta_6 \log(Firm\ age) + \beta_7 Salesgrowth + \beta_8 Leverage + \beta_9 \frac{R\&D}{Total\ Assets} + \beta_{10} \frac{CAPX}{Total\ Assets} + \beta_{11} Assetintangibility + \varepsilon i \quad (2)$$

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \beta_4 Tobin'sq + \beta_5 ROA + \beta_6 \log(Firm\ age) + \beta_7 Salesgrowth + \beta_8 Leverage + \beta_9 \frac{R\&D}{Total\ Assets} + \beta_{10} \frac{CAPX}{Total\ Assets} + \beta_{11} Assetintangibility + \beta_{12} Year + \varepsilon i \quad (3)$$

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \beta_4 Tobin'sq + \beta_5 ROA + \beta_6 \log(Firm\ age) + \beta_7 Salesgrowth + \beta_8 Leverage + \beta_9 \frac{R\&D}{Total\ Assets} + \beta_{10} \frac{CAPX}{Total\ Assets} + \beta_{11} Assetintangibility + \beta_{12} Year + \beta_{13} Industry + \varepsilon i \quad (4)$$

In model 5, industry-specific characteristics are added to investigate whether industry competition, uniqueness and growth opportunities have an impact on the likelihood of becoming the victim of a data breach. In this regression the industry-fixed effects are omitted, but the year-fixed effects are kept in the regression.

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \beta_4 Tobin'sq + \beta_5 ROA + \beta_6 \log(Firm\ age) + \beta_7 Salesgrowth + \beta_8 Leverage + \beta_9 \frac{R\&D}{Total\ Assets} + \beta_{10} \frac{CAPX}{Total\ Assets} + \beta_{11} Assetintangibility + \beta_{12} Uniqueindustry + \beta_{13} Herfindahlindex + \beta_{14} Industrytobin's\ q + \beta_{15} Year + + \varepsilon i \quad (5)$$

In model 6, industry indicators are added as dummy variables based on the two-digit SIC code. The manufacturing industry is omitted, as it serves as the reference group. This regression is used to identify whether companies in certain industries are more likely to become the target of a data breach.

$$P(D_{data\ breach} = 1) = \propto_0 + \beta_1 \log(Firm\ size) + \beta_2 Blockownership + \beta_3 Fortune500 + \beta_4 Tobin'sq + \beta_5 ROA + \beta_6 \log(Firm\ age) + \beta_7 Salesgrowth + \beta_8 Leverage + \beta_9 \frac{R\&D}{Total\ Assets} + \beta_{10} \frac{CAPX}{Total\ Assets} + \beta_{11} Assetintangibility + \beta_{12} Electric\&gas\&sanitary + \beta_{13} Finance + \beta_{14} Mineral\&construction + \beta_{15} Publicadministration + \beta_{16} Service + \beta_{17} Transport\&communication + \beta_{18} Wholesal\&retail + \beta_{19} Year + \beta_{20} Industry + \varepsilon i \quad (6)$$

4.2 Impact of data breaches on shareholder wealth

Based on the methodology of Mackinlay (1997) and Brown and Warner (1985) an event study is performed to measure the impact of a data breach on a target's value. Event studies have been widely used for analyzing the impact of a corporate event on firm value. According to the efficient market hypothesis, a sudden release of information will immediately be reflected in the share price (Fama, 1970). Therefore, the change in a stock price is a good proxy for the impact of a data breach. To avoid uncertainty about the announcement date, Factiva and the PRC database are used to cross-check the date when the data breach was disclosed for the first time.

A measure of the abnormal return is required to estimate the impact of a data breach. The normal return is defined as the expected return, conditional on that the event did not occur. The abnormal return is calculated as:

$$AR_{i\tau} = R_{i\tau} - E(R_{i\tau} \mid X_{i\tau}) \quad (7)$$

$AR_{i\tau}$ is the abnormal return, $R_{i\tau}$ the actual return and $E(R_{i\tau} \mid X_{i\tau})$ is the normal return. To determine the normal return, a market model as well as the Fama-French three-factor model are used. The expected return can be calculated in several ways. For example, by using the market model, the Fama-French three factor model and the mean adjusted model. The market model compares the event window returns with an expected market return over the same event window. The mean adjusted model compares the returns with the mean market return over an estimation period. This study focuses on the market model, as this model assumes a linear relationship between the market return and the stock return. In addition, the three-factor model is used to seek for any differences in the outcomes of the model used.
The market model is defined as:

$$AR_{i\tau} = R_{i\tau} - \alpha_i - \beta_i R_{m\tau} \quad (8)$$

Where $R_{i\tau}$ is the period- t return and $R_{m\tau}$ is the return on the market portfolio. For the $R_{m\tau}$ the CRSP equal weighted index is used and α_i and β_i are estimated over a pre-event estimation window. In line with the model of Akey et. al (2018), the estimation window that is used is set equal to 100 days: 150 days before until 50 days before the announcement. This is based on the fact that typically two months prior to the announcement, the event could already be reflected in the stock price.

The Fama-French three-factor model is defined as:

$$AR_{i\tau} = R_{i\tau} - Rf_{\tau} + \beta_1(R_{m\tau} - Rf_{\tau}) + \beta_2(SMB_{\tau}) + \beta_3(HML_{\tau}) + \varepsilon_{i\tau} \quad (9)$$

Where $R_{i\tau}$ is the period- t return and Rf_{τ} is the risk-free rate of the return at time t. $R_{m\tau}$ is the return on the market portfolio. For the $R_{m\tau}$ the CRSP equal weighted index is used. α_i and β_i are estimated over a pre-event estimation window, which is the same period as was used for the market model. The SMB (small minus big) factor is an indicator of firm size based on market capitalization. Fama and French (1993) noticed that for smaller companies the returns were on average higher than for larger companies, therefore they include the size premium to the model. The SMB factor is derived by subtracting the equal-weighted average return on a large stock portfolio from the equal-weighted average return on a small stock portfolio. They also discovered that on average firms with high M/B ratios have lower returns than firms with lower M/B ratios. This finding suggests that value stocks have higher returns than growth stocks. Therefore, they also include the HML factor (high minus low) to the model to take into account this difference. This factor can be seen as a value premium.

The HML factor is computed as the equal-weighted average returns for the low M/B portfolio minus the equal-weighted average returns of the high M/B ratio portfolio.

Once the output of the market model and the three-factor model for all individual firms is obtained, the average daily abnormal stock returns of all individual firms are aggregated to obtain the sample aggregated abnormal returns for a specific day:

$$AAR_{\tau} = \frac{1}{N}\sum_{i=1}^{N} AR_{i,\tau} \quad (10)$$

Where AAR_{τ} is the average abnormal return on day t and N is the number of firms in the sample. The average abnormal returns can then be aggregated to obtain the cumulative abnormal return (CAAR) from the day(s) before the announcement until the day(s) after the announcement.

$$CAAR_i(\tau 1,\tau 2) = \sum_{\tau 1}^{\tau 2} AAR_{\tau} \quad (11)$$

The CAAR's are calculated for different event windows. As a start, a one-day (day 0) event window is used to measure the announcement effect. Thereafter, other event windows are studied to account for potential delayed stock price reactions. T-tests are performed to test the significance of the CAAR's.

$$tCAAR = \frac{CAAR(\tau 1,\tau 2)}{S} \quad (12)$$

The standard deviation is calculated using the following formula:

$$S = \sqrt{\frac{1}{N(N-d)}\sum_{i=1}^{N}(CARi - CAAR)^2} \quad (13)$$

Where CAR_i is the cumulative abnormal trading returns per individual firm and d is the degrees of freedom. In this study the words 'mean CAR' and 'CAAR' are used interchangeably.

When performing a t-test, it assumed that the abnormal returns are normally distributed. Furthermore, the returns must be independently distributed. As the events in my sample overlap, it could be that the stocks are correlated. This violates the cross-sectional independency assumption. Lastly, it assumes that the abnormal returns are identically distributed. However, as the volatility of stocks differ, this indicates that the variances of the abnormal returns possibly differ and the assumption does not hold. To test the robustness of the t-test, I also perform a non-parametric test, the Corrado rank test (1989). According to Campbell and Wasley (1996), a non-parametric test is preferred over the parametric test as it does not assume a normal distribution. Furthermore it is robust against event-induced volatility and cross-correlation. I compare the single day rank statistics with the statistics of the t-test. A disadvantage of the non-parametric test is the calculation of CARs with event-windows of multiple days. Although, Campbell and Wasley (1996) invented a new version in which the CARs for multiple days can be calculated, I compare the results based on single days.

To perform a rank test, I transform the abnormal returns into ranks. The formula is given by:

$$K_{i,t} = \frac{\text{rank}(ARi,t)}{1+Mi+Li} \quad (14)$$

Where, M_i is the number of values and L_i is the number of matched returns in the event window. To compute the Corrado t-statistic the following formula must be used:

$$Trank_{,t} = \frac{Kt-0.5}{Sk} \quad (15)$$

Where $K_\tau = \frac{1}{Nt}\sum_{i=1,}^{N} K_{i,\tau}$ and Kt is the average rank for all companies. The standard deviation is calculated by:

$$Sk = \sqrt{\frac{1}{L1+L2}\sum_{t=T0}^{T2}\frac{Nt}{N}(Kt-0.5)^2} \quad (16)$$

4.3 Event study on trading Volumes

Next, an event study is performed on trading volumes to analyze the buying and selling behavior of investors surrounding the announcement of data breaches. The methodology is based on the paper of Campbell and Wasley (1996). By looking into trading behavior, insights could be obtained on whether informed trading takes place in the stock market. Several event windows are used prior to the announcement of data breaches. The benchmark for normal trading volume is based on the market model as well as on the mean-adjusted model. The market model is somewhat better than the mean-adjusted model in discovering abnormal stock volumes, as the model has a higher power than the mean-adjusted model. The traded volume is transformed using a natural logarithm. It is preferred to use the logarithmic transformed variable, as the transformed variable shows more symmetry in predicting the error terms (Ajinkaya and Jain, 1989). The small constant of 0.000255 is added before the log transformation, to overcome problems associated with taking the log of a zero trading volume.

The log-transformed trading volume is computed as follows:

$$V_{i,\tau} = \ln\left(\frac{Ni,\tau}{Si,\tau} * 100 + 0.000255\right) \quad (17)$$

Where $V_{i,\tau}$ is the log-transformed trading volume relative to the amount of shares outstanding on day t. Ln is the natural logarithm, Ni,τ is the stock volume traded on day t and Si,τ is the number of shares outstanding on day t.

The first benchmark used in this study is the mean-adjusted model. The estimation window is the same as for the stock returns and is (-150,-50). Normal trading volume is computed as follows:

$$NV_{i,\tau} = \frac{1}{101}\sum_{-150}^{-50} V_{i,s} \quad (18)$$

The results from the mean-adjusted model are compared with the results of the market model. For the market model, first all firms listed on a specific index are obtained. Firms in this study are listed on the NYSE or the NASDAQ. Secondly, to compute the market trading volume, the following formula is used:

$$V_{m,t} = \frac{1}{N}\sum_{i=1}^{N} V_{i,\tau} \quad (19)$$

Where $V_{m,t}$ is the market trading volume of a particular market index on day t and N is the number of firms included in the index.

Accordingly, normal trading volume is defined by the following formula:

$$NV_{i,\tau} = \alpha_i + \beta_i V_{m,t} + \varepsilon_{i\tau} \quad (20)$$

Where α_i is the intercept, β_i is the market beta and $\varepsilon_{i\tau}$ is the error term on day t.

Following the above-mentioned formulas, abnormal trading volume is computed with the following formula:

$$AV_{i,\tau} = V_{i,\tau} - NV_{i,\tau} \quad (21)$$

Once the output of the mean-adjusted or the market model is obtained, the average daily abnormal stock volumes of all individual firms are aggregated to obtain the sample aggregated abnormal volumes for a specific day:

$$AAV_{\tau} = \frac{1}{N}\sum_{i=1}^{N} AV_{i,\tau} \quad (22)$$

The average abnormal volumes over the event window can then be aggregated to obtain the cumulative average abnormal volume (CAAV) from the day(s) before the announcement until the day(s) after the announcement.

$$CAAV(\tau1,\tau2) = \sum_{\tau1}^{\tau2} AAV_{\tau} \quad (23)$$

Eventually, to test whether the cumulative average abnormal trading volumes are significantly different from zero, the following t-test is performed:

$$tCAAV = \frac{CAAV(\tau1,\tau2)}{S} \quad (24)$$

The standard deviation is calculated using the following formula:

$$S = \sqrt{\frac{1}{N(N-d)}\sum_{i=1}^{N}(CAVi - CAAV)^2} \quad (25)$$

Where CAV_i is the cumulative abnormal trading volume per individual firm. The words 'mean CAV' and 'CAAV' could be used interchangeably. In this study the CAVs are not investigated on an individual level, solely on an aggregated level. I assume that the abnormal trading volumes are normally distributed with mean zero. Moreover, the assumption is made that the distribution of abnormal trading volumes is independent and identical.

4.4 Determinants of the negative stock price reaction

To test the determinants of the stock price reaction, an OLS regression is performed. The dependent variable equals the CAR in the time window (-1,1) obtained from the market model. In model 1 (as defined in formula 26), dummies are included to measure whether the loss of financial information and whether a repeated breach has an impact on the magnitude of the stock price reaction.

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \varepsilon i \quad (26)$$

In model 2 (as defined in formula 27), several breach type indicators are added to test whether the nature of the breach has an impact on the magnitude of the stock price reaction.

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \beta_3 Creditcard + \beta_4 Hack + \beta_5 Insider + \beta_6 Portableloss + \beta_7 Stationaryloss + \varepsilon i \quad (27)$$

In model 3 (as defined in formula 28), several firm-specific characteristics are added as control variables. In model 4 (as defined in formula 29), year-fixed effects are added to capture any variation that happens over time. In model 5 (as defined in formula 30), industry-fixed effects are added to capture any variation that occurs between industries.

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \beta_3 Creditcard + \beta_4 Hack + \beta_5 Insider + \beta_6 Portableloss + \beta_7 Stationaryloss + \beta_8 \log(Firmsize) + \beta_9 Log(Firmage) + \beta_{10} ROA + B_{11} Leverage + \beta_{12} Salesgrowth + \beta_{13} Tobin'sq + \beta_{14} Blockownership + + \varepsilon i \quad (28)$$

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \beta_3 Creditcard + \beta_4 Hack + \beta_5 Insider + \beta_6 Portableloss + \beta_7 Stationaryloss + \beta_8 \log(Firmsize) + \beta_9 Log(Firmage) + \beta_{10} ROA + B_{11} Leverage + \beta_{12} Salesgrowth + \beta_{13} Tobin'sq + \beta_{14} Blockownership + B_{15} Year + \varepsilon i \quad (29)$$

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \beta_3 Creditcard + \beta_4 Hack + \beta_5 Insider + \beta_6 Portableloss + \beta_7 Stationaryloss + \beta_8 \log(Firmsize) + \beta_9 Log(Firmage) + \beta_{10} ROA + B_{11} Leverage + \beta_{12} Salesgrowth + \beta_{13} Tobin'sq + \beta_{14} Blockownership + B_{15} Year + B_{16} Industry + \varepsilon i \quad (30)$$

Lastly, in model 6 (as defined in formula 31), industry-specific characteristics are added to investigate whether industry competition, industry uniqueness and growth opportunities have an impact on the magnitude of the

stock price reaction. In this regression the industry-fixed effects are omitted, but the year-fixed effects are kept.

$$Y\ (CAR\ data\ breach) = \propto_0 + \beta_1 Financialloss + \beta_2 Repeatedbreach + \beta_3 Creditcard + \beta_4 Hack\ + \beta_5 Insider\ + \beta_6 Portableloss\ + \beta_7 Stationaryloss + \beta_8 \log(Firmsize) + \beta_9 Log(Firmage) + \beta_{10} ROA + B_{11} Leverage + \beta_{12} Salesgrowth + \beta_{13} Tobin'sq + \beta_{14} Blockownership + \beta_{15} Uniqueindustry + \beta_{16} Herfindahlindex + \beta_{17} Industrytobin's\ q + B_{18} Year + +\varepsilon i \quad (31)$$

4.5 The effect of data breaches on option trading

To examine how data breaches affect volume and open interest of option trading an event study is performed based on the mean-adjusted model. A benchmark period of (-150,-50) is used. This is similar to the event study on equity returns and volumes. In line with the study of Jayaraman, Frye and Sabherwal, (2001), the logarithmic transformation is applied to the option volume and the open interest. The reason for this transformation is that the amount of options traded on a daily basis varies enormously. The number of options traded on a daily basis is defined as $V_{i,\tau}$: ln (1+ number of options traded per day). The same transformation is used for open interest. For both the traded volume and the open interest the total group is divided into call and put options. One would expect the volume and open interest of put options to increase. However, it is also possible to perform a profitable strategy by selling call options prior to the announcement of a data breach (Jayaraman, Frye and Sabherwal, 2001). Therefore, this study investigates the effect of data breaches on both call and put options. The same procedure as for trading volume is used to obtain abnormal option volume and open interest, and a t-test is used to test for significance.

5. Results

This section explains the empirical results obtained in this study. Section 5.1 discusses the empirical results obtained from the probit model to address which firm- and industry-specific characteristics are associated with firms becoming the target of a data breach. Section 5.2 examines the results of the event study on equity trading. Section 5.3 looks into the firm-specific characteristics that influence the magnitude of the stock price reaction. In addition, section 5.4 discusses the effects of data breaches on option trading.

5.1 Likelihood of experiencing a data breach

Firstly, to compare firms that experienced a data breach and firms that did not, the firm- and industry-specific characteristics of both groups are studied. Table II shows an overview of the firm- and industry-specific characteristics of these so-called targets and non-targets of data breaches. When a firm experiences more than one data breach in a given year, all breaches are taken as one. In those cases, the first data breach is kept in the sample. This results in 313 data breaches, for 216 unique companies. The characteristics of the targets are compared with 56.835 firm year observations of non-targets. To account for outliers, all continuous variables are winsorized at the 1st and 99th percentiles. This is also in line with the study of Kamiya et al. (2018).

To test the differences between both groups, t-tests are performed for the mean and Wilcoxon signed-rank tests for the median. The results of the mean and median differences tests suggest that the firm- and industry-specific characteristics of targets and non-targets differ significantly.

The results contradict the results of Akey et al. (2018) and Lending et al. (2018), which report that firms have a higher chance of experiencing a data breach when they are smaller and when they have fewer growth opportunities, measured by the M/B ratio. They mention that large firms invest more in the protection of the cyber environment, as large firms have a greater risk of being the target of a cyber attack than smaller companies. As large companies have more important information and more money, there is more to be gained from. The results of the two sample t-test indicate that firms that experience a data breach are larger (measured by total assets) and older. Furthermore, based on the median of those targets a higher percentage of shares outstanding is owned by institutional block holders. Lastly, a higher percentage of targets compared to non-targets is included on the list of Fortune 500 companies. These results indicate that for larger more visible companies, all the investments in protecting the cyber environment might not always result in the prevention of data breaches.

In addition, firms that experience a data breach have higher growth opportunities based on Tobin's q and a higher return on assets compared to non-targets. When comparing more firm-specific characteristics, targets have more leverage, however, they invest less in research and development and spend less on capital expenditures. Targets also have fewer tangible assets compared to non-targets. Focusing on industry-specific characteristics, firms that experience a data breach are concentrated in industries that are less competitive (measured by Industry Herfindahl index), but also in industries that have fewer growth opportunities.

In general, based on the results of the two-sample t-test, firms that are more visible – as measured by Firm size, Institutional block ownership and presence on the list of Fortune 500 companies – are more likely to become the victim of a data breach.

To do a more in-depth analysis on whether certain characteristics influence the likelihood of firms experiencing a data breach, a probit regression is performed. For this probit regression, the same sample is used as for the two-sample t-test. Table IV presents the results from the probit regression, in which the dependent variable takes on the value of one in case a firm experiences a data breach and zero if not. Firm- and industry-specific characteristics are measured one year before the breach.

In model 1, several proxies for firm visibility are added. Firm visibility is measured by Firm size, Institutional block ownership and presence on the list of Fortune 500 companies. The results indicate that firms, which are more visible, are more likely to become the victim of a data breach. All three variables are positive and significant at the 1% level. To interpret the coefficients of the probit regression, the partial derivative is computed to get the marginal effect. The marginal effect of an increase in one unit of firm size is 0.13%, indicating that an increase with one unit of log(Firm size) increases the likelihood of firms experiencing a data breach with 0.13%. An increase of 1% institutional block ownership increases the likelihood with 0.37%. When firms are included on the list of Fortune 500 companies, they are 0.17% more likely to become the victim of a breach. All three coefficients are significant at the 1% level.

In model 2, several firm-specific variables are added as control variables. The magnitude of the coefficients of Firm size, Institutional block ownership and presence On the list of Fortune 500 are roughly the same and still significant at the 1% level. This regression also indicates that firms that have a higher valuation, measured by Tobin's q, have a higher chance of becoming the victim of a data breach. The variable is significant at the 1% level. In addition, older firms are more likely to experience a breach than younger firms and firms with fewer tangible assets are more likely to experience a data breach. Firm age and Asset intangibility are both significant at the 1% level.

In model 3, year-fixed effects are included to capture any variation in the outcomes that happen over time. The magnitude and the direction of the independent variables that measure firm visibility are still significant at the 1% level. In addition, the control variables Tobin's q, Firm age and Asset intangibility are still positive and significant at the 1% level.

Model 4 includes both year-fixed and industry-fixed effects. The industry-fixed effect is included to capture any variation between industries. Firm size and Institutional block ownership are still positive and significant. However, Institutional block ownership is now only significant at the 5% level. The Fortune 500 indicator is no longer significant, suggesting that presence on the list of Fortune 500 companies is clustered within certain industries. Firm age is still significant at the 1% level and asset intangibility at the 5% level. This regression indicates that firms that spend more on R&D are more likely to become the victim of a data breach. The coefficient is significant at the 5% level. As can be seen in model 4, the industry-fixed effects have a large impact on the outcomes of the regression. The Fortune 500 indicator is no longer significant. This indicates that the

within firm effect is smaller than across industries. Furthermore, the Pseudo R-squared increases from 0.180 up to 0.241.

In model 5 an indicator for industry uniqueness, the Industry Herfindahl index, and the Industry Tobin's q are added, while the Industry-fixed effects are removed. Firm size, Institutional block ownership and the indicator for Fortune 500 companies are the same in magnitude and direction as in model 3. However, Firm size and Institutional block ownership are significant at the 1% level and the Fortune 500 indicator at the 5% level. Tobin's q, Firm age and Asset intangibility are all positive and significant at the 1% level. In addition, the indicator for industry uniqueness is positively significant at the 5% level. Indicating that firms which are active in an unique industry are more likely to experience a data breach. See Table G in the Appendix for a more detailed explanation of the variables. Furthermore, the Industry Herfindahl index is significantly positive at the 1% level, signifying that firms which are active in less competitive industries are more likely to experience a data breach. Based on these results, firms that experience a data breach are active in industries that face less product competition.

In model 6, the industry-specific characteristics are replaced by seven industry indicators, based on the two-digit SIC codes. Adding these indicators makes it possible to test whether firms operating in certain industries are more likely to experience a data breach. In this regression the Manufacturing industry is omitted, as it serves as the reference group. The results indicate that firms operating in the Finance industry, Service industries, Transport and Communication industries and Wholesale and Retail industries are more likely to experience a data breach. All those industry indicators are positive and significant. Service industries and Wholesale and Retail are significant at the 1% level and Finance and Transport and Communication at the 5% level. Firms operating in the Mineral and Construction industry are less likely to become the victim of a data breach. The industry indicator is negative and significant at the 10% level. In this regression, Firm size and Institutional block ownership are still significant at the 1% level. Presence on the list of Fortune 500 companies is no longer significant, as it seems that presence on the list of Fortune 500 companies is clustered within certain industries. Tobin's q and Firm age are positive and significant at the 1% level. In addition, R&D/assets and Asset intangibility are significant at the 5% level.

Judging by the pseudo R-squared of the regressions, model 4 fits the data best. This indicates that there is substantial variation within industries that is not captured in the other models.

In general, according to the two-sample t-test and the probit regressions, Firm size has a positive impact on the likelihood of a firm experiencing a data breach. This is in contrast with the results of Akey et al. (2018) and Lending et al. (2018). Their studies indicate that smaller firms are more likely to become the victim of a data breach. However, the results are in line with the study of Camp and Lewis (2006) and Kamiya et al. (2018). These studies suggest that cyber attacks are more likely to occur at larger companies than at smaller companies.

The results of this study show that firms that have a higher percentage of institutional block ownership and that are included on the list of Fortune 500 companies, are more likely to experience a data breach.

These factors indicate that firms that are more visible are more likely to experience a data breach. Therefore the first hypothesis, stating that data breaches are more likely to occur at firms that are less visible, is rejected. In addition, firms that have a higher percentage of growth opportunities are more likely to experience a data breach. These results are contrary to the results of Akey et al. (2018), however, they are in line with the results of Kamiya et al. (2018). The coefficients for Firm age and Asset intangibility, which are positive and significant at the 1% level, are also in line with Kamiya et al. (2018).

Table IV : Likelihood of experiencing a data breach

The Table presents the results the probit regression. The dependent variable takes the value of one if a firm experiences a data breach and 0 if a firm does not experience a data breach. The sample consists of 57.146 firm year observations. Data is obtained from Compustat in the period 2005-2017. The standard errors are adjusted for heteroscedasticity and clustering at the firm level.

	Dependent variable = Data breach (indicator)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log (firm size)	0.195***	0.196***	0.204***	0.254***	0.209***	0.251***
	(0.020)	(0.023)	(0.024)	(0.029)	(0.025)	(0.028)
Block ownership	0.576***	0.470***	0.529***	0.445**	0.513***	0.447***
	(0.149)	(0.165)	(0.167)	(0.177)	(0.168)	(0.173)
Fortune500 (indicator)	0.272***	0.237***	0.226***	0.151	0.210**	0.122
	(0.076)	(0.081)	(0.084)	(0.097)	(0.090)	(0.093)
Tobin's q		0.091***	0.101***	0.067***	0.101***	0.076***
		(0.018)	(0.019)	(0.022)	(0.020)	(0.021)
ROA		0.365	0.302	0.248	0.263	0.354
		(0.263)	(0.262)	(0.313)	(0.272)	(0.315)
Log (firm age)		0.114***	0.121***	0.202***	0.124***	0.185***
		(0.038)	(0.038)	(0.046)	(0.040)	(0.044)
Sales growth		-0.063	-0.050	-0.011	-0.042	-0.029
		(0.069)	(0.067)	(0.075)	(0.069)	(0.076)
Leverage		0.090	0.132	0.052	0.134	0.156
		(0.156)	(0.156)	(0.172)	(0.161)	(0.151)
R&D / assets		-0.275	-0.275	1.442**	-0.221	1.175**
		(0.589)	(0.587)	(0.651)	(0.638)	(0.593)
CAPX / assets		0.892	0.774	0.358	0.699	0.656
		(0.704)	(0.731)	(0.868)	(0.720)	(0.790)
Asset intangibility		0.629***	0.606***	0.413**	0.557***	0.428**
		(0.184)	(0.191)	(0.208)	(0.191)	(0.193)
Unique industry (indicator)					0.160**	
					(0.069)	
Industry Herfindahl Index					1.215***	
					(0.326)	
Industry Tobins's q					-0.026	
					(0.074)	
Electric, gas and sanitary						-0.241
						(0.192)
Finance						0.272**
						(0.107)
Mineral and construction						-0.444*
						(0.236)
Public administration						-0.328
						(0.298)
Service industries						0.635***
						(0.086)
Transport and commun.						0.291**
						(0.147)
Wholesale and retail						0.803***
						(0.098)
_cons	-4.294***	-5.331***	-5.693***	-6.830***	-5.795***	-6.401***
	(0.158)	(0.267)	(0.308)	(0.486)	(0.331)	(0.365)
Obs.	57.052	56.857	56.857	51.039	56.857	56.857
Pseudo R-squared	0.151	0.169	0.180	0.241	0.186	0.224
Industry Dummy	N	N	N	Y	N	N
Year Dummy	N	N	Y	Y	Y	Y

5.1.1 Difference between different types of data breaches

An additional two-sample t-test is performed to examine if there are differences in firm- and industry-specific characteristics between two groups of data breaches. The first group includes: credit card fraud, an insider who breaches information, physical loss of documents, loss of portable devices and the loss of stationary devices. The second group consists of cyber attacks. The difference between both groups is tested to find out whether the results of the probit regressions are driven by one group in particular. The results in Table V, show that, compared to firms in group two, firms in group one are older (significant at the 10% level), have lower growth opportunities (1% level) measured by Tobin's q and a lower return on assets (5% level). According to the results, firms in group one have a higher presence on the list of Fortune 500 (5% level). Furthermore, capital expenditures, the unique industry indicator and industry Tobin's q are higher for group 2. The results are significant at the 10% level.

Table V : Difference between different types of data breaches

This Table presents the descriptive statistics from the characteristics of Cyberattack and other breaches used in this study. The Table reports statistics for 313 firm-year observations that are the victim of a data breach in the following fiscal year. Data is comprised from Compustat in the period 2005-2017. Differences in means and medians between firms that experience a data breach are measured by t-tests and Wilcoxon signed-rank tests.

	Data breaches (excl. cyberattack) (N=161): A		Cyberattack (N=152): B		Test of difference (A-B)	
	Mean	Median	Mean	Median	Mean	Median
Firm size	55.419	20.600	51.957	11.668	3.462	8.932
Firm age	33.050	30	29.625	25	3.425*	5.000*
Tobin's q	1.740	1.435	2.137	1.647	-0.398***	-0.212***
ROA	0.041	0.041	0.059	0.048	-0.018**	-0.007
Sales growth	0.102	0.070	0.079	0.049	0.024	0.021*
Leverage	0.255	0.224	0.247	0.209	0.008	0.15
R&D / assets	0.019	0.000	0.020	0.000	-0.002	0.000
CAPX / assets	0.032	0.024	0.039	0.030	-0.007*	-0.006*
Asset intangibility	0.810	0.890	0.811	0.883	-0.001	0.007
Block ownership	0.138	0.109	0.148	0.122	-0.010	-0.013
Fortune 500 (indicator)	0.708	1.000	0.599	1.000	0.109**	0.000**
Industry Herfindahl index	0.089	0.043	0.097	0.056	-0.008	-0.013
Unique industry (indicator)	0.484	0.000	0.586	1.000	-0.101*	-1.000*
Industry Tobin's q	1.462	1.464	1.546	1.519	-0.085*	-0.055**

*** p<0.01, ** p<0.05, * p<0.1

5.2 Event study on equity trading

In this paragraph, the results of the event study on stock price returns around the announcement of data breaches are presented. An event study with different event windows is performed and comparisons of the mean CARs of several sub-samples are made to find out if there are any differences in stock price reactions between groups.

5.2.1 Event study on the complete sample

In this paragraph, the results of the event study on stock price returns around the announcement of data breaches are presented. The sample consists of 313 events (216 unique firms) in the period 2005-2018. Table VI reports the mean CARs for different event windows. As shown in Table VI panel A, the mean CARs (-1,1), (-2,2), (-3,3) and (-5,5) are significantly negative affected in both the market model and the Fama-French three factor model, suggesting that the results are not dependent on the model chosen. The results are -0.52%, -0.56%, -0.66% and -0.58% for the market model and -0.44%, -0.56%, -0.74% and -0.75% for the Fama-French three factor. The decline in stock prices in the above-mentioned event windows are significant at the 1% level, except for the (-5,5) event window, which is significant at the 5% level. The results are in line with prior studies, which show that data breaches have a significantly short-term negative impact on stock prices. This indicates that investors do not already take data breaches into account when valuing a company before they buy the particular stock.

Furthermore, a Corrado rank test is performed to test the robustness of the event study on trading returns. As can be seen in Table D in the Appendix, the abnormal returns on the T=-2, T=0 and T=1 are significantly negative for both the t-test and the Corrado Rank test. At the day of the announcement and the day after the announcement based on the t-test the abnormal returns are significant at the 10% and 5% level respectively, whereas with the Corrado rank test the abnormal returns are significant at the 5% and 1% level. This shows that with the Corrado rank test, the significance levels are even higher compared to the t-test.

Table VI : Event study on trading returns and volume

This Table reports the results of the event study on the test on significance of the cumulative abnormal returns (CARs) around data breach dates. The sample consists of 313 events (216 unique firms) in the period 2005-2017. The abnormal returns are calculated based on the market model and the Fama-French three factor model. To calculate the abnormal returns, the CRSP equally weighted index as an indicator for normal returns is used. The estimation window is 100 days, 150 days before until 50 days before the announcement. Panel A: presents the mean cumulative abnormal returns for target firms. Panel B: presents the results of an event study on trading volumes. Panel C: presents the results of a difference test between data breaches that included the loss of financial information and data breaches that did not include financial information. Panel D: presents the results of a difference test between cyberattacks and other types of data breaches. Other data breaches include: Payment card fraud, Insider who stole information, Physical loss of documents, Lost or stolen portable device and lost or stolen stationary device.

Panel A. Event study on target firms

		Market model		Three factor model	
Event window (%)	**Sample size**	**Mean**	**P-value**	**Mean**	**P-value**
CAR (-1,1)	313	-0.521	0.001***	-0.435	0.006***
CAR (-2,2)	313	-0.564	0.007***	-0.562	0.007***
CAR (-3,3)	313	-0.660	0.008***	-0.735	0.005***
CAR (-5,5)	313	-0.578	0.084*	-0.746	0.021**
CAR (-5,0)	313	-0.401	0.06*	-0.226	0.013**

Panel B. Event study on trading volume

		Market model		Mean-adjusted model	
Event window	**Sample size**	**Mean**	**P-value**	**Mean**	**P-value**
CAV (-40,-1)	290	-0.248	0.267	-0.014	0.955
CAV (-30,-1)	290	-0.146	0.431	0.025	0.898
CAV (-20,-1)	290	-0.179	0.202	-0.067	0.641
CAV (-10,-1)	290	-0.153	0.057*	-0.103	0.194

Panel C. Comparison of CARs between data breaches with and without financial information loss

	Financial information loss (N= 206)		No financial information loss (N= 107)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.672	0.001***	-0.214	0.352	1.384	0.167
CAR (-2,2)	-0.628	0.013**	-0.412	0.262	0.619	0.499
CAR (-5,5)	-0.286	0.493	-1.058	0.060*	-1.099	0.273

Panel D. Comparison of CARs between data breaches that are a cyberattack and other data breaches

	Cyberattack (N=152)		Other data breach (N=161)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.732	0.002***	-0.308	0.062*	1.349	0.178
CAR (-2,2)	-0.655	0.026**	-0.457	0.049**	0.480	0.632
CAR (-5,5)	-1.253	0.012**	0.116	0.610	2.062**	0.040**

5.2.2 Informed trading

Given the results in Table VI panel A, which show that the stock prices of firms that experience a breach decrease significantly, it could be profitable for investors to anticipate on this and take on short positions or sell the target firm's shares. To investigate whether informed trading plays a role around the announcement of data breaches, trading volumes for longer periods prior to the event are analyzed. The event windows used for this analysis are (-10,-1), (-20,-1). (-30,-1) and (-40,-1). As can be seen in Table VI panel B, none of these tests are significant. This indicates that investors do not anticipate on inside knowledge and do not trade accordingly prior to the announcement of a data breach. This is in line with the studies of (Gatzlaff and McCullough, 2008; Johnson et al., 2017 and Kamiya et al., 2018). Their studies show no linkage between data breaches and trading activity in the run-up to the announcement. As depicted in Figure I graph II, there is no significant increase in trading volume surrounding the announcement of data breaches. Graph I and II of figure I are based on the market model, however the results of the mean-adjusted model show a similar pattern. The study of Karpoff (1987), indicates that the correlation between volume and a price decrease is negative. After the announcement of a data breach one would expect to find an increase in trading volume, which causes the drop in stock prices. Table VII gives an overview of the trading volume on a daily basis in the event window (-3,3). Surprisingly, there is no significant increase in trading volume surrounding the announcement of data breaches. Most of the coefficients are even negative surrounding the announcement of a data breach. This indicates that in general, based on trading volume the market underreacts to the announcement of a data breach. In his study, Karpoff (1987) also mentions that the large cost of short sales could potentially provide an explanation for why trading volume increases more for price increases than for price decreases of the same amount. Ying (1966) also finds that a low trading volume is followed by a fall in price, while a large increase in volume can be accompanied by either a large increase or decrease of a company's stock price. In Figure II graph I and II, the abnormal trading returns of the top 25 largest price drops in the (-3,3) event window is compared with the accompanied trading volumes. Returns and volumes are calculated with the market model. The results reveal a significantly positive increase in trading volume at the 10% level. This indicates that only for large price drops, trading volume increases significantly.

Figure I: Cumulative Abnormal Returns and Cumulative Abnormal Volumes

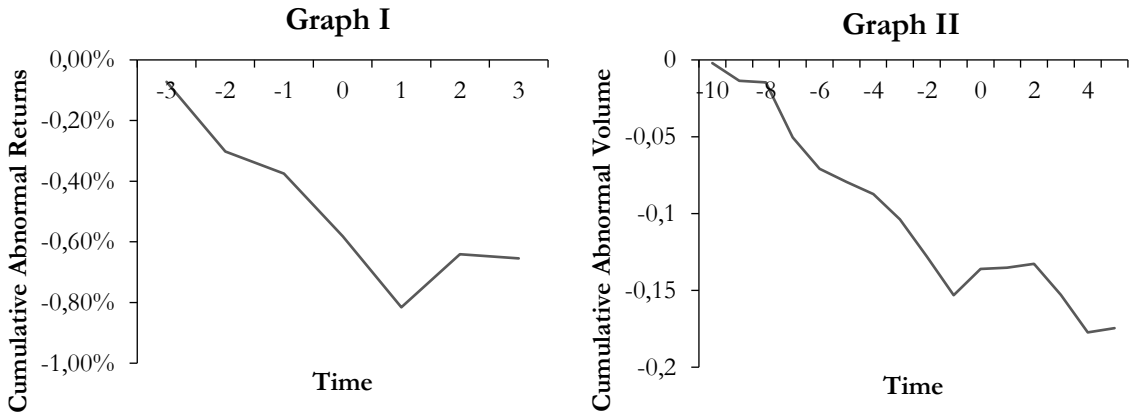
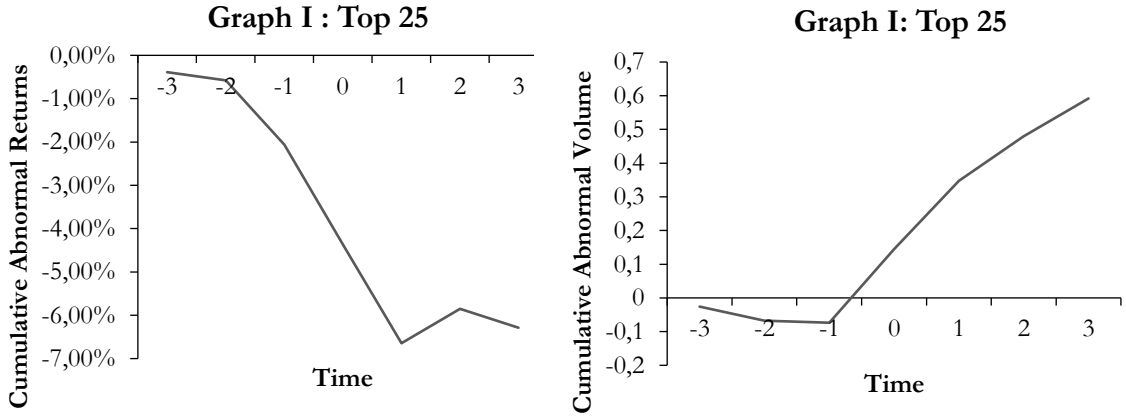


Table VII : Event study on trading volume – daily trading volume

		Market model		Mean-adjusted model	
Event days	**Sample size**	**Mean**	**P-value**	**Mean**	**P-value**
-3	290	-0.016	0.181	-0.013	0.286
-2	290	-0.024	0.070*	-0.020	0.122
-1	290	-0.025	0.047**	-0.022	0.087*
0	290	0.017	0.294	0.017	0.295
1	290	0.009	0.951	0.004	0.787
2	290	0.002	0.876	0.005	0.730
3	290	-0.020	0.208	-0.019	0.233

Figure II: Cumulative Abnormal Returns and Cumulative Abnormal Volumes top 25



5.2.3 Event study on data breaches that involve the loss of financial information

In panel C of Table VI, the sample is divided into data breaches that involve the loss of financial information and data breaches that did not involve the loss of financial information. The difference in stock price reactions for the (-1,1), (-2,2) and (-5,5) event windows is tested. As can be seen in the Table, the stock price reactions of data breaches that involve the loss of financial information are highly significant in the (-1,1) and (-2,2) event windows, with corresponding mean CARs of -0.67% and -0.63%, respectively. Data breaches that did not involve the loss of financial information were only significant in the (-5,5) event window, however, only at the 10% level. The mean CAR (-5,5) is not significant for firms that experienced a loss of financial information. The t-tests, used to test whether the mean CARs of both groups differ, were not significant.

5.2.4 Event study on cyber attacks compared to other data breaches

In panel D of Table VI, the sample of total data breaches is divided into a group that experienced a cyber attack and a group that experienced other data breaches such as: credit card fraud, an insider who breaches information, physical loss of documents, loss of portable devices and the loss of stationary devices experience. The difference in stock price reactions is tested for the (-1,1), (-2,2) and (-5,5) event windows. As can be seen in the Table, the stock price reactions of cyber attacks are significant in the (-1,1), (-2,2) and (-5,5) event windows, with corresponding mean CARs of -0.73% ,-0.66% and -1.25%, respectively. The (-1,1) is significant

at the 1% level, whereas the (-2,2) and (-5,5) event window is significant at the 5% level. The mean CAR of the group other data breaches was only significant in the (-1,1) and (-2,2) event windows. The t-tests, to test whether the mean CARs of both groups in each event window differ, is only significant at the 5% level in the (-5,5) event window.

Although, the results of the t-tests are only significant in the (-5,5) event window, one can observe that the stock price reaction between both groups differs. In Table E in the Appendix, the complete sample is divided by type of breach: Hacking or malware (49%), Credit card fraud (3%), Insider breaches information (17%), Physical loss of documents (8%), Loss of portable devices (20%) and Loss of stationary devices (3%). As can be seen from the Table, only the sub-group Hacking or malware is significant. This indicates that the negative stock price reaction of the complete sample of data breaches is driven by cyber attacks.

In Table F in the Appendix, a t-test is performed to test whether the stock price reaction of the different types of breaches differs. While only the subgroup Hacking or Malware causes a significant stock price reaction, the difference between the mean CARs of cyber attacks and all other data breaches is not significant.

Overall, evidence is found to support the second hypothesis stating that abnormal returns are generated surrounding the announcement of data breaches. Therefore, in the short term, the announcement of data breaches is costly for firms that experience those breaches. If the stock price is expected to fall after the announcement, investors could potentially benefit by trading in advance of the breach. However, the results in this section do not show any evidence that informed trading takes place. Moreover, the additional results do not show any evidence that data breaches that involve the loss of financial information have a larger impact on a firm's stock price. The two-sample t-test to test the difference in stock price reactions is not significant. Furthermore, the total sample was divided into firms that experienced a cyber attack and firms that experienced another data breach. The difference test was only significant in the (-5,5) event window, indicating that cyber attacks have a larger impact on a company's stock price than other data breaches. To deliver a more in-depth analysis, an OLS regression is performed to seek for the determinants of the stock price reaction.

5.3 Determinants of the cumulative abnormal returns

In this section, the results of the ordinary least squares regression (OLS) are reported. This regression is used to analyze the determinants of the cumulative abnormal returns. The dependent variable is the mean CAR of the (-1,1) event window, obtained with the market model. OLS regressions (1), (2) and (3) are performed without industry- and year-fixed effects, in regression (4) year-fixed effects are added and in regression (5) both year-fixed and industry-fixed effects are added. In regression (6) the industry-fixed effects are replaced by several industry-specific variables. In all regressions, the standard errors are adjusted for heteroscedasticity.

In model 1, an indicator for the loss of financial information and an indicator for a repeated breach within one year are included. As can be seen in Table VIII, both coefficients are negative, but have no significant impact on the stock price reaction.

In model 2, dummies are added for the nature of the breach, in order to test whether the type of breach has an impact on the stock price reaction of a target. The sample is divided into the following groups: Credit card fraud, Hack or Malware, Insider breaches information, Physical loss of documents and Stationary device loss. None of the dummies is significant, indicating that the type of breach has no influence on the magnitude of the stock price reaction. In this regression, the indicator for loss of financial information becomes significant at the 10% level, indicating that a data breach that involves the loss of financial information has a significantly more negative stock price reaction.

In model 3, several independent variables are added as control variables: log (Firm size), log (Firm age), ROA, Leverage, Sales growth, Tobin's q and Institutional block ownership. As can be seen in the Table, Leverage has a significant and negative impact on the stock price reaction with a coefficient of -0.007. The coefficient is significant at the 10% level. In addition, Sales growth has a significant negative impact on the stock price reaction, with a coefficient of -0.018 and significant at the 5% level. This indicates that the stock prices of firms with a higher percentage of sales growth in the previous year react more negatively to a data breach. None of the other determinants are significant, which indicates that these have no impact on the magnitude of the reaction. This contradicts the results of Gatzlaff and McCullough (2008) and Cavusoglu et al. (2004). According to the first study, growth opportunities have a significantly negative influence on the magnitude of the stock price reaction, while firm size mitigates the negative impact on a company's wealth. In line with this finding, the second study finds that the market reaction is significantly higher for smaller firms than for larger companies. However, the results in this study offer no evidence to support these results.

Model 4 includes Year-fixed effects and shows that the indicator for the Loss of financial information is statistically significant at the 5% level. This indicates that the stock price of firms in which the data breach includes the loss of financial information declines with 0.7% more than of firms at which the data breach does not include the loss of financial information. Leverage is also significant at the 5% level, with a coefficient of -0.009. However since the Year-fixed effect is included, Sales growth is no longer significant, indicating that the difference in Sales growth could be driven by fluctuations in the economic cycle that happen over time.

Model 5 includes both Year-fixed effects and Industry-fixed effects, capturing both the variation over time and between industries. The indicator for the Loss of financial information is still significant, however in this regression it is only significant at the 10% level, with a negative coefficient of -0.007. In line with model 4, Leverage is negative and significant at the 5% level.

In model 6, the Industry-fixed effects are removed and the Industry Herfindahl index, an indicator for industry uniqueness and the industry Tobin's q are included. None of these variables are significant, which indicates that the stock price reaction is not impacted by industry competitiveness, uniqueness of the industry and the valuation of a given industry. In this regression the indicator for the loss of financial information is negative and significant at the 5% level. In line with regressions (4-5), leverage has a negative impact on the stock price reaction, with a coefficient of -0.011 and significant at the 5% level.

Analyzing the R-squared of all regressions, model 5 fits the data best. This indicates that there is substantial variation within industries that is not captured in the other regressions.

Overall, some evidence was found to support hypothesis three, which states that the impact on stock price reactions is larger for data breaches that involve the loss of financial information. Model 5 has the highest R-squared and in this regression the coefficient of financial loss is -0.007 and significant at the 10% level. This is in line with the results of Campbell et al., (2003), Hovav and D'Arcy (2003) and Kamiya et al. (2018). Their studies show that when a data breach includes valuable or confidential information, the stock price reaction is significantly higher.

In addition, evidence is found to support hypothesis four. According to the results of the OLS regression, the magnitude of the stock price reaction is not influenced by the nature of the breach. This is in line with the studies of Cavusoglu, Mishra and Raghunathan (2004) and Gatzlaff and McCullough (2008), who also find no relation between the nature of the breach and the magnitude of the stock price reaction.

Table VIII : Determinants of the cumulative abnormal returns

This Table reports the results of the OLS regression. The cumulative abnormal returns are based on the event window one day before until one day after the announcement. The CAR is based on the market model. The standard errors are adjusted for heteroscedasticity and clustering at the firm level.

	CAR (-1,1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Financial information loss	-0.005	-0.006*	-0.007*	-0.007**	-0.007*	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Repeated data breach within	-0.002	-0.002	-0.000	-0.001	-0.003	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Credit card fraud		-0.001	-0.005	-0.009	-0.005	-0.008
		(0.017)	(0.017)	(0.017)	(0.020)	(0.017)
Hack		-0.000	-0.002	-0.003	0.004	-0.003
		(0.005)	(0.005)	(0.005)	(0.007)	(0.005)
Insider		0.007	0.005	0.006	0.012	0.006
		(0.006)	(0.006)	(0.006)	(0.008)	(0.006)
Portable device		0.007	0.006	0.007	0.012	0.007
		(0.005)	(0.005)	(0.005)	(0.008)	(0.005)
Stationary device loss		0.004	0.003	0.008	0.006	0.009
		(0.009)	(0.010)	(0.011)	(0.010)	(0.011)
Log (firm size)			0.001	-0.006	-0.010	-0.010
			(0.023)	(0.024)	(0.027)	(0.023)
Log (firm age)			-0.003	-0.000	-0.001	0.001
			(0.009)	(0.009)	(0.011)	(0.009)
ROA			-0.018	-0.013	-0.010	-0.018
			(0.021)	(0.019)	(0.023)	(0.019)
Leverage			-0.007*	-0.009**	-0.015**	-0.011**
			(0.004)	(0.005)	(0.007)	(0.005)
Sales growth			-0.018**	-0.016	-0.016	-0.017
			(0.009)	(0.010)	(0.011)	(0.011)
Tobin's q			0.001	0.001	0.002	0.001
			(0.001)	(0.001)	(0.002)	(0.002)
Block ownership			-0.014	-0.012	-0.015	-0.013
			(0.011)	(0.011)	(0.013)	(0.011)
Unique industry						-0.003
						(0.003)
Industry Herfindahl index						0.010
						(0.012)
Industry Tobins's q						-0.005
						(0.004)
_cons	-0.001	-0.003	0.012	-0.002	0.010	0.003
	(0.003)	(0.005)	(0.011)	(0.016)	(0.026)	(0.017)
Observations	313	313	313	313	313	313
R-squared	0.007	0.020	0.046	0.081	0.185	0.090
Industry Dummy	N	N	N	N	Y	N
Year Dummy	N	N	N	Y	Y	Y

Standard errors are in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

5.4 Event study on option trading

As the results of the event study on equity trading find no increase in volume, it could be of interest to further investigate the option market to find out whether data breaches influence the amount of option trading. The sample consists of 259 events (174 unique firms) for volume and 258 events for open interest in the period 2005-2017. There was no option data available after 2017.

5.4.1 Event study on put option volume and open interest

Table IX panel A reports the mean CAVs for different event windows. As shown in Table IX, the mean CAVs (-1,1), (-2,2) and (-5,5) are not significant. Furthermore, the event windows (-40,-1), (-30,-1), (-20,-1) and (-10,-1) do not show any positive significant increase in put option trading volume nor in put option open interest. These longer event windows are used to find evidence of informed trading. So far, there was only one study that investigated the effects of data breaches on option trading. The study of Mitts and Talley (2018) shows a significant increase in put option trading ahead of the announcement of a data breach. This study shows evidence on informed trading in the option market prior to the announcement of a data breach. The difference in findings between this study and the above-mentioned study could be caused by the difference in the sample construction as well as the methodology. Mitts and Talley (2018) use the propensity-score matching technique and have a final sample of almost 51 matched firms. Their study also finds an increase in put option open interest. However, in line with the results of put option volume found in this study, no increase was found in put option open interest either.

5.4.2 Event study on call option volume and open interest

Table X reports the mean CAVs for different event windows. As shown in in the Table, the mean CAVs (-1,1), (-2,2) and (-5,5) are not significant. The event windows (-40,-1), (-30,-1), (-20,-1) and (-10,-1) do not show a positive significant increase in call option trading volume or in open interest. So far, there was only one study that investigated the effects of data breaches on option trading. However, Mitts and Talley (2018) only investigate the effect on put option volume and open interest.

Investors could also benefit financially by trading in the call option market. However, the results show no evidence of an increase in call option volume and call option open interest.

In summary, no evidence is found in support of data breaches affecting the put and call option market. None of the results are significant.

Table IX : Put option volume and open interest

Panel A. Put option trading volume

Event window	Sample size	Mean	P-value
CAV (-1,1)	259	0.202	0.313
CAV (-2,2)	259	0.136	0.647
CAV (-5,5)	259	0.206	0.724
CAV (-10,-1)	259	-0.019	0.966
CAV (-20,-1)	259	0.306	0.685
CAV (-30,-1)	259	0.235	0.818
CAV (-40,-1)	259	-0.064	0.960

Panel B. Put option open interest

Event window	Sample size	Mean	P-value
CAV (-1,1)	258	-0.014	0.899
CAV (-2,2)	258	-0.009	0.962
CAV (-5,5)	258	-0.043	0.915
CAV (-10,-1)	258	-0.255	0.651
CAV (-20,-1)	258	-0.625	0.354
CAV (-30,-1)	258	-0.859	0.375
CAV (-40,-1)	258	-0.936	0.438

Table X : Call option volume and open interest

Panel A. Call option trading volume

Event window	Sample size	Mean	P-value
CAV (-1,1)	259	0.137	0.457
CAV (-2,2)	259	-0.072	0.869
CAV (-5,5)	259	-0.220	0.691
CAV (-10,-1)	259	-0.185	0.682
CAV (-20,-1)	259	-0.552	0.560
CAV (-30,-1)	259	-0.709	0.500
CAV (-40,-1)	259	-0.874	0.509

Panel B. Call option open interest

Event window	Sample size	Mean	P-value
CAV (-1,1)	258	-0.042	0.660
CAV (-2,2)	258	-0.067	0.673
CAV (-5,5)	258	-0.232	0.506
CAV (-10,-1)	258	-0.280	0.362
CAV (-20,-1)	258	-0.434	0.457
CAV (-30,-1)	258	-0.478	0.497
CAV (-40,-1)	258	-0.588	0.568

6. Conclusion

This study analyzed which firm- and industry-characteristics influence the likelihood of firms experiencing a data breach. Furthermore, it provided information about the effect of data breaches on stock price returns. Since, data breaches affect stock prices significantly, this study also tried to assess whether investors benefit financially by trading on inside information in the equity market. Moreover, it attempted to identify which firm- and industry-specific characteristics have an impact on the magnitude of the stock price reaction. Finally, evidence was sought of informed trading surrounding the announcement of data breaches in the option market.

First, this study provides empirical evidence that a firm's visibility is positively related to the likelihood of it experiencing a data breach. Firms that are larger and that have a higher percentage of shares held by institutional block holders are more likely to experience a data breach. In addition, inclusion on the list of Fortune 500 companies is also positively related to the likelihood of experiencing a data breach. These findings contradict the first hypothesis, which states that less visible firms are more likely to experience a data breach. An explanation for these contradictory findings could be that the studies on which the first hypothesis was based use a different sample, with fewer and different types of data breaches. Another important explanation for the findings in this study could be that, since large firms have more valuable information and more money, there is more to be gained from a data breach and therefore it is more profitable for hackers to focus on these companies.

Secondly, strong evidence was found to support that in the short-run, abnormal stock returns are generated surrounding the announcement of data breaches. The cumulative abnormal stock returns in different short-term event windows were significant and negative. Furthermore, based on the OLS regression, evidence was found revealing that the negative stock price reaction was larger for data breaches that included the loss of financial or other valuable information compared to data breaches that did not include the loss of valuable information. This is in line with current literature, which suggests that the loss of valuable information significantly impacts the magnitude of the stock price reaction. When, dividing the total sample into six different types of data breaches, only the group that was defined as cyberattacks was found to have a significant impact on a company's stock price. However, based on a difference test, mixed results were found on whether the nature of a breach impacts the magnitude of the stock price reaction. Moreover, based on the OLS regression, no evidence was found that the nature of a breach impacts the magnitude of the stock price reaction.

Since evidence was found that data breaches affect a company's stock price significantly, it could be possible for investors to benefit financially by acting based on insider knowledge. However, based on trading volume in the equity market, no evidence was found of informed trading related to data breaches.

Lastly, while no evidence was found of informed trading in the equity market, investors could potentially seek for financial benefits in the option market. However, no evidence was found that the volume and open interest in the option market increases surrounding the announcement of data breaches.

7. Limitations and recommendations

This section discusses the most important limitations of this research and proposes recommendations for future research.

The first limitation of this research came from manually matching data breaches with the CRSP database to find out whether companies were listed. As a result, many observations were deleted from the sample. The final sample contained 313 events for 216 distinct firms. While the sample was large enough to draw conclusions, it was still fairly small. As a solution, this same study could be done in five years from now. In five years from now, a much larger dataset is expected to be available since the number of data breaches is increasing and regulation about the announcement of data breaches is becoming stricter. Another solution could be to expand the dataset with data breaches that have happened outside the US.

Furthermore, the database of the Privacy Clearing House was used as a starting point of this research. As this paper performed an event study on equity and option trading, the exact announcement date of a data breach was enormously important. However, many different announcement dates were found when matching data breaches from the PRC database and news articles in which the specific data breach was published. In this study the first date mentioned was used as the announcement date, whether it was in the PRC database or in a specific newspaper. However, this difference of announcement dates might have had a negative influence on the accuracy of the event studies.

Moreover, in this study six different types of data breaches were investigated. However, this study makes no distinction between the strength and size of a particular data breach. In future research, it is advisable to take into account the size and impact of a data breach, since these factors vary enormously. When doing the event study it could be interesting to divide the total sample into different groups by the size or impact of the breach. The impact or size of a breach is hard to measure, however a possible proxy for the size of a data breach could be the number of times the breach was mentioned in professional newspapers. The announcement effect is not as clear as the announcement effect for mergers and acquisitions or earnings announcements, and therefore making a distinction between the impacts of data breaches would be an interesting addition to this study.

Lastly, regulation around the announcement of data breaches is not uniform. States have different rules concerning the announcement of data breaches. This leads to a gap between the moment a company notices a data breach and the moment it is announced. Future research should take into account the period between noticing the data breach and the moment the breach is announced. This time gap could potentially impact the stock price reaction.

8. Appendix

Table A : Chronological distribution of data breaches
The Table presents the distribution of 313 data breaches by calendar year and industry. Industry is based on the two-digit SIC code.

YEAR	Mineral Construction (10-17)	Manufacturing (20-39)	Transport Communication (40-48)	Electric Gas Sanitary (49)	Wholesale Retail (50-59)	Finance (60)	Services (70-89)	Public administration (99)	Total
2005	0	4	0	0	0	0	0	0	4
2006	0	4	4	0	5	11	5	1	27
2007	1	8	2	1	5	10	5	0	31
2008	0	6	2	1	3	6	3	0	19
2009	1	0	1	0	0	3	0	0	10
2010	0	2	2	0	11	11	11	0	31
2011	0	5	1	0	5	6	5	0	22
2012	0	7	2	0	5	4	5	0	29
2013	0	5	1	1	5	6	5	0	22
2014	0	7	4	1	7	9	7	0	35
2015	0	2	3	0	8	4	8	0	20
2016	0	4	2	0	6	5	6	0	27
2017	0	4	0	0	5	6	5	0	22
2018	0	3	1	0	3	3	3	0	14
Total	2	61	25	4	68	84	68	1	313

Table B: Correlation matrix test for multicollinearity

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Log (Firm size)	1.000								
(2) Log (Firm age)	0.277	1.000							
(3) Fortune 500	0.608	0.268	1.000						
(4) ROA	0.350	0.182	0.132	1.000					
(5) Leverage	0.245	0.023	0.119	-0.065	1.000				
(6) Sales growth	-0.094	-0.217	-0.060	-0.063	-0.014	1.000			
(7) Tobin's q	-0.275	-0.137	-0.071	-0.213	-0.077	0.186	1.000		
(8) Block ownership	0.068	0.188	-0.036	0.101	0.004	-0.068	-0.023	1.000	
(9) Asset intangibility	-0.116	-0.109	-0.104	-0.084	-0.280	0.020	0.086	-0.004	1.000

Variables	(1)	(2)	(3)
(1) Unique industry	1.000		
(2) Industry Herfindahl index	-0.331	1.000	
(3) Industry Tobin's q	0.297	-0.119	1.000

Table C : Breusch-Pagan / Cook Weisburg test for heteroscedasticity

This Table presents the Breusch-Pagan / Cook Weisburg test for heteroscedasticity with the dependent variable CAR (-1,1) based on the market model.

	Coefficient	St.Error	t-value	p-value
Financial information loss (indicator)	-0.005	0.003	-1.50	0.135
Repeated data breach within one year	-0.002	0.004	-0.57	0.566
Firm size	0.009	0.029	0.30	0.763
Log (firm age)	-0.006	0.011	-0.57	0.568
Fortune 500 (indicator)	0.007	0.004	1.88	0.061*
ROA	-0.029	0.023	-1.23	0.221
Leverage	-0.009	0.006	-1.52	0.131
Sales growth	-0.015	0.008	-1.81*	0.071
Tobin's q	0.002	0.002	1.10	0.274
Block ownership	-0.009	0.011	-0.85	0.396
Industry Herfindahl Index	0.003	0.016	0.17	0.863
Industry Tobin's q	-0.006	0.004	-1.06	0.291
Unique industry (indicator)	-0.003	0.003	-0.88	0.379
Constant	0.015	0.010	1.48*	0.140

Breusch-Pagan / Cook-Weisberg test for heteroscedasticity
Ho: Constant variance
Variables: fitted values of CAR
chi2(1) = 56.07
Prob > chi2 = 0.0000

Table D :Robustness check – Corrado rank test

Market model				
Event days	**Sample size**	**AAR%**	**P-value t-test**	**Corrado Rank test**
-3	313	-0.076	-0.430	0.456
-2	313	-0.225	0.030**	0.032**
-1	313	-0.074	0.400	0.217
0	313	-0.204	0.069*	0.043**
1	313	-0.234	0.012**	0.010***
2	313	0.178	0.054*	0.100
3	313	-0.012	0.925	0.204

Table E: Event study on different data breaches

Event window (%)	**Sample size**	**Mean**	**t-test**	**P-value**
Hacking or Malware				
CAR (-1,1)	152	-0.746	-3.092	0.001***
CAR (-2,2)	152	-0.678	-2.309	0.011**
CAR (-5,5)	152	-1.313	-2.660	0.004***
Credit card fraud				
CAR (-1,1)	10	-0.833	-0.486	0.319
CAR (-2,2)	10	-1.247	-0.836	0.212
CAR (-5,5)	10	4.619	1.195	0.869
Insider breaches information				
CAR (-1,1)	53	-0.150	-0.448	0.328
CAR (-2,2)	53	-0.468	-0.918	0.182
CAR (-5,5)	53	-0.647	-0.980	0.166
Physical loss of documents				
CAR (-1,1)	25	-0.585	-1.162	0.128
CAR (-2,2)	25	-0.240	-0.313	0.379
CAR (-5,5)	25	0.274	0.209	0.582
Loss of portable device				
CAR (-1,1)	63	-0.216	-0.803	0.213
CAR (-2,2)	63	-0.540	-1.321	0.096
CAR (-5,5)	63	0.135	0.253	0.600
Loss of stationary device				
CAR (-1,1)	10	-0.496	-0.597	0.283
CAR (-2,2)	10	0.365	0.226	0.587
CAR (-5,5)	10	-0.929	-0.642	0.268

Table F: Comparison between cyberattacks and other types of data breaches

	Hack or Malware (N=152)		Credit card fraud (N=10)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.746	0.001***	-0.833	0.319	-0.085	0.933
CAR (-2,2)	-0.678	0.011**	-1.247	0.212	-0.375	0.716
CAR (-5,5)	-1.313	0.004***	4.619	0.869	1.522	0.161
	Hack or Malware (N=152)		Insider breaches info (N=53)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.746	0.001***	-0.150	0.328	1.438	0.153
CAR (-2,2)	-0.678	0.011**	-0.468	0.182	0.356	0.722
CAR (-5,5)	-1.313	0.004***	-0.647	0.166	0.808	0.421
	Hack or Malware (N=152)		Physical loss (N=25)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.746	0.001***	-0.585	0.128	0.289	0.775
CAR (-2,2)	-0.678	0.011**	-0.240	0.379	0.5331	0.598
CAR (-5,5)	-1.313	0.004***	0.274	0.582	1.138	0.264
	Hack or Malware (N=152)		Loss of portable devices (N=63)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.746	0.001***	-0.216	0.213	1.468	0.144
CAR (-2,2)	-0.678	0.011**	-0.540	0.096	0.274	0.785
CAR (-5,5)	-1.313	0.004***	0.135	0.600	1.996	0.048
	Hack or Malware (N=152)		Loss of stationary devices (N=10)		Test of difference (a-b)	
Event window (%)	**Mean**	**P-value**	**Mean**	**P-value**	**t-test**	**P-value**
CAR (-1,1)	-0.746	0.001***	-0.496	0.283	0.288	0.779
CAR (-2,2)	-0.678	0.011**	0.365	0.587	0.634	0.541
CAR (-5,5)	-1.313	0.004***	-0.929	0.268	0.251	0.806

Table G : Variable construction

Variable	Definition	Source
Asset intangibility	(1 – PPE) / total assets	
CAPX / Total assets	Capital expenditures / total assets	Compustat
CAR	Cumulative abnormal return	CRSP
Credit card (indicator)	Fraud that involves debit and credit card fraud. Excludes hacking	Privacy Rights Clearinghouse
Electric, gas and sanitary industry	One for companies active in industries with SIC code: 49	Compustat
Finance industry	One for companies active in industries with SIC code: 60	Compustat
Financial loss (indicator)	One for companies at which the data breach involved the loss of financial information	Privacy Rights Clearinghouse
Fortune 500 (indicator)	Dummy if a firm is included in the list of US Fortune 500 companies	Compustat / Fortune500.com
Hack (indicator)	Hacking or malware initiated by an outside party	Privacy Rights Clearinghouse
HML	½ *(Small value + Big value) – ½ *(Small growth + Big growth) (M/B) cutoffs are bottom 30% and top 70%	CRSP
Industry	One for the two-digit SIC code in which the firm is active	Compustat
Industry Herfindahl index	Sum of squared market shares based on two-digit SIC code	Compustat
Industry Tobin's q	Median Tobin´s q of all firms in the same industry based on the two-digit SIC code	Compustat
Insider (indicator)	Someone from the inside intentionally breaches information	Privacy Rights Clearinghouse
Institutional block ownership	Percentage of total shares outstanding owned by shareholders who own more than 5%	Compustat
Leverage	Total debt / total assets	Compustat
Log (Firm age)	Logarithm of number of years available in Compustat	
Log (Firm size)	Log (Total assets)	Compustat
Mineral and Construction industry	One for companies active in industries with SIC codes: 10-17	Compustat
Portable loss (indicator)	Lost or stolen portable device	Privacy Rights Clearinghouse
Public administration industry	One for companies active in industries with SIC code: 99	Compustat
R&D / Total assets	Research and development / total assets	Compustat
Repeated breach (indicator)	One for companies at which the data breach was a subsequent breach	Privacy Rights Clearinghouse
ROA	Net income / total assets	Compustat
Sales growth	$Sales_{t-1}/Sales_t$	Compustat
Service industries	One for companies active in industries with SIC codes: 70-89	Compustat
SMB	1/3 * (Small Value + Small neutral + Small growth) – 1/3 *(Big value + Big neutral + Big growth) Big stock are top 90% and small bottom 10%.	CRSP
Stationary loss (indicator)	Lost or stolen stationary device	Privacy Rights Clearinghouse
Tobin's q	(Total assets – common equity + market value of equity)/ total assets	Compustat
Transport and Communication industry	One for companies active in industries with SIC codes: 40-48	Compustat

Unique industry	One if a company is active in the top quartile of the two-digit SIC industries based on the median of product uniqueness. Product uniques is defined as : Selling expense / sales	Compustat
Wholesale and retail industry	One for companies active in industries with SIC codes: 50-59	Compustat
Year	One for the fiscal year before the data breach	Compustat

9. References

Acquisti, A., Friedman, A., and Telang, R. (2006). Is there a cost to privacy breaches? An event study. Paper presented at the International Conference on Information Systems (ICIS), Milwaukee, WI.

Ajinkaya, B.B., Jain, P.C. (1989). The behavior of daily stock market trading volume. Journal of Accounting and Economics, 11, 331-359.

Akey, P., Lewellen, S., and Liskovich, I. (2018). Hacking Corporate Reputations. Working paper. University of Toronto, Toronto.

Arcuri, M. C., Brogi, M., Gandolfi, G. (2017). How does cybercrime affect firms? The effect of information security breaches on stock returns. First Italian Conference on Cybersecurity, Venice.

Augustin, P., Brenner, M., Grass, G., and Subrahmanyam, M.G. (2018). How do informed investors trade in the options market? Working paper. Canadian Derivatives Institute, Montreal.

Bernard, T.S., Hsu, T., Perlroth, N. and Lieber, R. (2017,September 7). Equifax Says Cyberattack May Have Affected 143 Million in the U.S.. The New York Times. Retrieved from www.nytimes.com.

Bianchi, D., and Tosun., O. (2018). Cyberattacks and stock market activity. Working paper. University of Warwick, Coventry.

Blau, B. M., and Tew, P. L. (2014). Short sales and class-action lawsuits. Journal of Financial Markets, 20, 79-100.

Brown, S.J., Warner, J.B. (1980). Measuring security price performance. Journal of Financial Economics, 8, 205-258.

Camp, L.J., and Lewis, S. (2006). Economics of Information Security. Springer Science & Business Media, New York.

Campbell, K., Lawrence, A., Gordon, M., Loeb, P., and 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.

Campbell, C.J., Wasley, C.E. (1996). Measuring Abnormal Daily Trading Volume for Samples of NYSE/ASE and Nasdaq Securities Using Parametric and Nonparametric Test Statistics. Review of Quantitative Finance and Accounting, 6, 309-326.

Cavusoglu, H., Mishra, B., and Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce, 9, 70-104.

Chakravarty, S., Gulen, H., Mayhew, S. (2004). Informed trading in stock and option markets. The Journal of Finance, 59(3), 1235–1257.

Christophe, S., Ferri, M., and Angel, J. (2004). Short-selling prior to earnings announcements. Journal of Finance, 59(4), 1845–1875.

Christophe, S., Ferri, M., and Hsieh, J. (2010). Informed trading before analyst downgrades: evidence from short sellers. Journal of Financial Economics, 95, 85–106.

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

Easley, D., and O'Hara, M. (2004). Information and the cost of capital. Journal of Finance, 59, 1553–1583.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383-417.

Fama. E., K.R. French. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33, 3-56.

Garg, A., Curtis, J., Halper, H. (2003b). Quantifying the financial impact of IT security breaches. Information Management and Computer Security, 11(2), 74-83.

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

Ge, L., Hu, J., Humphery-Jenner, M., and Lin, T. (2014). Informed Options Trading Prior to Bankruptcy Filings. Working paper. Lee Kong Chian School of Business, Singapore.

Hao, X., Lee, E., Piqueira, N., (2013). Short sales and put options: Where is the bad news first traded? Journal of Financial Markets 16 (2), 308–330.

Hilary, G., Segal, B., and Zhang, M.H. (2016) Cyber-risk disclosure: Who cares? Working paper. Georgetown University, Georgetown.

Horowitz, J.L., Savin, N.E., (2001). Binary response models: logits, probits and semiparametrics. Journal of Economic Perspective 15 (4), 43–56.

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

International Standards Organization (2016). ISO/IEC 27050. Retrieved from: www.iso.org.

Jayaraman, N., Frye, M.B., Sabherwal, S. (2001). Informed Trading around Merger Announcements: An Empirical Test Using Transaction Volume and Open Interest in Option Market. The Financial Review, 37, 45-74.

Jarrell, G., and Poulsen, A.B. (1989). Stock Trading before the Announcement of Tender Offers: Insider Trading or Market Anticipation? Journal of Law, Economics, and Organization, 5(2), 225-48.

Johnson, M., Kang, M.J., and Lawson, T. (2017) Stock price reaction to data breaches. Journal of Finance Issues, 16, 1-13.

Kamiya, S., Kang, J.-K.,Kim, J., Milidonis, A., and Stulz, R. M. (2018). What is the Impact of Successful Cyberattacks on Target Firms? Tech. rep., National Bureau of Economic Research.

Karpoff, J.M. (1987). The relation between price changes and trading volume: A Survey. The Journal of Financial and Quantitative Analysis, 22, 109-126.

Keown, A.J., & Pinkerton, J.M. (1981). Merger Announcements and Insider Trading Activity: An Empirical Investigation. Journal of Finance, 36(4), 855-869.

Larcker, D.F., Reiss, P.C., and Brian T. (2017). Critical update needed: Cybersecruity expertise in the boardroom. Working paper. Stanford University, Stanford.

Lending, C.,Minnick, K., and Schorno, P. J. (2018). Corporate Governance, Social Responsibility and Data Breaches. Financial Review 53(2): 413-455.

Lewis, J. (2018). Economic Impact of Cybercrime, No Slowing Down. McAfee.

MacKinlay, C.A. (1997). Event Studies in Economics and Finance. Journal of Economic Literature, 35, 13-39.

Mitts, J., and Talley, E.L. (2018). Informed Trading and Cybersecurity Breaches. Working paper. Columbia Law School, New York.

Myung, K., and Dorantes, C. (2006). The impact of information security breaches on financial performance of the breached firms: an empirical investigation. Journal of Information Technology Management, 17, 13-22.

Poteshman, A. M., (2006). Unusual option market activity and the Terrorist attacks of September 11, 2001. The Journal of Business, 79(4), 1703-1726.

U.S. Securities and Exchange Commission. (2010). Option trading. Retrieved from: https://www.sec.gov/fast-answers/answersoptions

U.S. Securities and Exchange Commission. (2018a). Former Equifax Executive Charged With Insider Trading. Retrieved from: https://www.sec.gov/news/press-release/2018-40

U.S. Securities and Exchange Commission. (2018b). Former Equifax Manager Charged With Insider Trading. Retrieved from: https://www.sec.gov/news/press-release/2018-115

Spanos, G., and Angelis, L. (2016).The Impact of Information Security Events to the Stock Market: A Systematic Literature Review. Computers & Security, 58, 216-229.

Tsiakis, T., and Stephanides, G. (2005). The economic approach of information security. Computers & Security, 24 (2): 105-108.

Von Solms, R., Van Niekerk, J. (2013). From information security to cybersecurity. Computers & Security, 38, 97–102.

Womack, K. (1996). Do Brokerage Analysts' Recommendations have investment value? The Journal of Finance, 51(1), 137-167.

Ying, C.C. (1966). Stock Market prices and Volumes of Sales. Econometrica, 34, 676-686.