From Shares to Shields: The Role of Employee Ownership in Mitigating Data Breach Risks

Mahtab Karimi*

December 2024

Abstract

Employees have a comparative advantage in monitoring their peers and other employees—an advantage that managers and executives do not have. I call this mutual monitoring and investigate whether it can protect firms from risks where a small mistake by one individual could cause significant harm, such as data breaches. To explore this, using AI(BERT models), I conduct a comprehensive textual analysis to identify distinct data breach incidents reported by the same firm. I then examine how the ratio of active participants in employee stock ownership plans (ESOPs) to total employees (the active ratio) affects the probability of a data breach incident. My findings indicate that a higher active ratio is associated with a lower probability of such incidents. Moreover, by analyzing the extensive and intensive margins of ESOP ownership, I find that two firms with the same ESOP value per employee can experience different levels of protection against data breaches due to variations in their active ratio; the distribution of ESOP assets within the firm matters in protecting against data breach incidents. Next, using a staggered difference-in-differences model, I analyze how the first noticeable data breach in an industry impacts the active ratio of peer firms within the same industry. I find that ESOP firms increase their active ratio by 3 to 4 percentage points following the industry shock, driven by a 7 to 10 percent increase in the number of ESOP owners within those firms. Notably, this change remains persistent after excluding financial firms and industry shocks coinciding with the years 2007 and 2008.

*I am grateful to my advisors Jarrad Harford, Eric Zivot, Jonathan Karpoff, Jing Tao, and Melissa Knox for invaluable guidance and support throughout this project. I am thankful to Douglas Kruse for sharing the data on employee stock ownership plans. I also thank both Douglas Kruse and Ethan Rouen for sharing information and insights on the structure of ESOPs. I thank Philip Bond, Doron Levit, Stephan Siegel, Jordan Nickerson, German Gutierrez, Ahmed Guecioueur, and the participants of the Ph.D. seminar at the Foster School of Business. I thank Ali Karimirad and Ken Inosaki for their helpful comments.

All errors are my own. Mahtab Karimi: Department of Economics, University of Washington, Savery Hall, 410 Spokane Lane, Seattle, WA 98105; mahtak@uw.edu

Acknowledgment:

I confirm that this research has not received any external funding or financial support. Additionally, I have no financial relationships or conflicts of interest to disclose.

" ... No junior employee at Microsoft or Intel can improve the value of her heavyweight employer to such a degree that it will make it worthwhile for her to work harder once stock options are offered. Nevertheless, given the sensitivity of the "knowledge industry" to leakage of its intellectual property, all employees can add much to the company's value by standing on guard against such loss. ... Stock options privatize the firm's monitoring task into the hands of its employees." Hannes, 2006

1 Introduction

Companies often state in their 10-K and proxy statements that they provide broad stock-based compensation to enhance productivity. An increase in employee ownership, along with the alignment of employees' interests with those of shareholders, is believed to improve employee performance and overall company productivity. Nevertheless, the economic literature presents mixed results regarding the relationship between productivity and stock-based compensation. This raises another question worthy of further investigation: what if companies use this type of compensation as a risk management tool rather than as a direct means of increasing productivity? Put differently, what if firms implement stock-based compensation to internalize the negative externalities associated with employee actions, ultimately leading to a lower likelihood of risky outcomes that can be caused by employees actions?

To answer this general question, I focus on one of the key benefits of stock-based compensation: mutual monitoring. This concept is based on the idea that employees have a comparative advantage in monitoring their peers and other employees—an advantage that managers and executives typically lack. By offering employees company stock and making them partial owners, firms encourage employees to act more responsibly, both in their own actions and in overseeing those of their peers. I refer to this mechanism as the mutual monitoring channel throughout this work. Mutual monitoring can be particularly helpful in situations where a mistake by one person could lead to significant losses for the entire firm, like data breaches. In other words, I investigate if the mutual monitoring can protect the firm against risky and costly outcomes like data breaches. Specifically, I ask, do companies offer employee stock-based compensation to encourage employees to monitor one another and protect the company against costly and risky outcomes, such as data breaches?

Data breaches are costly, and damage firms' reputations. IBM's 2024 annual report on data breaches identifies three primary root causes: malicious or criminal attacks account for 55%, IT failures for 23%, and human error for 22%. The 22% attributed to human error directly results from individual mistakes, while the other two categories could be indirectly related to human errors and the actions of system users within

the firm. In a survey[1] 71% of business leaders indicated that they think next cybersecurity breach will come from the inside.

Given these facts, firms seem to recognize the significant role employees play in these incidents. Ownership in the company can transfer some of the costs of data breaches to employees. If employees who are owners of the company are more careful with their own actions and those of their peers to avoid harming the firm, then there is a strong possibility that companies leverage stock-based compensation as a defense against cybersecurity issues arising from their own employees' actions.

Given the anecdotal consensus that "the company can't even determine who the unauthorized party that breached its network was,"[2] this channel could be even more important. within this framework, employees are likely to work more carefully because they are owners or because their colleagues are monitoring them. Moreover, Kamiya et al.,2021 reports that the unconditional probability of a data breach incident in a given year within the sample of all public firms from 2005 to 2017 is 0.047%, which can result in an average loss of $1.25 billion for shareholders. Therefore, it is reasonable to consider a data breach incident as a tail event—an event with a low probability but high potential damage.

So far, I have introduced data breaches as the risky outcome and presented evidence showing that employees play a significant role in their occurrence, as well as firms' awareness of this fact. Now, I outline the approach I will use to estimate mutual monitoring in this study. To define a mutual monitoring measure, I focus on Employee Stock Ownership Plans (ESOPs). These are retirement stock-based benefit plans offered to employees by their company. The main variable or my mutual monitoring measure is called active ratio and is defined as the ratio of active participants—employees who are hired, actively work for the company, and have an ESOP—to total employees. Therefore, in this work, I use the active ratio and mutual monitoring measure interchangeably.

ESOPs have three characteristics that make them plausible within the structure of my argument. First, compensation is broad-based and offered to employees at various levels, not exclusively to executives. Second, compensation is determined by the company and given to the employees; employees themselves do not have a discretion in their ownership. Third, employees do not have the option to trade their stock or time the exercise of their options. In other words, employees do not have access to their stock before retirement or leaving the company.

I now explain why ESOP holders have a stake in mitigating the risk of a data breach.1"originatingScript":"m2","payload":"
0daf-4d55-9a53-30698dd33e84e10caf","muid":"157ba760-832a-454e-a988-20d44822999cc56055","sid":"1e52bfef-

[1] https://www.prnewswire.com/news-releases/the-threat-from-within-71-of-business-leaders-surveyed-think-next-cybersecurity-breach-will-come-from-the-inside-301733317.html

[2] https://www.myinjuryattorney.com/consumer-privacy-data-breach-lawyers/if-your-information-has-been-compromised-in-a-data-breach/

b7dc-436e-909f-5958227b0dbf850907""originatingScript":"m2","payload":"guid":"7e2c7b8e-0daf-4d55-9a53-30698dd33e84e10caf","muid":"157ba760-832a-454e-a988-20d44822999cc56055","sid":"c7be0e43-e5f2-49ce-adb9-24d96c2fef2c3cf8e9" Kamiya et al., 2021 reports that firms experience a 0.8% decrease in CAR during the three-day window surrounding data breach announcements. In my sample, the average market value of ESOP assets per participant equals $130,000, translating to an average loss of approximately $1,000 per participant. Moreover, in the case of a tail event, when people try to avoid bad outcomes, they tend to focus on the worst-case scenarios. As a result, employees may be particularly concerned about outcomes that are worse than average. In a subsample of incidents where the breach is followed by an FBI investigation, litigation, or lawsuits, the change in CAR can reach -8%, translating to an average loss of $10,000 per employee. In the worst-case scenario, post-attack costs can become so significant that the company is forced to file for bankruptcy.

In summary, data breaches are events that, when they occur, can cause significant harm to a company and its shareholders. Employees play a role in these incidents. ESOPs activate mutual monitoring among employee owners and help reduce the associated negative externalities. Therefore, rather than studying the effect of ESOPs on productivity, I, for the first time, investigate whether firms use broad-based stock compensation to mitigate the risk of data breaches. Specifically, I ask three questions: First, does a higher active ESOP employee ratio lead to a lower likelihood of a data breach? Second, do companies that have been the target of a data breach take actions to increase their active ratio to be more protected? Finally, does a noticeable breach in one company lead other ESOP companies in the same industry to take actions to be more prepared for a possible future breach?

I use two datasets in this study: data breach reports collected by the Privacy Rights Clearinghouse (PRC) from 2005 to 2023, and information on ESOP plans extracted from IRS Form 5500 for the period 2010 to 2023. I begin by using the EDGAR API [3] provided to identify all public and private firms in the data breach incident dataset that have more than 100 employees. I then use the ESOP plan information to map ESOP firms to the public firms identified in the PRC database. A key challenge at this stage is that the PRC database includes repeated incident reports. Different companies often report the same incident to multiple authorities, such as the attorneys general or consumer agencies. Additionally, companies sometimes release more information about the same breach over time, reporting it in stages. To identify distinct incidents in the dataset, I implement machine learning tools. My work is the first to employ machine learning, offering a tractable and expandable approach for identifying distinct data breach incidents.

I utilize an Sentence-BERT model and define the similarity between two data breach incidents reported

[3]Electronic Data Gathering, Analysis, and Retrieval, It's the filing system used by the U.S. Securities and Exchange Commission (SEC) for companies and individuals to submit documents required by federal securities laws.

by the same firm as the cosine of the angle between their embedded vectors estimated by S-BERT. I then create an index with the value of the time gap between two incidents divided by their similarity measure. My rationale is that two data breach incidents reported close together but with low similarity in their descriptions are likely to be distinct events. Since the similarity measure is a number less than one, dividing the time gap by this measure amplifies the value of the time gap. Therefore, a larger distinction index indicates that the two incidents are more distinct.

After identifying the distinct incidents using the distinction index, I address the first question. My hypothesis is that the higher the active ratio of ESOP employees in a firm, the more employee monitors the firm has, the lower the probability of a data breach. The results of the probit model align with this prediction. After controlling for executive compensation, employee wages, and the value of employees' ESOP ownership, I find that a higher active ratio is associated with a lower probability of a data breach. However, firm size can weaken this protective role of ESOP employees, as larger firms often have more external relationships, such as with contractors or vendors.

Moreover, I sort the firms into three quantiles, once by the value of ESOP assets per employee and once by the active ratio. I find no significant difference between the first and third quantiles of ESOP assets per employee after controlling for the active ratio, while the active ratio is always negatively associated with the probability of a data breach. However, there is a significant difference between the first and third quantiles of the active ratio, whereas the value of ESOP assets per employee is not significant. In other words, this comparison shows that two firms with the same level of ESOP value per employee can have different levels of protection from data breaches due to differences in their active ratio. I refer to this analysis as an extensive vs. intensive margin analysis of ESOP ownership.

To determine whether data breach target firms increase their active participants ratio following the first breach they experience, as I hypothesize, I use a propensity score matching method and match each firm-year of a target firm with a firm-year of a non-target firm. I then utilize a staggered difference-in-difference to examine the change in the active ratio after a data breach. In this analysis, I control for two important sources of heterogeneity: one arising from industry conditions and the other from local labor market conditions. I find that the active ratio increases in targeted firms following their first breach. However, this effect disappears after accounting for industry and local labor market conditions.

Next, I examine the industry-wide effects of data breaches on non-targeted firms within each industry. First, I focus on all public and private firms with more than 100 employees identified in the PRC dataset, then I retain only incidents with more than 500 impacts- the number of people whose data been exposed- within each 4-digit SIC code. This is what I define as an industry shock. I then map these shocks to all public ESOP firms that have never experienced a data breach in the sample based on their 4-digit SIC codes.

Next, I implement a two-way fixed effects model to estimate the change in the active ratio for each 4-digit industry after the shock.

I find that, in the two years after the breach until five years after the shock, ESOP firms increase their active ratio by 3 to 4 percentage points. Looking more closely, I investigate what drives this change. It appears that when a noticeable breach occurs for the first time in an industry, peer firms in the same industry increase the number of their ESOP owners by 7 to 10 percent. Notably, this change remains persistent after excluding financial firms and industry shocks coinciding with the years 2007 and 2008. However, in financial firms, there is a significant decrease in wages and the value of ESOP ownership per participant after the shock. Moreover, firms also react by initially increasing the salary component of executives' compensation by 3 percentage points. However, after two periods, there is a decrease of 2 to 3 percentage points in non-equity incentive compensation.

In summary, and in a broader context, my findings show that a higher active ratio reduces the risk of a data breach due to the mutual monitoring activated by broad-based stock compensation. Moreover, firms deliberately increase their active ratio following the first noticeable data breach in their industry, which lowers the probability of subsequent breaches.

It's important to emphasize that my research primarily investigates whether companies use employee stock-based compensation as a defense against data breaches, similar to vaccination, rather than medicine or insurance. In other words, I am not suggesting that companies utilize broad-based employee stock compensation to address damages or offset probable costs associated with data breaches. This work empirically and directly study the relationship between mutual monitoring and broad-based stock compensation. The policy implications of my work suggest that focusing on the benefits of ESOPs as a risk-mitigating tool, rather than solely examining their relationship to productivity, can highlight the advantages of these plans more effectively and make them more compelling for companies.

1.1 Contribution to the Literature

Mutual monitoring, if mentioned in the prior literature, has been used as a possible link to explain the relationship between broad-based compensation and the outcome. A closely related strand is Freeman et al., 2008 which provides evidence, based on a survey, that employees monitor their peers in firms with some form of group incentive plans. Additionally, Core and Guay, 2001 suggest that direct monitoring becomes more difficult as firm size increases, so larger firms tend to offer more stock options to indirectly monitor their employees. This is how they interpret the positive relationship between firm size and stock option grants. My work empirically and directly study the relationship between mutual monitoring/self-controlling

and broad-based stock compensation.

In more detail, the extant literature categorizes the relationship between broad-based employee equity plans and firms into three primary areas: the effects of broad-based plans on performance and retention (e.g., Hall and Murphy, 2003, Oyer, 2004, Oyer and Schaefer, 2005, Hochberg and Lindsey, 2010, Kim and Ouimet, 2014, and Aldatmaz et al., 2018); the effects of broad-based plans on firms' governance (e.g., Pagano and Volpin, 2005, Call et al., 2016, Masulis et al., 2020, and Wu et al., 2023); and the effects of the plans on other operational aspects, such as innovation and ESG ratings (e.g., Babenko and Tserlukevich, 2009, Babenko et al., 2011, Chang et al., 2015, Babenko and Sen, 2016, and Kong et al., 2023). I go over each category in the following paragraphs.

The effects of broad based plans on performance and retention. Hall and Murphy, 2003 primarily examine executive stock options but also discuss broad-based options. They suggest that companies offer stock options because the perceived cost is significantly lower than the actual economic cost, due to accounting practices. Oyer, 2004 suggests that firms offer stock-based compensation to index employees' compensation to the value of their outside employment opportunities, which leads to higher retention. Oyer and Schaefer, 2005 calibrate results of a standard moral hazard model and find that stock options granted to middle managers are positively correlated with retention, but they reject an incentive-based explanation for offering the stock. Hochberg and Lindsey, 2010 use the incentive-based compensation of firms that are geographically close but not in the same industry as a determinant of offering broad-based plans. They conclude that offering broad-based options leads to a significant increase in return on assets. Kim and Ouimet, 2014 specifically focus on ESOP plans. They show that adopting an ESOP where employees have high bargaining power leads to higher wages, which they interpret as a sign of improved performance following the adoption of the ESOP. One specific difference between my work and this study is that I am focusing on the variation within ESOP firms, rather than between ESOP and non-ESOP firms, as this work does. In my study, the main variable is a continuous variable, whereas in this work, the main variable is an indicator of whether ESOP plans are offered or not. In a different study, Aldatmaz et al., 2018 find that employee turnover decreases in the years following a large broad-based employee stock option grant in a firm.

The effects of broad based plans on governance. Pagano and Volpin, 2005 demonstrate that a manager with a small equity stake can deter hostile takeovers by using two methods to align employees with them: offering high wages or providing membership in an employee stock ownership plan. This approach aligns the incentives of both employees and the manager, suggesting that the firm's workers will act against raiders and support the incumbent CEO. Masulis et al., 2020 show that employee voting rights derived from membership in ESOP plans allow managers to pursue value-destroying acquisitions. Call et al., 2016 show that companies tend to grant more options during periods of misreporting compared to non-violating firms and their own

non-violation periods. They suggest that companies prone to financial misreporting may use the strategic issuance of stock options to non-executive staff as a tactic to buy silence.

The effects of the plans on other operational aspects, such as innovation and ESG ratings. Babenko and Tserlukevich, 2009 study the tax benefits of broad-based stock options for firms. Babenko et al., 2011 show that the cash flow generated by exercising stock options leads to an increase in firm investment. Chang et al., 2015 demonstrate that broad-based employee stock compensation can encourage greater innovation among employees. Babenko and Sen, 2016 find that employees' contributions to ESPP plans[4] are associated with an increase in the number of patents filed by the company. Kong et al., 2023 show that offering ESOPs to non-executive employees leads to increased ESG ratings.

One difference between my work and the other works mentioned above is that I focus on a non-value measure of ESOPs and show that it remains effective even after controlling for the dollar value of ownership. Unlike previous works, the channel I examine is not about employees trying to become wealthier, but about protecting what they own, regardless of its value. This approach sheds new light on employee stock bonus plans and their roles in the firm.

Additionally, my work contributes to the existing literature on the consequences of data breaches for companies. The closest paper to my work in this field is Kamiya et al., 2021, which identifies factors influencing data breach likelihood and explores their impact on shareholders, management, and peer firms. However, my research differs in its focus. Rather than directly examining the consequences of data breaches, I explore how these events can serve as catalysts for mutual monitoring within ESOP firms. In essence, data breaches are treated as shocks that activate a specific mechanism, mutual monitoring, in these firms. The direct implications of data breaches are a secondary consideration in my analysis.

The extant literature on the effects of data breaches can be categorized into: the effects of data breaches on shareholder wealth (e.g., Acquisti et al., 2006, Gatzlaff and McCullough, 2010, Kamiya et al., 2021); the effects of data breaches on other financial outcomes, such as innovation (Ettredge et al., 2018, Huang and Wang, 2021, SUN, 2021, Wang et al., 2023, He et al., 2020); the effects of data breaches on management and governance (Banker and Feng, 2019, Zhang et al., 2024, Kamiya et al., 2021, Hsu and Wang, 2014, Lending et al., 2018,Hartmann and Carmenate, 2021, Ashraf, 2022); and the effects of data breaches on a firm's labor force (Akey et al., 2021, Bana et al., 2022). I go over each category in the following paragraphs.

The effects of data breaches on shareholder wealth. There are several papers in information science and technology science investigating this question. Acquisti et al., 2006 find that there is a negative impact of privacy breaches on a company's market value on the announcement day of the breach. They also provide evidence suggesting that the negative impact increases with the number of individuals affected by the breach.

[4]Employee Stock Purchase Plan

Similarly, Gatzlaff and McCullough, 2010, and Kamiya et al., 2021 demonstrate that data breaches have a negative and statistically significant impact on shareholder wealth.

The effects of data breaches on other financial aspects, such as innovation. Ettredge et al., 2018 show that firms mentioning the existence of trade secrets have a significantly higher probability of being breached compared to firms that do not. Huang and Wang, 2021 indicate that companies experiencing a reported data breach are subject to higher loan spreads and are more frequently required to offer collateral. SUN, 2021 show a positive link between being a target in M&A deals and the adoption of data breach disclosure laws at the state level. Wang et al., 2023 find that there is a positive link between the adoption of new IT technology and the risk of data breaches. He et al., 2020 demonstrate that there is a 10 percent reduction in R&D activities in the year subsequent to a data breach. This decrease is particularly pronounced in firms where R&D is not a central aspect of their business model. The research also highlights a reduction in the number of patents filed two years after the breach and an increase in cash reserves the following year.

The effects of data breaches on management and governance. He et al., 2020 find that breaches caused by system deficiencies increase the chance of a CIO's turnover by 72 percent. Zhang et al., 2024 find that companies reduce overall CEO pay following multiple data breaches, specifically in the non-cash incentive portion of the CEO's compensation. Conversely, the paper shows that companies increase the total compensation for non-CEO executives after repeated data breaches, with this increase primarily focused on the non-cash incentive component. Kamiya et al., 2021 provide evidence that a CEO's bonus and option awards decrease following a data breach. Hsu and Wang, 2014 show that board size, average age/tenure, and age heterogeneity could reduce the likelihood of security breaches, while the proportion of independent directors and tenure heterogeneity could increase it. Lending et al., 2018 show that socially responsible companies with smaller boards and greater financial expertise are less likely to be breached. Ashraf, 2022 finds that data breaches experienced by a firm in the same industry one quarter before leads to a reduction in material weaknesses for non-breached firms in the current period.

The effects of data breaches on a firm's labor force. Akey et al., 2021 show that salaries increase after a firm suffers a data breach. Bana et al., 2022 demonstrate that, following a data breach, companies tend to hire more cybersecurity and public relations professionals.

Specifically regarding industry-wide shocks, Kamiya et al., 2021 demonstrates that the cumulative abnormal returns (CAR) of industry peers are impacted as a result of an incident. Ashraf, 2022 finds that data breaches experienced by a firm in the same industry with a similar product market one quarter prior lead to a reduction in future material weaknesses for non-breached firms in the current period. Different from previous studies, my work is the first to comprehensively investigate the effects of data breaches as industry shocks on ESOP participation, rank-and-file employees, and executives.

In addition to leveraging data breach incidents to explore the effects of mutual monitoring, my work is the first to employ machine learning, offering a tractable and expandable approach for identifying distinct data breach incidents. Previous works mentioned above all use a manual approach to construct the samples they need. Given the increasing frequency and reporting of data breach incidents, my approach is highly effective for handling datasets of any size.

2 Institutional Details

2.1 Employee stock ownership plan (ESOP)

An employee stock ownership plan (ESOP)[5] is a retirement program that primarily invests in employer stock. After passage of ERISA[6] in 1974, this type of plan became widely accepted. The plan is designed to provide employees with income at retirement or termination while offering tax advantages to the company. The key feature of an ESOP is its stock bonus component. This means contributions are typically made in company stock, and benefits from an ESOP are usually distributed in whole shares of employer stock, but fractional shares can be paid in cash. Employees' ESOP ownership amount is determined by the company as a function of factors like tenure, and wage. Because an ESOP is a defined contribution plan, it is subject to all of the rules and regulations under Code 401(a). However, because an ESOP must, by definition, be a stock bonus plan, a plan that fails to qualify as an ESOP under the code, may still be qualified as a stock bonus plan.

ESOPs are designed so that employees receive more benefits the longer they work for the company. In other words, employees acquire portions of their shares over time, which are known as vested shares, representing the portion of the ESOP that employees own. The contribution of these shares must be structured to become fully vested by the time of retirement. If an employee leaves the company earlier, they will only own the vested portion of their ESOP account.

An ESOP usually takes two forms: non-leveraged and leveraged. In a non-leveraged ESOP, the company contributes stock or cash to the ESOP each year, which then holds the stock. The stock is allocated to participants' accounts based on the formula outlined in the plan. In a leveraged ESOP, a bank loans money to the ESOP through a promissory note, which is guaranteed by the company. The ESOP uses the loan proceeds to buy stock from the company or existing shareholders. The company makes contributions to the ESOP trust to repay the loan, and the ESOP trustees repay the loan according to the schedule. Finally, participants receive stock or cash when they retire or leave the company.

[5] This section is based on information from Chapter 8 of IRS documents.
[6] Employee Retirement Income Security Act of 1974

An ESOP functions through a trust fund, where companies either issue new shares, borrow funds to buy shares, or contribute cash to purchase shares. An ESOP trustee, an impartial third party, holds legal ownership of company stock and must act in the best interests of plan participants. Trustees, who can be either independent or company insiders, are selected by the board of directors and are responsible for protecting participants and improving the ESOP.

Employer contributions to an ESOP are generally deductible by the employer if they do not exceed 25% of the participants' aggregate compensation. Dividends paid regarding the ESOP's employer securities are considered earnings on account balances, and therefore do not count towards the annual addition limitation and can be deductible if paid out to the plan participant.

Employees do not pay tax at the time of contributions into the ESOP. They are taxed at the time of distributions, and the rates they are taxed on is favorable to the participant. The ESOP distributions can be rolled into an IRA[7] or other retirement plans accumulating gains over time taxed as capital gains later. ESOP distributions are taxed as ordinary income, but if employees receive a lump-sum distribution in the form of stock, they will generally pay ordinary income tax on the value of the employer's contributions to the plan, plus capital gains tax on the appreciation in stock value when the stock is sold.

In this study, I use data on ESOP firms obtained from IRS Form 5500, which employers maintaining a pension or welfare benefit plan under ERISA are required to file. I define an ESOP as any form with items '2O' or '2P' marked in the filling.[8] Additionally, I classify a plan as an ESOP if its name includes the terms 'Employee Stock Ownership Plan' or 'ESOP'. This analysis focuses on the stock bonus feature of ESOPs, rather than their qualification under section 401(a). Consequently, I do not examine the specific qualifications of the plans under the Code.

2.2 Security breach notification laws

Security breach notification laws are a set of regulations that oblige companies to take a specific set of actions in the case of a data breach incident. In the United States, security breach notification laws were first enacted in California in 2003, with other states enacting similar laws thereafter. Table 1 shows when each state implemented these laws. The first part of each regulation provides the definition of a security breach incident. For example, California defines a security breach as "an unauthorized acquisition of computerized data that compromises the security, confidentiality, or integrity of personal information." Definitions in other states are similar, though with some variations.

Generally, the law specifies the conditions under which an entity must notify victims, attorneys general,

[7] Individual Retirement Account

[8] According to IRS regulations, '2O' indicates an ESOP other than a leveraged ESOP, while '2P' refers to leveraged ESOPs.

consumer agencies, or all consumers nationwide. It provides a detailed definition of personal information and includes guidelines on the specifics of the notification, such as the details to be included in the notice and the types of services, like credit monitoring, that the company should offer to victims.

In this work, the data on data breaches comes from privacyrights.org. This dataset covers breaches reported to attorneys general and the Department of Health and Human Services from January 1, 2005, to September 28, 2023, and includes 35,167 incident reports. For my analysis, I treat all incidents in the dataset as data breach incidents, regardless of their classification under legal definitions of a security breach. For example, a physical theft of printed names and Social Security numbers of a group of consumers is included in my sample as a data breach incident. However, it is unclear whether such an incident would be classified as a security breach according to the provided definition. Therefore, I include all reported incidents in my analysis.

3 Data

This research employs two primary data sources: security breach data and ESOP data, which are discussed in detail in the following two sections. Additionally, the study incorporates data from several other sources: Compustat for firm fundamentals, CRSP for stock market-related variables, BoardEx for board information, ExecuComp and the EDGAR database for executive compensation, and Thomson for institutional ownership information. Comprehensive descriptions of all variables and their data sources are provided in the Appendix.

3.1 Data breach incidents

The data on data breaches comes from privacyrights.org, covering breaches reported to attorneys general and the Department of Health and Human Services from January 1, 2005, to September 28, 2023.[9] This dataset includes 35,167 incident reports. My goal is to identify all public firms that experienced a data breach, as well as all private firms and their corresponding SIC codes.

The main identifier for each incident in this dataset is Name of Entity. I drop all observations that include the words 'university' or 'school' in the name of the entity. This helps to delete incidents that happened in educational centers. Next, I drop all observations where all of the variables related to the date of the incident are missing.

Next, using the EDGAR API, I try to match each entity name in the dataset to its CIK, SIC, and other important identifiers. I write my queries in three steps. First, I use the full name of the entity. Second, I remove phrases such as 'Co', 'Corporation', and 'LLC' from the end of the name. Third, I remove the last

[9]This is the time of the report; there are incidents that occurred before 2005 but were reported later.

word of the entity's name. I am doing this to account for the various ways the name of an entity could have been mentioned. The result of this process includes all possible matches for a firm that has experienced a data breach to any name in the SEC dataset. For example, a company like 'Apple,' identified as a data breach company, will be matched to both 'Apple' and 'Applebees' within the SEC database. This shows that the result needs further cleaning. Nevertheless, the API result is an improvement over the raw data breach reports, as it provides the CIK and additional characteristics for a subset of the firms experiencing data breaches.

Any entity name that is not matched to SEC data falls into one of two categories: it is either not a publicly traded company or the way its name appears in the original data breach dataset is not standard. To clean up this part of 4,500 entity names(not reports), I carefully examine each entity name, removing any nonprofit, government agency, local business, or company with fewer than 100 employees based on data from Pitchbook.com and other available sources. If I can't find information about the number of employees for an entity, I remove it from the dataset.[10]

After examining the 4,500 names of entities to identify public firms and incorporating results from the EDGAR API, I end up with 530 public firms that have had at least one incident report since 2005. From the set of private firms, I keep only those with a maximum impact reported greater than 500. This results in a sample of 1,805 private firms for which I manually find the SIC code. Figure 1 illustrates the number of public and private entities in the top ten industries, ranked by the affected entities, within each 2-digit SIC code from 2005 to 2023. As the graph shows, Health Services and Business Services have the highest levels of data breach incidents, with 15% and 9% of all incidents occurring in those industries, respectively

3.2 ESOP data

Data on ESOP firms come from IRS Form 5500 and include all IRS filings from 2005 to 2023. After mapping this data to the EDGAR database and Compustat, adding all different variables, and removing observations where the number of employees reported in Compustat is lower than the number of participants in the plan, I end up with a sample of 5,397 firm-year observations covering 634 public ESOP firms from 2005 to 2023. As Figure 2 shows, Depository Institutions and Electric Services have the highest number of firms offering ESOP.

By mapping the public ESOP CIKs with the public data breach CIKs, as explained in the last part, I can determine which public ESOPs have experienced a data breach. This mapping leads to a sample of 578 data breach incidents covering 126 ESOP firms. Table 2 provides a summary of the sample selection process

[10]The number of employees is used here only as a criterion to exclude small corporations and is not related to my research question.

for identifying data breach incidents mapped to ESOP firms.

3.3 S-BERT Model

To explain why I use an S-BERT model in this study and how it benefits my research, I first provide an overview of the raw data breach incidents dataset.

As mentioned in the dataset's documentation and after reviewing the original notices of many incidents, I found that it is very likely that the same incident has been reported more than once. Finding these cases is challenging because each data breach typically involves three different dates: the day the breach occurred, the day the company discovered the breach, and the day it disclosed the incident. Even if the date the breach occurred is accurate, the reporting dates to different authorities may vary. This means that these dates cannot be used directly to identify different breaches.

Secondly, there are five different types of records impacted in the dataset, which could be reported inconsistently by different sources:

- Records Impacted From Source: The number of records impacted as indicated by the government data source.
- State Records Impacted from Source: Number of state residents impacted as per the source data.
- Total Records Impacted From Letter: Total number of records impacted as stated in the breach notification letter.
- State Records Impacted from Letter: Number of state residents impacted as mentioned in the breach notification letter.
- Max Records Impacted: The maximum estimated number of records impacted by the breach.

Varying reporting dates, and inconsistent numbers of impacts leave me with no option but to use breach descriptions to identify distinct incidents. The sample of all incidents involving ESOP firms, as detailed at the end of Section 3.2, consists of 578 incidents covering 126 ESOP firms. Table 3 represents the statistics of 578 incident reports for 126 public ESOP firms. As the table shows, each ESOP firm in this sample experiences an average of 4.59 incidents.

To identify distinct incidents based on their descriptions, I read two breach descriptions and determine whether they pertain to the same incident. The challenge is that a human can only classify two incidents as similar or not similar; a binary classification. However, this approach presents two problems: for instance, if I conclude that incidents A and B are different and also find that incidents B and C are different, this

does not necessarily imply that A and C are different. Furthermore, if I recognize that A and C are similar, which one should I retain in the final sample: A or C? This highlights the need for a non-binary measure of similarity between two data breach descriptions.

To assess the similarity of the two data breach descriptions, I use semantic textual similarity with BERT.[11] BERT is a machine learning model designed to understand each word by jointly conditioning on both left and right context. It is pre-trained on English Wikipedia(2.5 billion words) and Google's BooksCorpus (800 million words). However, I employ a sentence BERT (S-BERT) model instead of a classic BERT model. The key difference between BERT and S-BERT is that S-BERT introduces an additional pooling layer on top of BERT. This means that S-BERT models understand text on a sentence-by-sentence basis, rather than viewing it merely as a sequence of words.

I begin by utilizing a classic pre-trained S-BERT model and training it on the STS-B Multi-MT dataset, which contains approximately 5,000 sentence pairs covering various topics. This is a standard dataset used in natural language processing (NLP) tasks. In this dataset, the similarity level of each pair is rated on a scale from 1 to 5, with 5 indicating the highest level of similarity and 1 the lowest. Since the data breach descriptions don't include technical terms, this dataset is a good starting point for training the model. For the model's loss function, I use cosine similarity, where the difference between two texts is equal to the cosine of the angle between them; a higher cosine value indicates greater similarity. I adjust the similarity levels in the STS-B Multi-MT dataset to be less than one by dividing each by 5, since my loss function is cosine similarity. This preprocessing step normalizes the data before training the model.

In the next step, I fine-tune my model to specifically address data breaches. To do so, first, I identify which data breach descriptions contain the name of the targeted firm. I find these instances by searching for the name of the entity and its variations in the text of each breach explanation. Then, I create a sample of 2,000 pairs of data breach descriptions divided into two parts: the first part includes pairs from the same firm, and the second part includes pairs from two different firms, with each description mentioning the name of the firm involved. Subsequently, I use my previously trained model to predict the similarity between these 2,000 pairs. However, for pairs in the second group, regardless of the degree of similarity, I adjust the similarity measure by dividing it by 10. My rationale is that, despite any semantic similarity between two breach descriptions, if the text clearly indicates they belong to different firms, then they should not be considered similar at all.

I keep the results from these initial 2,000 pairs and then select a new sample of 3,000 pairs from the initial dataset, STS-B Multi-MT. Using this new sample, I use the same S-BERT model, treating this collection of 5,000 pairs as the new training set. I utilize the outcomes from this model to assess the similarity across all

[11] Bidirectional Encoder Representations from Transformers

potential pairs of data breach descriptions that belong to the same ESOP firm. Table 4 provides an overview of the entire training process.

Next, I find all unique pairs of incidents within each ESOP firm that reports more than one incident, excluding pairs already used in the training process. Then, I use the S-BERT model to estimate the similarity for each pair. In this work, what I refer to as the similarity measure is the result of the S-BERT model, which quantifies the similarity between the descriptions of two data breach incidents.

Table 5 displays the model's predictions for the similarity between the two data breach texts. It's important to remember that this similarity measure represents the cosine of the angle between the two texts' vectors. In other words, a mean of 0.31 in Table 5 indicates that for an average firm in my dataset, which has more than one data breach reported, if I select two of these data breach descriptions and generate their associated vectors, then the cosine of the angle between those two vectors is 0.31. I utilize this similarity measure to create an index for distinguishing between two incidents.

3.3.1 Finding distinct incidents

As previously discussed, my main goal was to find an efficient, consistent, and traceable method to identify distinct data breach reports for each firm. To achieve this, I focused on automating the manual process I used to assess the differences between incidents. Based on my experience, I evaluate whether two incidents are distinct by examining their text and timing. In the previous section, I developed a method to assess the similarity between two data breach descriptions. In this part, I incorporate a time variable into the similarity assessment and create a distinction index.

To create the index, I first add the pairs of data breach incidents that were excluded during fine-tuning to the rest of the pairs. Table 6 presents summary statistics for the time difference between two data breach incidents within the same firm. it is worth noting that a "pair" refers to two data breach incident reports from the same firm. The mean time difference of 961.5 days indicates that, on average, the absolute time difference between two incident reports within the same firm is 961.5 days.

Next, I divide the time gap by the similarity measure. The formula for the index is as follows:

$$Distinction\ Index = \frac{Time\ gap}{Similarity\ measure}$$

Or, formally, the Distinction Index for a pair$(b_t, b_{t'})$ of two data breach incident reports, with one occurring at time t and the other one at time t', is defined as:

$$Distinction\ Index(b_t, b_{t'}) = \frac{|t - t'|}{Similarity\ measure(b_t, b'_t)}$$

My rationale is that two data breach incidents with closely reported dates but low similarity in their descriptions are likely to be different incidents. Since the similarity measure is a number less than one, dividing the time gap by this measure increases the value of the time gap. Therefore, a larger distinction index indicates that the two incidents are more distinct. The only exception is when the index is negative, which means the similarity measure is negative. This implies that the two vectors are very divergent. Thus, regardless of the time gap, a negative distinction index indicates that the two data breaches are distinct.

Finally, to identify distinct incidents, I sort the reports within each firm from earliest to latest. An incident is considered distinct if its distinction index with the next report is a local peak on the positive side or a local minimum on the negative side of the graph. In other words, for four subsequent incidents $b_{t-1}, b_t, b_{t+1}, b_{t+2}$ at times $t-1, t, t+1, t+2$ incident b_t is recognized distinct If:

$$Distinciton\ index(b_t, b_{t+1}) > 0$$

$$and$$

$$Distinciton\ index(b_{t-1}, b_t) < Distinciton\ index(b_t, b_{t+1})$$

$$and$$

$$Distinciton\ index(b_t, b_{t+1}) > Distinciton\ index(b_{t+1}, b_{t+2})$$

OR

$$Distinciton\ index(b_t, b_{t+1}) < 0$$

$$and$$

$$Distinciton\ index(b_{t-1}, b_t) > Distinciton\ index(b_t, b_{t+1})$$

$$and$$

$$Distinciton\ index(b_t, b_{t+1}) < Distinciton\ index(b_{t+1}, b_{t+2})$$

In other words, an incident b_t is categorized as distinct if its distinction from the subsequent incident is a local maximum on the positive side of the graph and a local minimum on the negative side. For example, Figure 3 shows Ameriprise Financial Inc.,[12] with 67 reports since 2010. Using my method, I select the first

[12]Formerly American Express Financial Corp.

report, all peaks in the middle, and the one before the last report,[13] resulting in 21 distinct incidents over 13 years, marked in red. Table 7 lists the identified distinct incidents for Ameriprise, which appear to be unique, suggesting the effectiveness of the algorithm. After applying this method to the entire sample, I identify 231 distinct incidents across 126 public ESOP firms. As Table 8 shows, after identifying distinct incidents, each ESOP firm experiences, on average, fewer than 2 data breaches in this sample. Before applying the method, the average number was over 4, as shown in Table 3, which is relatively high.

Applying this approach to the sample of 969 reports of public non-ESOP firms-404 distinct firms- results in 504 distinct incidents. The primary advantage of this method is that it yields results without making assumptions about the distribution of distinct incidents. Figure 4 illustrates the distribution of reports before and after utilizing the mentioned method for both ESOP and non-ESOP firms. As shown in the figure, the method preserves the initial distribution of reports, indicating that it does not introduce bias towards specific years' reports.

In summary, I started with a sample of 578 data breach incident reports across 126 ESOP firms and identified 231 distinct incidents. I use these distinct incidents to determine firm-year observations with or without a data breach incident in my analysis. In summary, I began with a sample of 578 data breach incident reports across 126 ESOP firms and identified 231 distinct incidents. I use these distinct incidents to determine firm-year observations with or without a data breach incident in my analysis.

4 Results

4.1 Likelihood of experiencing a data breach incident

I hypothesize that firms with higher levels of employee ownership will experience increased peer monitoring or self-control, which should lead to a lower likelihood of data breaches. Unlike prior literature, which uses dollar value measures, I focus on the number of employees classified as 'owners.' By definition, an active participant is an employee who works for the company and has contributions made on their behalf to the ESOP plan. Therefore, the primary proxy for employee ownership is the ratio of active participants in the ESOP plan to the total number of plan employees. This variable is intended to capture the effect that each rank-and-file employee, as an owner, can have as a monitor. The higher the number of active participants relative to the total number of employees, the greater the monitoring.

To start with, I identify the fiscal year corresponding to each distinct incident identified previously. Then, I compare the characteristics of firm-years that lead to an incident with those of firm-years that

[13]The distinction index is not defined for the last report, as it does not have a subsequent incident, and therefore it is not shown on the graph.

do not experience any incidents afterward. It is worth mentioning that I do not impose any condition on the number of incidents per year, so it does not matter how many incidents are assigned to the same year. Table 9 presents summary statistics for 125 firm-year observations with an incident (83 distinct public ESOP firms) and 4,255 firm-year observations from 551 distinct ESOP firms that never experience a data breach over the period 2005 to 2023. For my analysis, I select three groups of variables: first, variables related to rank-and-file employees, including ESOP plan participation and wages; second, variables related to executive compensation and its different elements; and third, variables used by Kamiya et al. (2021), which include firm fundamentals and governance characteristics. All continuous variables are winsorized at the 1st and 99th percentiles.

According to Table 9, firms experiencing a data breach incident are larger and have a greater presence among Fortune 500 companies. This means that these companies stand out more. Another significant difference is that firms with an incident have a higher number of active participants in their ESOP plan; however, the ratio relative to the total employee population is not significantly different. These firms also have higher levels of ESOP plan assets and wages paid to employees, which is not surprising given their larger size. In the group of executive compensation, executives in firms experiencing a data breach receive a lower portion of their compensation as salary and a higher portion in stock and non-equity incentive compensation.

Furthermore, ESOP firms with an incident tend to have lower levels of R&D expenses, less financially constraint and are from industries with a higher Herfindahl index. This higher Herfindahl index indicates that these firms are in industries with lower levels of market competition. To define a financially constraint firm, I use the approach developed by Whited and Wu, 2006. I construct the WW index as a function of cash flow, dividends, long-term debt, total assets, industry sales growth, and firm sales growth. Firms are then ranked based on this index; the higher the index, the more constrained the firm is.

In the group of corporate governance characteristics, I find that the proportion of firms with a risk committee on the board is not significantly different between the two groups of firms in my sample. A board is considered to have a risk committee if the term 'risk' is included in its name. However, ESOP firms experiencing a data breach tend to have a higher number of board committees.

Table 10 presents the correlation between the main variable of this study, the active ratio, and a group of other important variables. Figure 5 shows the distribution of this variable in the sample. Except for 160 observations with a ratio close to one, the remaining values are spread out relatively evenly.

Table 11 presents the distribution of industry of firm-years with at least one incident from 2010 to 2023. As the table shows, among ESOP firms, the highest number of data breach incidents occurs in the finance industry, followed by manufacturing and wholesale trade. Additionally, the overall trend of incidents over time is increasing. A key point of this table for my analysis is that it provides evidence that data

breach incidents in this sample are not restricted to specific industry-year shocks. Instead, each industry can experience incidents in different years, and various industries experience incidents in each year. Table 12 presents the distribution of targeted firm-year observations across each state. This table is important for my analysis because it demonstrates that data breach incidents in this sample are not restricted to specific state-year shocks.

To directly investigate the role of ESOP ownership in the likelihood of becoming a target of a data breach, I estimate the following probit model:

$$\begin{aligned} P(\text{Data breach in ESOP firm } i \text{ in year } t=1) = \alpha + \beta_1 \left(\frac{\text{Active participants}}{\text{total employees}}\right)_{i,t-1} + \\ + \beta_2 \log\left(\frac{\text{ESOP}}{\text{active participants}}\right)_{i,t-1} + X_{i,t-1} + \\ + Industry\ FE + Year\ FE + \epsilon_{i,t} \end{aligned} \tag{1}$$

Based on my hypothesis, the higher the active employees ratio, the better the monitoring or self controlling, and the lower the likelihood of a data breach. In other words, I expect to see a negative sign for β_1 in Equation 1. I also add the average ESOP assets each employee has. This is supposed to capture the strength of employee ownership; ideally, with no free riding, the higher the employees' ownership, the greater their incentive to monitor their peers, and the lower the probability of a breach. In other words, I expect to see a negative sign for β_2 as well. Other control variables, shown by $X_{i,t-1}$ are variables from Table 9. I also add industry and year fixed effects. All variables, except Tobin's q which is estimated for two years prior,[14] are measured for one year prior to the breach; if a breach happens in year t, control variables from year $t-1$ are used on the right side of Equation 1.

Table 13 reports results of estimating Equation 1. I start by regressing the breach indicator on the active ratio alone; the coefficient is negative. In Regression (2), I drop the active ratio and only include executive equity compensation, defined as the mean of equity compensation—stock plus options—for each executive relative to their total pay and wage, along with the other independent variables. I find that firms that stand out—such as those with larger sizes and Fortune 500 membership—are more likely to be targeted by data breaches. A positive coefficient on ROA implies that data breach incidents are more common in firms with higher profitability. A negative coefficient on Q suggests that breached ESOP firms have lower future growth opportunities.

In Regression (3), I include only ESOP-related variables and drop executive compensation. As the table shows, both active participants and the ESOP ownership value are negatively correlated with the probability

[14] Because it is highly correlated with past stock performance.

of a data breach. In Regression (4), I include both the set of ESOP variables and executive compensation together. It is important to include both executive compensation and ESOP variables together, as executives might also be participants in the ESOP. This approach will control for any potential correlations between ESOP participation and executive pay.

In Regression (5), I include the interaction term of the active participants ratio and ESOP ownership to better separate their individual effects. As the table shows, both variables remain negative. Based on Table 10, the active participants ratio and executive equity compensation are negatively correlated, so it is possible that the effect of the active ratio is masked by the executive compensation. Therefore, in Regression (6), I add the interaction term of these two variables. As the table shows, the individual effect of the active participants ratio is negative; however, it becomes weaker as the executive equity increases.

Using similar reasoning to Regression (6), I include the interaction term of firm size and the active ratio in Regression (7). Larger firms often have more external relationships, such as with contractors or vendors, which can escalate the risk of the firm's information being leaked. As Regression (7) shows, a higher active ratio is associated with a lower probability of a data breach; however, firm size can weaken this protective role of ESOP employees.

In Table 14, I include variables that, based on the literature and intuitive reasoning, could influence broad-based stock option offerings by firms. This table presents only the coefficients for ESOP plans, employee and executive compensations, and corporate governance variables. All other control variables are included as detailed in Table 13, but their coefficients are not presented in this table. Regression (1) is the same as model (6) in Table 13 and is mentioned as a reference.

In Regressions (2) and (3), I include Industry-Year fixed effects to account for trends such as layoff waves across different industries. The threat of layoffs can impact employees' monitoring quality or even motivate them to take action against the firm they are about to leave. I find that the coefficient on the active ratio remains negative. Notably, the negative sign on employees ESOP ownership level and the positive sign on the wage variable suggest that firms adjust compensation for rank-and-file employees based on industry conditions. This also supports the idea that higher equity compensation strengthens employee monitoring. Another notable point in this model is the negative sign on the risk committee indicator. This suggests that firms establish a risk committee within their board to adapt to industry conditions, and this committee could be effective in reducing the likelihood of a data breach.

In Regressions (4) and (5), I include State-Year fixed effects to account for the possibility that firms adjust their stock offerings to remain competitive in the local labor market. For example, Kim and Ouimet (2014) use a variable measuring worker bargaining power to determine which ESOP firms are likely to experience higher performance. This variable is based on the number of employees in a firm's establishments within

the same industry that are located near each other. Similarly, Hochberg and Lindsey (2010) use the average non-executive incentive compensation across other firms in the same two-digit zip code to explain stock option offerings within a firm. Both studies support the idea that firms offer broad-based stock options to adapt to local labor market conditions. It is worth mentioning that ESOPs are not options, they are stock bonus plans. However, if companies use stock option plans to adjust their employees' compensation based on the local labor market and for retention, they can apply the same reasoning to use ESOPs or other stock bonus plans. Therefore, I use the geographical variables inspired by stock options literature. As Regression (4) shows, the state-year fixed effect is not strong enough to separate the individual effect of the active participants ratio. However, after accounting for the weakening effect of executive compensation in Regression (5) by adding the interaction term, I find that the individual effect of the active ratio is negative with a greater absolute value than in Regression (1), where there was no state-year fixed effect.

In Regression (6), I include State fixed effects in addition to state-year fixed effects. This is done for two reasons. First, to control for security breach notification laws, which were first adopted in California in 2003 and, by 2010, were implemented in most states. Although this is less of a concern in my sample, it is still addressed. The primary reason for including state fixed effects is to account for fundamental differences between states, such as religious beliefs and risk-taking behaviors. For example, Kumar et al. (2011) demonstrate that broad-based employee stock option plans are more prevalent in regions with a higher Catholic-to-Protestant ratio. Overall, adding State fixed effects does not change the effect of the active ratio on the probability of a data breach.

One point worth mentioning about Table 14 is the decrease in the number of observations after adding new indicators. Since I am running a probit model with many zeros (failures) as the dependent variable, many of these zeros will be perfectly predicted by the new indicators and thus removed from the model during estimation. The overall comparison between the variables in Tables 8 and 9 shows that this smaller sample does not affect the estimated coefficients. For example, the coefficient of the active ratio in Regression (4) of Table 14 is very similar to the coefficients in Regressions 13.

Before concluding this section, I closely investigate the effect of executive equity on the likelihood of a data breach. The previous table have demonstrated that higher executive equity is associated with a higher likelihood of a data breach. Zhang et al. (2024) and Kamiya et al. (2021) show a decrease in CEO pay following a data breach incident. This suggests that firms may already recognize how executives' risk-taking behavior can increase the firm's risk appetite, which, in turn, leads to a data breach. Therefore, they decide to adjust executive compensation after a breach. However, executive compensation is highly correlated with firm size, so the positive relationship between executive equity and the likelihood of a breach requires further investigation.

Table 15 presents the results of this investigation. Regression (1) includes the interaction term between executive equity and the active ratio, showing a positive effect of firm size and the interaction of the active ratio with executive equity on the likelihood of a data breach. In Regression (2), I introduce the interaction between firm size and the active ratio. This term accounts for part of the significant effect observed in the interaction between executive equity and the active ratio. Consequently, only the active ratio exhibits a negative correlation with breach probability, while other key coefficients are not significantly different from zero.

Regression (3) incorporates industry-year fixed effects into the previous model, none of the interaction terms are significant. The value of ESOP per participant stays negative. This variable, due to its correlation with executive equity, captures some of the protective effects of executive equity that could have been otherwise missed. The results from this regression support the notion that the effect of executive equity compensation on breach likelihood is primarily a size effect.

In Regression (4), I retain the interaction between firm size and the active ratio along with industry-year fixed effects. This model does not provide additional insights beyond those obtained in the previous regression. Regressions (5), (6), and (7) include the interaction term between firm size and executive compensation. Regression (5) includes only the interaction term between firm size and executive equity without industry-year fixed effects. The table shows that while the individual effects of firm size and executive equity are positive, their interaction term is negative, which contradicts expectations. This negative effect disappears when controlling for industry-year fixed effects in Regression (6). Finally, Regression (7) includes all interaction terms simultaneously. The results indicate a positive effect of executive equity compensation, but when all the interaction terms are included together, none of them turns out to dominate the others.

Next, I investigate the effects of ESOP ownership on the probability of a data breach from both an extensive and intensive margin perspective. The intensive margin refers to the effect of the value of ESOP assets per employee on the data breach incident, while the extensive margin refers to the effect of the active ratio (ESOP employees per employee) on the incident. This idea is driven by the following equality:

$$\frac{ESOP}{employees} = \frac{Active}{employees} \times \frac{ESOP}{Active}$$

where ESOP represents the total assets of a firm's ESOP plan and employees refer to the employee population.

In Table 16, I divide the data into three quantiles and estimate a quantile regression model to analyze the relationships across those quantiles. In Regression (1), I sort the data based on $(ESOP)/emp$ each year and then examine the effect of the active ratio on the probability of an incident. If the role of active participants is completely driven by the ESOP value, then the higher quantile of $log(ESOP)/emp$ should

have a lower probability of a data breach relative to the lower quantile, and the active ratio should not have any effect. However, Regression (1) demonstrates the opposite. In other words, among firms with similar levels of $log(ESOP)/emp$, a higher active ratio can effectively protect the firm from a data breach.

In Regression (2), I sort the data based on the active ratio in each year and examine the effects of $log(ESOP)/emp$ on the probability of a breach. First, I observe a negative coefficient for the highest quantile of the active ratio relative to the lowest. Second, I do not see any effect of $log(ESOP)/emp$ on the likelihood of an incident. The results of columns 1 and 2 prove that the effects in the prior tables are not driven by size or ESOP value. In other words, two firms with similar levels of $log(ESOP)/emp$ can have different exposures to an incident based on their active ratio. This table provides insight into why the distribution of ESOP assets within the firm matters, showing that it is more than just an average value.

Overall, the results in this section suggest that a higher ratio of ESOP participants to total employees is associated with the likelihood of a data breach due to the peer monitoring and self-control effects that owner employees can have. This value holds in the presence of the ESOP value per participant, which is also negatively associated with the probability of a data breach. Moreover, two factors can weaken the monitoring effect: executive equity and firm size. Additionally, I provide evidence that the distribution of ESOP assets within the firm is effective in protecting against data breach incidents. In other words, two firms with the same level of average ESOP ownership among all employees can have different levels of exposure to data breach incidents, depending on their different active ratios.

4.2 Effects of a data breach on ESOP participation

In this section, I examine what happens to a firm's ESOP participation following its first data breach incident. Based on my hypothesis, if firms use ESOPs as a way to impose peer monitoring after a data breach, the number of active participants relative to the firm's employee population in the plan should increase compared to the period before the breach.

In the sample 88 ESOP firms have experienced a data breach incident at least once. For each treatment firm, I use propensity score matching to identify a control firm that has not experienced a data breach. The propensity score is calculated using the logit regression of data breach (an indicator that takes the value of one if a firm experiences an incident in a given year, and zero otherwise) on firm size, stock performance, stock return volatility and leverage. I also require that the both treated and matched firm come from the same fiscal year.

The control group is chosen based on the closest propensity score without replacement from the set of all firms that have never experienced a data breach between 2005 and 2023. The perfectly matched control

group for my sample would be a firm from the same industry, with each control firm chosen only once. However, this will lead to a small set of matched firms. Instead, I do the matching without replacement for firm-year pairs with no industry criteria and will control for industry fixed effects later in regressions.

Based on these matching criteria, my model identifies matches for 69 data breach firms, which consist of 59 distinct control firms. Table 17 presents descriptive statistics for the matched sample. I find no significant difference between firms with an incident and their matching non-breached counterparts, suggesting that the matching approach identifies control firms that are very similar to the treatment firms. To create the final difference in difference sample, I include all years of data for the treatment firms, as well as all years of data for any firm that is chosen as a control firm at least once in the matching process. To ensure a reasonable number of observations after the first breach, I exclude firms that experienced their first breach after 2020. This results in a sample of 1,667 ESOP firm-year observations from 2005 to 2023.

Since the units do not get treated all at the same time, I need to define a group unit based on the treatment timing. In my work, one group consists of all firms that have been targeted by a data breach incident for the first time in the same year, with the year ranging from 2005 to 2020, or firms that have never been targeted. With this definition, the structure of this experiment is summarized as follows:

- **Treatment**: the first data breach incident in a firm
- **Treatment group**: All firms targeted by a data breach for the first time in the same year(2005-2020)
- **Control group**: The closest propensity score without replacement from the set of all firms without a data breach between 2005-2020.

I estimate the following difference-in-difference model:

$$Active\ ratio_{itg} = \alpha + \beta D_{g,t} + \alpha_i + \alpha_t + \epsilon_{itg} \tag{2}$$

where $Active\ ratio_{i,t,g}$ equals the number of active employees in the ESOP plan relative to the total number of employees for firm i in fiscal year t which belongs to treatment group g (i.e., all firms experiencing their first data breach in year g or never). $D_{g,t}$ is an indicator that equals one for the time period after the treatment for each treatment group g. In other words and generally, $D_{g,t}$ equals $Treat_g Post_t$, an indicator that equals one if $t >= g$ and the observation belongs to treatment group g. It is zero otherwise. α_t is fiscal year fixed effect, α_i represents the firm fixed effect, $\epsilon_{i,t,g}$ is the error term.

Since each firm belongs to only one treatment group, I do not include any treatment group fixed effects here, as they would be redundant after including firm fixed effects. The parameter of interest in Equation (2) is β, which represents the difference between the change in the treatment group and the change in the control

group relative to the pre-treatment period. If my hypothesis holds, I should observe a positive value for β, indicating that firms experiencing a data breach will have a greater change in their active ratio compared to firms with no incident in the sample.

A key point worth mentioning before estimating Equation 2 is that the classic structure of difference-in-difference assumes that once a firm is treated, it remains treated afterward. This structure might be a good fit for policy settings where agents adopt a policy and are expected to commit to it. However, it may not be a reasonable assumption in the context of my work. In this sample, some firms have observations for up to 11 years after experiencing their first incident. Under the classic assumption, this implies that firms will remember a breach from 11 years ago and continue to act based on it, which is highly unlikely. One straightforward way to address this issue is to exclude any treatment observations beyond a specific time in the sample. While this approach reduces the sample size, it also removes potentially useful information. Lin and Zhang (2022) extends a TWFE model in which distant periods relative to the treatment time are assumed to have a zero treatment effect. Wooldridge (2023) similarly proposes that, in a difference-in-differences framework, treatment groups can exit the treatment. While a treatment effect can be estimated for all time periods, it may not align with what the researcher intends to define as a treatment effect. Therefore, rather than excluding observations from model, I exclude time periods from the treatment.

At first, I start by the breach period and one period after it. Table 18 shows the results from estimating Equation 2. Regression (1) shows that the coefficient on the active ratio, assuming that the treatment takes effect one period after the first breach, is positive. This finding aligns well with my hypothesis. In Regression (2), I assume that the treatment stays on for five periods after the first breach and then turns off. As Table 18 shows, the key coefficient remains positive. A key point about Regression (2) is that the difference between the treatment and control groups becomes insignificant starting from two periods after the breach, and it remains insignificant till five periods after the incident. This suggests that five periods is an appropriate cutoff to assume the treatment turns off, and any subsequent periods can be considered as non-treated for the treatment group.

In Regression (3), I assume the treatment remains on permanently throughout the sample. In Regressions (4), I add state-year fixed effects to the specifications in column (2). As mentioned previously, these fixed effects, in line with the prior literature, are intended to account for the impact of local labor markets on offering stock bonus or stock option plans to employees. As the table shows, the results are not significant anymore. In Regressions (5) and (6), I add industry-year fixed effects to the specification. As the results show, the positive coefficient is no longer significant. These results imply that the difference between the treatment and control groups is largely captured by state-year and industry-year fixed effects.

In Table 19, I estimate Equation 2 to identify the effect of the incident on firms' other labor force variables.

In Regression (1), I examine the change in the average value of ESOP ownership per participant and find a negative change. Next, in Regression (2), I examine the change in the employee population, and it does not show any significant change. In Regression (3), I study the change in the number of active participants in the ESOP and find that the change is closer to a decrease or no change. In Regression (4), I look at the change in wages after the breach and cannot find a significant change.

In Table 20, I analyze the changes in executive compensation elements following a data breach. I begin by examining the equity portion of the compensation, and as the table shows,there is no significant effect. In Regression (2), I add industry-year fixed effects but still do not observe any effect. In Regressions (3) through (5), I examine changes in salary, bonus, and non-equity incentives but find no significant changes, except for an increase in bonus in period three after the first shock.

To establish causal inference for the negative coefficient of the value of ESOP ownership per participant, I need to verify if parallel trends hold. In other words, I must demonstrate that, in the absence of the first breach, the treatment firms would have followed the same trend as the control firms. It is not shown here, but there is a negative trend before the shock, indicating that parallel trends do not hold.

Overall, this section provides evidence that the variation in firms' reactions to a data breach is largely captured by state-year and industry-year conditions. This serves as a segue into the next section, where I demonstrate that the first noticeable breach in each industry leads to significant changes at the firm level.

4.3 Data breach incidents as industry-wide shocks

"The best way to get management excited about a disaster plan is to burn down the building across the street."
Dan Erwin, Security Officer, Dow Chemical Co

In this section, I examine how the active ratio changes after the first data breach in the firm's 4-digit SIC industry. I expect that ESOP firms will increase their active ratio after the first noticeable breach in a public or private firm within their industry.

When a firm decides to respond to an external data breach, it does not matter whether the breach occurred in a public or private firm. What matters is that the threat is serious enough to prompt action. Therefore, all public and private firms with reported incidents are included in the sample. To ensure the breach is noticeable, I focus on incidents where the number of affected impacts exceeds 500. I then keep the earliest breach in each 4-digit SIC code, meaning an industry shock is defined as the first time a public or private firm in the industry experiences a breach with at least 500 impacted cases. Table 6 shows the distribution of the first incident in each of the 403 4-digit SIC codes of the breaches dataset. For example, in 2016, around 50 industries experienced a data breach with more than 500 affected individuals for the first

time. Because I need observations for both before and after the shock for each firm, I focus on industry shocks occurring after 2005 and before 2020 in the sample. Matching the shocks with ESOP firms results in a sample of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. Table 21 presents the number of observations and industries in each treatment group.

In summary, the structure of the experiment is as follows:

- Experiment: How does the active ratio change after the first data breach in the firm's 4-digit SIC industry?
- **Treatment**: The earliest breach in each 4-digit SIC code- public or private firm- with at least 500 impacts.
- **Treatment group**: All firms belonging to industries experiencing their first noticeable data breach in the same year (2005-2022), 75 4-digit SIC codes
- **Control group**: 962 observations from industries with no noticeable data breach incident, 70 4-digit SIC codes

To find the average treatment effect, I estimate a staggered difference-in-differences regression. To establish causal inference, I need to account for the fact that assigning firms to treatment and control groups might not be random; certain firms or industries may have a higher likelihood of being targeted by a data breach. To address this issue, I first exclude any firm that has ever been a target of a breach from the sample. Second, I include firm fixed effects in the estimation to control for time-invariant unobservable characteristics. Then, I run the following difference-in-differences regression:

$$Active\ ratio_{itg} = \alpha + \beta D_{g,t} + \alpha_s \times \alpha_t + \alpha_t + \alpha_i + X_{it} + \epsilon_{it} \tag{3}$$

where $Activeratio_{i,t}$ equals the number of active employees in the ESOP plan relative to the total number of employees for firm i in fiscal year t which belongs to treatment unit g (i.e., all firms belonging to industries experiencing their first data breach in year g or never, each firm i is assigned to only one g).

$D_{g,t}$ equals $Treat_g \times Post_t$; an indicator that equals one if firm i belongs to an industry that experienced its first breach in year g and $t >= g$. α_t is fiscal year fixed effect, α_i represents the firm fixed effect, $\alpha_s \times \alpha_t$ shows state-year FE and controls for local labor market conditions, X_{it} represents a group of control variables for each firm that are unlikely to be affected by the shock, and $\epsilon_{i,t}$ is error term.

Since each firm belongs to only one treatment group, I do not include any treatment group fixed effects here, as they would be redundant after including firm fixed effects. The parameter of interest in equation 3 is β, which shows the change in the active ratio following the first noticeable data breach in the industry relative to the change in the control group firms over the same time period. I expect to see a positive sign for the β here.

Table 22 presents the results of estimating Equation 3. If I do not impose any restrictions on the timing of the first breach within an industry, I end up with time horizons extending up to 15 years after the initial breach. Similar to the previous section, assuming that treatment remains active for 15 years is unreasonable in this context. Therefore, in Panel A, Regression (1), I limit the sample to observations within five years after treatment and exclude any longer time horizons. As shown in Regression (1), firms within the industry exhibit a 3% increase in active ratio relative to the control group during periods two to five.

In Regression (2), I analyze the same sample but include all the periods after the shock but assume that the treatment effect goes off after 10 periods. As shown in the table, firms within the industry exhibit no significant reaction during the first two periods (period zero and period one). However, beginning in period two and continuing through period five, the cumulative effect is positive. In Regression (3), I include all observations but assume that the treatment effect turns off after fifteen periods. The critical insight here is that starting from period six, the difference between the treatment and control groups begins to diminish. In other words, Regressions (2) and (3) suggest that five periods is a reasonable time horizon to assume that the treatment turns off without having to drop observations from the model. Therefore, in Regression (4), I assume the treatment remains on for five periods, then turns off. As the table illustrates, firms in industries experiencing a notable data breach for the first time increase their active ratio by 3% more than the change in the control group's active ratio over the same period. It is important to note that none of the firms in this sample have actually experienced a data breach and they are reacting to an external incident.

In the remainder of the table, I aim to address two questions: First, what is driving the change in the active ratio? Second, how do other related variables, such as wages, change after the shock? For all dependent variables, I assume the treatment effect fades after five periods. Additionally, since I am interested in the dynamics, I present the changes over each of the five periods following the treatment rather than a cumulative value.

As Panel B of Table 22 shows in Regressions (1), the change in ESOP value per participant is affected by the shock negatively after one period. The next variable, the change in employee population in columns (2), does not show a significant change. In Regressions (3) of Panel B, I examine the change in the number of active participants in the ESOP plan after the shock. As the results show, there is a 7% increase in the treatment group relative to the control group after two periods, and this increase persists thereafter. In

Regression (4), I examine the change in wages within the sample. I find a 5% decrease in wages in period 3, which reduces to 3% in the next period. Regression (5) shows no evidence of a change in the value of ESOP assets.

In sum, it appears that the 3 percentage points change in the active ratio demonstrated in Panel A is driven by an increase in the number of active ESOP employees after a data breach, which also leads to a decrease in the ESOP ownership value for each employee.

Next, I investigate the change in executive compensation of peer firms after the shock in the industry. Table 23 shows the change in different elements of executive compensation following the first breach in the industry. Regression (1) shows no change in the executive equity portion of compensation (the sum of stocks and options) during the first period and the following period after the breach. Regression (2) shows an increase in salary during the first and second period after the shock. Regression (3) shows no change in bonuses. Regression (4) shows a decrease in non-equity incentives from two years after the breach until three years after the breach.

The key identification assumption in a difference-in-differences model is common counterfactual trends between the treatment and control groups; i.e. in the absence of the intervention the treatment and control groups would have experienced the same changes in outcomes. Therefore, I estimate Equation 4 to show the trends before and after the shock.

$$Active\ ratio_{itg} = \alpha + \sum_{e=-4}^{-2} \delta_e \cdot D^e_{g,t} + \sum_{e=0}^{5} \beta_e \cdot D^e_{g,t} + \alpha_s \times \alpha_t + \alpha_t + \alpha_i + X_{it} + \epsilon_{it} \tag{4}$$

where $Active\ ratio_{i,t,g}$ equals the number of active employees in the ESOP plan relative to the total number of employees for firm i in fiscal year t which belongs to treatment unit g (i.e., all firms belonging to industries experiencing their first data breach in year g or never). $D^e_{g,t}$ is an indicator for firm i which belongs to treatment group g being e periods away from the first breach in its industry at time t. α_t is fiscal year fixed effect, α_i represents the firm fixed effect, $\alpha_s \times \alpha_t$ shows state-year FE and controls for local labor market conditions, X_{it} represents a group of control variables for each firm that are unlikely to be affected by the shock, and $\epsilon_{i,t}$ is error term. A non significantly different from zero coefficient for δe, 2 to 4 periods before the treatment, indicates that the change in the active ratio in the treatment and control groups would follow the same trend when treatment is absent.

Figure 7 illustrates the dynamics of change in the significant coefficients estimated above. Except for an increasing trend in the value of ESOP ownership per participant, other variables do not exhibit a significant trend before the shock. In the case of ESOP ownership value per participant, there was an increasing trend before the shock, but the direction completely reverses after the shock, shifting from positive to negative. The

absence of a trend before the shock, combined with controlling for time-invariant unobservable characteristics, suggests a causal effect of the first noticeable industry shock on the active ratio, increasing it by 3 percentage points.

As Table 21 shows, a big portion of the sample experiences its first data breach around 2007 and 2008, coinciding with the financial crisis. To investigate whether the previous results are driven by the financial crisis shock or financial firms, I re-estimate Equation 3 for three subsamples: First, I exclude all firms that experienced a shock in 2007 or 2008. Second, I exclude all financial firms—those with SIC codes starting with 6—from the sample. Third, I exclude both the industry shocks of 2007 and 2008 as well as all financial firms. As Tables 24, 25 and 26 show, the increase in the active ratio and ESOP employees after the shock persists in the sample, even after excluding financial shocks or financial firms. Finally, in Table 27, I examine the changes in financial firms experiencing the shock in 2007 or 2008. As the results show, the decrease in ESOP ownership per participant is substantial, driven by a decline in the value of ESOP assets after 2007 and 2008. However, even in this case, firms increase the number of ESOP owners, thereby raising their active ratio.

Overall results of this section aligns well with Boasiako and Keefe, 2021, which find that the implementation of compulsory disclosure laws related to data breaches is associated with an increase in cash reserves held by companies. In other words, these laws prompt firms to take the threat of a data breach more seriously. A similar effect can occur after a noticeable data breach within the same industry. To respond to this threat, firms increase the number of ESOP owners by 7 to 10 percent and raise their active ratio by 3 to 4 percentage points. In the case of financial firms experiencing a shock during 2007 and 2008, there is a significant decrease in the value of ESOP assets, wages, and the ESOP ownership value per participant.

5 Robustness check

In this part, I re-estimate the models from Table 14 for two samples: all incident reports and the first report for each firm. As shown in Tables 28 and 29, the results are not affected by this change in the sample. The only point to note is that, relative to the sample of distinct incidents, these samples cover a slightly higher number of distinct firms because some firms do not have reported breaches after 2010 if their reports are not identified as an incident due to their similarity with reports from before 2010. In Table 30, I use the sample of distinct incidents but add lags of breach incidents and the active ratio as control variables. As the results show, the coefficient on the active ratio is not affected by these variables. In Table **??**, I incorporate a trend term into the regression to control for potential upward trends in both the active ratio and the probability of a data breach. In Table **??**, I re-estimate the results using a normalized value of the active ratio. As

demonstrated in the tables, the pattern of the key coefficient and its association with data breach likelihood remains consistent.

6 Conclusion

Employees have a comparative advantage in monitoring their peers and other employees, an advantage that managers and executives do not possess. This motivation can be strengthened by offering employees company stock, making them partial owners of the company. Such an approach, which encourages employees to be more careful with both their own actions and those of their peers, can be particularly helpful in situations where a small mistake by one person could lead to significant losses for the entire firm. I refer to this mechanism as the mutual monitoring channel in this work, and a data breach incident is one such O-ring problem for a firm.

To test whether mutual monitoring is effective in preventing data breach incidents, I focus on firms with employee stock ownership plans (ESOPs). Using the ratio of active ESOP participants to the firm's total employee population, termed the active ratio, as a monitoring measure, I find that a higher active ratio is associated with a lower likelihood of data breaches. However, this effect diminishes with firm size, and executive equity appears to have the opposite effect.

Moreover, by examining the extensive and intensive margins of ESOP ownership, I find that two firms with the same level of ESOP value per employee can have different levels of protection from data breaches due to variations in their active ratio. I refer to this as an extensive vs. intensive margin analysis of ESOP ownership; I provide evidence that the distribution of ESOP assets within the firm is effective in protecting against data breach incidents.

Second, I examine the change in a firm's active ratio following its first data breach incident. The increase in the active ratio after the shock is explained by state-year and industry-year fixed effects. Third, using a staggered difference-in-differences structure, I find that following the first noticeable breach in the industry, ESOP firms increase their active ratio by 3 to 4 percentage points. This effect is primarily driven by a 7 to 10 percent increase in the number of ESOP owners in the firm.

Notably, this change remains persistent after excluding financial firms and industry shocks coinciding with the years 2007 and 2008. However, in financial firms, there is a significant decrease in wages and the value of ESOP ownership per participant after the shock. Combining all my findings, I provide evidence that firms maintain a higher active ratio in response to industry data breach shocks, which helps protect them against future data breach threats.

Tables

Table 1: Timing of the first implementation of security breach laws in the U.S.
The table shows the year in which each state first enacted a security breach law. The group on the right represents the states where ESOP firms with an incident in my sample are located.
Source: perkinscoie.com

ESOP firms with a breach			
State	**Adoption year**	**State**	**Adoption year**
AR	2005	AL	2018
AZ	2006	AK	2009
CA	2003	CO	2006
CT	2006	DE	2005
FL	2014	DC	2007
GA	2005	HI	2007
IA	2008	ID	2006
IL	2006	IA	2008
IN	2006	KS	2007
KY	2014	LA	2006
MA	2007	ME	2006
MD	2008	MS	2011
MI	2007	MT	2006
MN	2006	NE	2006
MO	2009	NH	2007
NC	2005	NM	2017
NJ	2006	ND	2005
NV	2005	OK	2008
NY	2005	OR	2007
OH	2006	PA	2006
RI	2023	PR	2006
SD	2018	SC	2009
TN	2005	VT	2012
TX	2009	WV	2008
UT	2007	WY	2007
VA	2019		
WA	2005		
WI	2006		

Table 2: Summary of the sample selection process
This table provides an overview of how data breach incidents corresponding to public ESOP firms were identified.

35,167 incidents		
Drop schools and universities		
Drop observations with a missing report date		
	Keep the Entity names	
	530 public firms	1,805 private firms
	126 ESOP firms	404 non ESOP
	578 reports	969 reports

Table 3: Distribution of 578 ESOP report incidents
Each ESOP firm in this sample experiences an average of 4.59 incidents.

Firms count	Mean	Std	min	25%	50%	75%	max
126	4.59	7.86	1	1	2	5	67

Table 4: Process of training and fine-tuning of the S-BERT model

1-Train an S-BERT model using the STS-B Multi-MT		
2-Create a sample of 2,000 pairs of data breach descriptions : 1,000 will include pairs from the same firm And 1,000 will include pairs from two different firms And Each description mentions the name of the firm involved.	Use the model trained in step 1 to find the similarity between the 2,000 pairs.	Divide the similarity measured for 1,000 pairs from two different firms by 10. Do not change the similarity measure for 1,000 pairs from the same firm.
		3-Use these 2,000 pairs along with an additional 3,000 pairs from the STS-B Multi-MT dataset to train another SBERT model. This is the final trained model.

Table 5: Summary Statistics of the **similarity measure** estimated by S_BERT model
Pairs Count represents the number of unique pairs of data breach descriptions within the same firm, where 'unique' means that pairs are considered the same regardless of order, implying that AB and BA are counted as a single pair.
A mean of 0.31 indicates that for an average firm in my dataset, which has more than one data breach reported, if I select two of these data breach descriptions and generate their associated vectors, then the cosine of the angle between those two vectors is 0.31.

Pairs Count	Mean	Std	min	25%	50%	75%	max
4,590	0.31	0.12	-0.10	0.22	0.30	0.39	0.76

Table 6: Summary statistics for the **number of days between two data breach incidents**
Pairs Count represents the number of unique pairs of data breach descriptions within the same firm, where 'unique' means that pairs are considered the same regardless of order, implying that AB and BA are counted as a single pair.
A mean of 961.5 indicates that, on average, the absolute time difference between two incident reports is 939 days.

Pairs Count	Mean	Std	min	25%	50%	75%	max
4,899	961.5	826	0	343.5	740	1,350	5,414

Table 7: Descriptions and dates of the distinct incidents identified by the machine learning method for Ameriprise Financial Inc.

Date of Breach	Description of Breach
2/12/2010	Ameriprise Financial Services Inc. experienced a physical data breach when an express mailing vendor lost a REIT application.This application, intended for the REIT transfer agent, contained the name, address, Social Security Number, and date of birth of one New Hampshire resident. The incident was reported to the New Hampshire Attorney General's Office on February 12, 2010, and the client was offered one year of credit monitoring from Equifax.
3/18/2015	A data breach occurred at Ameriprise Financial Services, Inc. on March 18, 2015. The breach was reported on April 24, 2015. The specific details of the breach, including how it happened and what information was impacted, are not provided. Only 2 records from the state were known to be impacted.
10/13/2015	A data breach at Ameriprise Financial Services, Inc. occurred on October 13, 2015, when an advisor's computer was accessed by a third party to fix an issue, allowing potential access to client files containing names, addresses, dates of birth, Social Security, an account numbers. The breach affected one resident of New Hampshire. The company provided credit monitoring services through Equifax and took steps to protect accounts from unauthorized activity.
6/15/2016	Ameriprise Financial Services, Inc. experienced a data breach where documents including a Maryland resident's personal information (name, SSN, DOB, driver's license number, account number) were lost in mail transit and have not been located. The breach was reported to the Maryland Office of the Attorney General and steps are taken to monitor credit and protect against unauthorized access.
10/20/2016	There is limited information available for this breach. It was reported on October 20, 2016. Ameriprise Financial Services, Inc. experienced a data breach affecting 477 records, with 8 records from the state of Indiana.
12/6/2016	On December 6, 2016, Ameriprise Financial Services, Inc. experienced a breach when an external hard drive in a franchise advisor's office was found accessible online. This incident affected three New Hampshire residents, exposing their names, addresses, dates of birth, account numbers, and Social Security numbers. The issue was resolved by taking the hard drive offline, and the affected individuals were offered credit monitoring services.
12/20/2016	Details about the breach are not specified, including what happened, the duration or extent. It was discovered on 10/17/16 and reported on 12/20/16. Only the regulatory requirement to report the breach to ME (Maine) has been disclosed.
2/27/2017	The breach involved Ameriprise Financial Services Inc. However, specific details regarding the breach, including the date and the nature of the exposed data, were not provided. Only two records were reported to be impacted by the source from the state of Indiana (IN).
6/7/2017	A breach occurred at Ameriprise Financial Services, Inc. on June 7, 2017. The specifics of the breach, including what information was affected and how the breach happened, are unknown. The breach was reported on July 14, 2017.
11/1/2017	The data breach incident for Ameriprise Financial Services Inc. occurred on 11/1/2017, and was reported on 1/17/2018. There is no description of the breach available and the specific details of the breach including the information types impacted are unknown. Only 56 records were impacted from the data source, with just one record from Indiana.
2/16/2018	An Ameriprise Financial Services ex-advisor uploaded a client list with personal and sensitive information to a personal email on February 16, 2018. The data included names, addresses, email addresses, dates of birth, account, and Social Security Numbers for one Maryland resident, who is also being offered credit monitoring services.
6/5/2018	There is limited information regarding the breach at Ameriprise Financial Services Inc., but it affected at least one individual based on the source provided and was reported on June 5, 2018.
8/7/2018	A data breach at Ameriprise Financial Services, Inc. was identified on August 7, 2018, but specific details regarding the nature of the breach, the type of information impacted, and the measures taken are not provided. The breach affected 2,984 records, which was reported on February 27, 2019.
9/25/2018	A breach notification from Ameriprise Financial Services, Inc. was disseminated, giving extensive guidance on how to prevent identity theft and what to do if one's personal information is compromised. The letter did not specify details of the breach, including the nature of the breach or what specific data was impacted. It was reported on September 25, 2018.
12/8/2018	There is no detailed description provided for the breach at Ameriprise Financial Services Inc. It is known that the breach occurred on December 8, 2018, and was reported on January 4, 2019. Only a single record was impacted according to the source.
7/13/2019	The nature of the breach, specifics of the incident, and the type of information impacted are unknown. A total of 41 records were impacted in this breach. It was reported on July 31st, 2019.

1/14/2020	On January 14, 2020, Ameriprise Financial Services, Inc. experienced a data breach when a third-party service provider sent out an email containing personal information of a Maryland resident to an incorrect recipient. The information included the client's name, Ameriprise ID, and Social Security Number. Ameriprise is providing credit monitoring services to the affected individual.
5/1/2020	A data breach was reported by Ameriprise Financial Inc. on October 23, 2020, which occurred on May 1, 2020. The details of the breach, including the nature and type of information compromised, are not provided. 696 records were impacted, but the specific states affected are not clear.
8/17/2020	A data breach occurred at Ameriprise Financial, Inc. on August 17, 2020. The exact details of the breach, including the manner in which the breach occurred and the specific types of information compromised, are not provided. It was reported on September 9, 2020. The breach affected 387 records but further details are unknown.
6/24/2021	A data breach occurred at Ameriprise Financial Inc on June 24, 2021. The specific details of the breach, such as how it happened and what information was impacted are unknown. The breach was reported on June 29, 2021.
3/28/2022	A data breach occurred at Ameriprise Financial, Inc. starting on March 28, 2022, and was reported on June 1, 2022. There is no available description of the breach, nor details about the specific information impacted. A total of 175 records were impacted.

Table 8: Distribution of 231 distinct incidents
Each ESOP firm in this sample experiences an average of 1.83 distinct incidents.

Firms count	Mean	Std	min	25%	50%	75%	max
126	1.83	2.51	1	1	1	2	21

7 Likelihood of experiencing a data breach incident

Table 9: Summary statistics.
The table shows summary statistics for a sample of 197 firm-year observations that experienced a data breach incident in the following fiscal year (83 distinct ESOP firms) and 4,255 firm-year observations from 551 distinct ESOP firms that never experience a data breach over the period 2005 to 2023. The appendix provides detailed descriptions of the construction of the variables. ***,** and * denote that t-tests (Welch's t-test) for mean differences in firm and industry characteristics between attacked and nonattacked firms are significant at the 1%, 5%, and 10% levels, respectively.

	Firm-years followed by incident (N =125): A		**Firms without incident (N = 4,255): B**		**Test of difference (A–B)**
Variable	**Mean**	**Median**	**Mean**	**Median**	**Mean Difference**
Active participants (000)	91.52	30.58	8.07	2.51	83.45 ***
Active participants/total employees	0.59	0.63	0.61	0.67	-0.02
ESOP assets per participant ($ thousand)	135.94	122.26	129.81	94.52	6.13
Total ESOP assets ($ billion)	6.59	3.44	1.42	0.23	5.17 ***
Wage ($ thousand)	90.30	75.91	77.59	70.62	12.70 **
Executive salary/Executive total pay	0.19	0.16	0.37	0.31	-0.18 ***
Executive stock /Executive total pay	0.38	0.39	0.24	0.23	0.14 ***
Executive option /Executive total pay	0.12	0.10	0.11	0.05	0.01
Executive bonus/Executive total pay	0.05	0.00	0.05	0.00	0.00
Executive non-equity/Executive total pay	0.22	0.21	0.17	0.18	0.04 ***
Total assets ($ billion)	116.39	40.52	19.67	3.94	96.72 ***
Asset intangibility	0.80	0.88	0.76	0.85	0.04 *
CAPX/assets	0.03	0.02	0.03	0.02	-0.00
R&D/assets	0.01	0.00	0.01	0.00	-0.01 ***
Leverage	1.09	0.76	0.91	0.68	0.18
Q	1.67	1.42	1.63	1.30	0.05
Tobin's q (previous year)	1.64	1.47	1.61	1.31	0.02
ROA	0.05	0.05	0.04	0.03	0.01 **
Financial constraint (indicator)	0.06	0.00	0.37	0.00	-0.32 ***
Stock retrun volatility	0.02	0.01	0.02	0.02	-0.01 ***
Stock performance	0.01	0.01	0.01	-0.02	0.01
Institutional block ownership	0.17	0.16	0.17	0.16	0.00
Industry's Tobin's q	1.50	1.40	1.49	1.32	0.01
Industry's Herfindahl index	0.10	0.04	0.05	0.03	0.04 ***
Risk committee (indicator)	0.02	0.00	0.01	0.00	0.01
Number of board committees	5.09	5.00	4.32	4.00	0.77 ***
Fortune 500 membership (indicator)	0.81	1.00	0.33	0.00	0.48 ***
sales growth	1.03	1.03	1.05	1.04	-0.02 *

Table 10: Correlation of important variables used in the analysis. The active ratio is defined as the ratio of plan active participants to the total number of employees in the ESOP firm. Star shows the significance level of 5% and less. The appendix provides detailed descriptions of the construction of the variables.

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**
(1) active ratio	1.00												
(2) log(ESOP/ participants)	-0.18*	1.00											
(3) Executives equity compensation	-0.27*	0.41*	1.00										
(4) log(Wage)	0.07*	0.28*	0.06*	1.00									
(5) log(Total assets)	-0.11*	0.42*	0.56*	0.18*	1.00								
(6) Q prev year	-0.23*	0.21*	0.26*	-0.12*	0.01	1.00							
(7) Stock performance	-0.02	0.06*	0.01	-0.02	-0.04*	0.02	1.00						
(8) Fiancial constraint	0.11*	-0.29*	-0.44*	-0.05*	-0.69*	-0.06*	0.00	1.00					
(9) Stock return volatility	0.08*	-0.32*	-0.24*	-0.01	-0.29*	-0.19*	0.08*	0.25*	1.00				
(10) Instituitonal block ownership	-0.03*	0.08*	0.11*	0.08*	0.03*	0.04*	-0.02	0.02	-0.04*	1.00			
(11)R&D/assets	-0.13*	0.10*	0.10*	0.07*	-0.13*	0.26*	0.03*	0.07*	0.07*	0.00	1.00		
(12) Risk committee	0.08*	0.00	-0.03	0.08*	0.06*	-0.08*	-0.01	-0.03*	0.02	-0.02	-0.05*	1.00	
(13) Number of board committees	-0.02	0.19*	0.25*	0.09*	0.48*	-0.02	-0.03*	-0.33*	-0.14*	0.03	-0.04*	0.05*	1.00

Table 11: Distribution of data breach incidents by year and industry.
The sample consists 123 firm-year observations that experienced at least one cyberattack in a fiscal year over the period 2010 to 2023. Numbers in parentheses show the range of the 2-digit SIC code for each category.

Fiscal Year	**Manufacturing (20-39)**	**Transport, communications (40-48)**	**Electric, gas, and sanitary services 49**	**Wholesale trade and retail trade (50-59)**	**Finance (60-69)**	**Service industries (70-89)**	**Nonclassifiable Establishments 99**	**Total**
2011	0	0	0	0	2	0	0	2
2012	1	0	0	0	0	1	0	2
2013	1	1	0	*1*	2	2	0	7
2014	2	2	0	3	4	2	0	13
2015	3	2	1	3	5	2	0	16
2016	5	0	1	3	8	2	1	20
2017	3	1	0	2	8	1	0	15
2018	2	0	0	3	9	1	1	16
2019	1	2	1	0	5	1	0	10
2020	2	0	1	2	4	1	0	10
2021	3	0	0	0	2	0	0	5
2022	5	0	0	0	2	0	0	7
Total	28	8	4	17	51	13	2	123

Table 12: Distribution of data breach incidents by year and state.
The sample consists of 123 ESOP firm-year observations that experienced at least one cyberattack in a fiscal year over the period from 2010 to 2023. The adoption year indicates the year when a security breach law was first implemented in the state.

	adoption year	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**Total**
AR	2005	0	0	0	0	0	1	0	1	0	1	0	0	3
AZ	2006	0	0	0	0	0	0	0	0	0	1	0	0	1
CA	2003	0	0	0	0	1	2	1	1	0	0	0	0	5
CT	2006	0	0	0	0	0	2	1	0	0	0	0	0	3
FL	2014	0	0	0	0	1	0	1	0	1	1	0	0	4
GA	2005	1	0	1	1	3	1	1	3	0	0	0	1	12
IA	2008	0	0	0	0	0	0	1	1	1	0	0	0	3
IL	2006	0	0	1	1	0	4	3	0	0	1	0	0	10
IN	2006	0	0	0	0	0	0	0	0	0	0	0	1	1
KY	2014	0	0	0	1	0	1	0	1	1	1	1	1	7
MA	2007	0	0	0	0	0	1	0	0	0	0	0	0	1
MD	2008	0	0	0	1	0	0	0	1	1	0	0	0	3
MI	2007	0	0	0	0	0	0	0	1	0	0	0	0	1
MN	2006	0	0	0	1	1	2	2	1	1	1	0	1	10
MO	2009	0	0	0	0	0	0	0	0	0	1	0	0	1
NC	2005	0	0	1	0	2	2	0	1	1	0	1	1	9
NJ	2006	0	1	1	0	1	0	0	0	0	1	0	0	4
NV	2005	0	0	0	0	0	1	0	0	0	0	0	0	1
NY	2005	1	0	1	3	2	1	2	2	1	1	2	0	16
OH	2006	0	1	0	2	1	1	0	1	1	0	1	1	9
RI	2023	0	0	1	0	1	0	1	0	0	1	0	0	4
SD	2018	0	0	0	0	0	0	0	0	1	0	0	0	1
TN	2005	0	0	0	1	0	0	0	0	0	0	0	0	1
TX	2009	0	0	1	1	1	0	1	1	0	0	0	0	5
UT	2007	0	0	0	0	0	0	0	1	0	0	0	0	1
VA	2019	0	0	0	0	0	1	0	0	1	0	0	1	3
WA	2005	0	0	0	1	0	0	1	0	0	0	0	0	2
WI	2006	0	0	0	0	2	0	0	0	0	0	0	0	2
Total		2	2	7	13	16	20	15	16	10	10	5	7	123

Table 13: Role of ESOP holdings in the likelihood of a data breach incident.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach in a given year, and zero otherwise. The sample consists of 125 firm-year observations that experienced a data breach in the following fiscal year (83 distinct ESOP firms) and the remaining firm-year observations that did not experience an incident during the period from 2005 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

	Dependent variable = Data breach incident (indicator)						
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Active ratio	-0.14***		-0.10*	-0.10*	-0.78*	-0.34***	-0.31***
	(0.05)		(0.05)	(0.05)	(0.42)	(0.10)	(0.09)
log(ESOP/active participants)			-0.12**	-0.13**	-0.11**	-0.12**	-0.12**
			(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Active ratio ×					0.06		
log(ESOP/active participants)					(0.04)		
Executive equity/Executive total pay		0.23		0.28	0.27	0.26	0.22
		(0.28)		(0.28)	(0.28)	(0.28)	(0.29)
Active ratio ×						0.51**	
Executive equity/Executive total pay						(0.20)	
Active ratio×							0.07**
log(Total assets)							(0.03)
log(wage)		0.16	0.17	0.17	0.18	0.15	0.12
		(0.25)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)
log(Total assets)		0.11**	0.14***	0.12**	0.12**	0.13**	0.12**
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Q prev year		-0.22**	-0.19**	-0.20**	-0.20*	-0.19*	-0.21**
		(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
ROA		2.62*	2.80*	2.76*	2.68*	2.54*	2.73*
		(1.50)	(1.53)	(1.53)	(1.55)	(1.54)	(1.56)
sales growth		-0.27	-0.35	-0.36	-0.37	-0.38	-0.37
		(0.32)	(0.32)	(0.32)	(0.32)	(0.32)	(0.32)
Stock performance		-0.05	-0.03	-0.01	-0.03	-0.02	-0.04
		(0.18)	(0.17)	(0.18)	(0.18)	(0.17)	(0.18)
Leverage		-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Financially constraint (indicator)		-0.09	-0.09	-0.09	-0.09	-0.10	-0.10
		(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
Stock return volatility		-2.02	-3.29	-3.46	-2.94	-4.31	-4.65
		(5.96)	(6.27)	(6.27)	(6.26)	(6.33)	(6.37)
Institutional block ownership		-0.73	-0.64	-0.71	-0.76	-0.79	-0.81
		(0.58)	(0.59)	(0.59)	(0.59)	(0.60)	(0.61)
Zero institutional ownership indicator		-0.18	-0.17	-0.17	-0.18	-0.19	-0.22
		(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
CAPX/assets		3.19	3.27	3.14	2.97	2.82	2.98
		(3.02)	(3.11)	(3.10)	(3.09)	(3.12)	(3.15)
Asset intangibility		0.08	0.12	0.15	0.10	0.13	0.13
		(0.55)	(0.55)	(0.56)	(0.57)	(0.57)	(0.58)
Fortune 500 (indicator)		0.37**	0.39**	0.38**	0.38**	0.35**	0.39**
		(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.17)
Risk committee (indicator)		-0.10	-0.18	-0.18	-0.15	-0.13	-0.13
		(0.20)	(0.20)	(0.21)	(0.21)	(0.22)	(0.21)
Number of board committees		0.01	0.01	0.01	0.01	0.01	0.01
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
log(age)		0.09	0.10	0.11	0.11	0.12	0.10
		(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
R&D/assets		-1.92	-1.54	-1.49	-1.43	-1.24	-0.92
		(3.83)	(3.82)	(3.81)	(3.72)	(3.59)	(3.48)
Zero R&D indicator		-0.18	-0.18	-0.17	-0.17	-0.17	-0.17
		(0.27)	(0.27)	(0.27)	(0.27)	(0.26)	(0.27)

Tax_high		-0.04	-0.06	-0.05	-0.05	-0.06	-0.07
		(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
Constant	-2.72***	-5.11*	-3.95	-3.99	-4.16	-3.79	-3.43
	(0.44)	(2.69)	(2.69)	(2.72)	(2.74)	(2.76)	(2.76)
Observations	4,289	3,963	3,984	3,963	3,963	3,963	3,963
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.176	0.241	0.245	0.245	0.247	0.248	0.249

Table 14: Role of ESOP holdings and local labor market in the likelihood of a data breach incident.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach in a given year, and zero otherwise. The sample consists of 125 firm-year observations that experienced a data breach in the following fiscal year (83 distinct ESOP firms) and the remaining firm-year observations that did not experience an incident during the period from 2005 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

	Dependent variable = Data breach incident (indicator)					
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Active ratio	-0.34***	-0.21***	-0.31**	-0.06	-0.40**	-0.40**
	(0.10)	(0.07)	(0.16)	(0.09)	(0.16)	(0.16)
log(ESOP/active participants)	-0.12**	-0.23***	-0.23***	-0.03	-0.03	-0.03
	(0.06)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Executive equity/Executive total pay	0.26	0.88**	0.87**	0.83*	0.88**	0.88**
	(0.28)	(0.36)	(0.37)	(0.43)	(0.42)	(0.42)
Active ratio ×	0.51**		0.21		0.71**	0.71**
Executive equity/Executive total pay	(0.20)		(0.30)		(0.30)	(0.30)
log(wage)	0.15	0.67	0.65	0.55	0.51	0.51
	(0.24)	(0.43)	(0.44)	(0.36)	(0.38)	(0.38)
Risk committee (indicator)	-0.13	-0.79**	-0.76**	0.03	0.15	0.15
	(0.22)	(0.31)	(0.30)	(0.40)	(0.42)	(0.42)
Number of board committees	0.01	-0.01	-0.01	0.00	0.00	0.00
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Constant	-3.79	-6.33	-6.16	-6.86	-6.19	-4.88
	(2.76)	(4.76)	(4.84)	(4.25)	(4.40)	(4.44)
Observations	3,963	1,242	1,242	1,107	1,107	1,107
Independent variables	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.248	0.308	0.308	0.302	0.306	0.306
Industry×Year FE	N	Y	Y	N	N	N
State×Year FE	N	N	N	Y	Y	Y
State FE	N	N	N	N	N	Y

Table 15: Role of executive equity compensation, firm size, and the active ratio in the likelihood of a data breach incident.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach in a given year, and zero otherwise. The sample consists of 125 firm-year observations that experienced a data breach in the following fiscal year (83 distinct ESOP firms) and the remaining firm-year observations that did not experience an incident during the period from 2005 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

	Dependent variable = Data breach incident (indicator)						
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Active ratio	-0.34***	-0.39***	-0.36**	-0.39***	-0.10**	-0.22***	-0.39**
	(0.10)	(0.10)	(0.17)	(0.14)	(0.05)	(0.07)	(0.19)
log(Total assets)	0.13**	0.12**	0.16**	0.17**	0.27***	0.29**	0.28**
	(0.05)	(0.05)	(0.07)	(0.07)	(0.09)	(0.12)	(0.12)
Executive equity/Execuitve total pay	0.26	0.23	0.86**	0.86**	1.27**	1.77**	1.72**
	(0.28)	(0.28)	(0.37)	(0.37)	(0.50)	(0.69)	(0.70)
Active ratio ×		0.05	0.06	0.06			0.06
log(Total assets)		(0.03)	(0.05)	(0.04)			(0.04)
Active ratio ×	0.51**	0.27	-0.10				-0.02
Executive equity/Execuitve total pay	(0.20)	(0.26)	(0.38)				(0.38)
Executive equity/Execuitve total pay×					-0.31**	-0.28	-0.27
log(Total assets)					(0.13)	(0.19)	(0.19)
log(ESOP/active participants)	-0.12**	-0.12**	-0.22***	-0.22***	-0.13**	-0.25***	-0.24***
	(0.06)	(0.06)	(0.08)	(0.08)	(0.06)	(0.08)	(0.08)
Constant	-3.79	-3.44	-5.55	-5.55	-4.18	-6.79	-6.04
	(2.76)	(2.77)	(4.95)	(4.96)	(2.73)	(4.90)	(5.10)
Observations	3,963	3,963	1,242	1,242	3,963	1,242	1,242
Independent variables	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.248	0.250	0.310	0.310	0.248	0.310	0.312
Industry×Year FE	N	N	Y	Y	N	Y	Y

Table 16: Quantile regression
The table presents estimates of quantile probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach in a given year, and zero otherwise. The sample consists of 125 firm-year observations that experienced a data breach in the following fiscal year (83 distinct ESOP firms) and the remaining firm-year observations that did not experience an incident during the period from 2005 to 2023. In Column 1, data is sorted based on log(ESOP/employees). In Column 2, sorting is based on active ratio. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table

Dependent variable = Data breach incident (indicator)		
	Sorted on (ESOP/emp)	**Sorted on active ratio**
VARIABLES	**(1)**	**(2)**
Active ratio	-0.41**	
	(0.17)	
log(ESOP/emp)		0.00
		(0.08)
Q2-Q1	0.19	-0.21
	(0.20)	(0.43)
Q3-Q1	0.06	-1.13***
	(0.25)	(0.41)
Executive equity/Executive total pay	0.86**	-0.03
	(0.43)	(0.62)
Q2-Q1×		0.68
Executive equity/Executive total pay		(0.76)
Q3-Q1×		2.15***
Executive equity/Executive total pay		(0.74)
Active ratio ×	0.74**	
Executive equity/Executive total pay	(0.32)	
Constant	-5.10	-4.10
	(4.36)	(4.40)
Observations	1,107	1,105
Independent variables	Y	Y
Industry FE	Y	Y
Year FE	Y	Y
State×Year FE	Y	Y
Pseudo R-squared	0.307	0.309

7.1 Effects of a data breach on ESOP participation

Table 17: Summary statistics of treatment and control observations. The table shows summary statistics for a sample of 122 firm-year observations that experienced a data breach in a fiscal year (62 distinct ESOP firms) and 62 firm-year observations (54 distinct ESOP firms) that did not experience a cyberattack over the period 2010 to 2023. The sample is created using propensity score matching. The appendix provides detailed descriptions of the construction of the variables. , and denote that t-tests (Welch's t-test) for mean differences in firm and industry characteristics between attacked and nonattacked firms are significant at the 1%, 5%, and 10% levels, respectively.

	Firm-years with data breach (N=69):		**Firm-years without data breach (N=69):**		**Test of difference:**
	A		**B**		**A-B**
Variable	**(mean)**	**(median)**	**(mean)**	**(median)**	
Total assets ($ billion)	41.13	14.34	36.55	15.53	4.57
Leverage	0.99	0.72	1.06	0.76	-0.06
Q	1.81	1.65	1.75	1.46	0.05
ROA	0.06	0.05	0.05	0.04	0.00
Financial constraint (indicator)	0.14	0.00	0.13	0.00	0.01
Stock retrun volatility	0.02	0.02	0.02	0.02	0.00
Stock performance	0.02	0.04	0.04	0.03	-0.02
Institutional block ownership	0.17	0.17	0.18	0.17	-0.00
Industry's Tobin's q	1.58	1.50	1.54	1.32	0.04
Asset.intangibility	0.77	0.85	0.77	0.88	0.00
CAPX/assets	0.03	0.03	0.03	0.02	0.00
R&D/assets	0.01	0.00	0.02	0.00	-0.01
Fortune 500 membership (indicator)	0.62	1.00	0.49	0.00	0.13
Industry's Herfindahl index	0.10	0.04	0.05	0.03	0.05 ***
sales growth	1.04	1.04	1.04	1.03	0.01

Table 18: Effect of data breach incidents on the active participants ratio
This table presents results of Equation 2. The sample results from a propensity score matching process without replacement. The propensity score is calculated using the logit regression of data breach incident (an indicator that takes the value one if a firm experiences a data breach involving financial information loss, and zero otherwise) on firm size, stock performance, stock return volatility, and leverage. I require both treated and matching firms to be in the same same fiscal year. The sample consists of 1,170 firm-year observations (61 treated firms that experience a cyberattack and 53 control firms that do not experience a cyberattack). The appendix provides detailed descriptions of the construction of the variables. **Active ratio** equals the number of active ESOP participants relative to firm employee population. **T** is an indicator for a treatment group. **breach year and the following year** is an indicator variable with value one for the year breach has happened in it or the year after that. It is zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	Active ratio					
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
T*	0.03*	0.03*	0.02	0.02	0.03	0.03
(breach year and the following year)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
T×Breach year+2		0.00	-0.01	0.01	-0.00	-0.00
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
T×Breach year+3		0.01	0.00	0.01	-0.00	-0.00
		(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
T×Breach year+4		-0.02	-0.03	-0.02	-0.02	-0.02
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
T×Breach year+5		-0.01	-0.02	-0.01	-0.02	-0.02
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
T×Breach year+6-11			-0.04*			
			(0.02)			
Constant	0.62***	0.62***	0.63***	0.64***	0.46**	0.30
	(0.13)	(0.13)	(0.13)	(0.17)	(0.18)	(0.24)
Observations	1,616	1,616	1,616	1,616	1,616	1,616
R-squared	0.89	0.89	0.89	0.93	0.94	0.94
Control variables	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	N	Y
Firm FE	Y	Y	Y	Y	Y	Y
industry*year FE	N	N	N	N	Y	Y
State*Year FE	N	N	N	Y	N	N

Table 19: Effect of cyberattacks on firms' labor force
This table presents results of Equation 2. The sample results from a propensity score matching process without replacement. The propensity score is calculated using the logit regression of data breach incident (an indicator that takes the value one if a firm experiences a data breach involving financial information loss, and zero otherwise) on firm size, stock performance, stock return volatility, and leverage. I require both treated and matching firms to be in the same same fiscal year. The sample consists of 1,170 firm-year observations (61 treated firms that experience a data breach and 53 control firms that do not experience a data breach). The appendix provides detailed descriptions of the construction of the variables. In colum (1), **log(ESOP/participants)** equals logarithm of ESOP assets per participant. In column (2), **log(emp)** equals logarithm of the firm employees population. In column (3), **log(active employees)** equals the logarithm of the ESOP active participants. In column (4), **log(wage)** equals the logarithm of wage per employee. **T** is an indicator for a treatment group. **breach year and the following year** is an indicator variable with value one for the year breach has happened in it or the year after that. It is zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	**log(ESOP/participant)**	**log(employees)**	**log(ESOP employees)**	**log(wage)**
T×	-0.13**	-0.00	0.06	0.00
(breach year and the following year)	(0.06)	(0.03)	(0.05)	(0.02)
T×Breach year+2	-0.07	-0.03	-0.08	0.01
	(0.08)	(0.03)	(0.08)	(0.02)
T×Breach year+3	-0.05	-0.03	-0.06	0.00
	(0.06)	(0.03)	(0.06)	(0.02)
T×Breach year+4	-0.04	0.00	-0.08	0.00
	(0.06)	(0.03)	(0.06)	(0.02)
T×Breach year+5	-0.04	-0.01	-0.06	-0.00
	(0.05)	(0.03)	(0.05)	(0.01)
Constant	13.61***	1.72***	0.82*	11.62***
	(0.41)	(0.30)	(0.46)	(0.14)
Observations	1,616	1,616	1,616	1,593
R-squared	0.96	1.00	0.99	0.99
Control variables	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Industry*Year FE	Y	Y	Y	Y

Table 20: Effect of data breach on executive compensation.
This table presents results of Equation 2. The sample results from a propensity score matching process without replacement. The propensity score is calculated using the logit regression of data breach (an indicator that takes the value one if a firm experiences a data breach involving financial information loss, and zero otherwise) on firm size, stock performance, stock return volatility, and leverage. I require both treated and matching firms to be in the same same fiscal year. The sample consists of 1,170 firm-year observations (61 treated firms that experience a data breach and 53 control firms that do not experience a data breach). **Executive equity** equals the mean of the equity portion (stock + options) relative to total compensation across all executives. **Executive salary** equals the mean of salary relative to total compensation across all executives. **Executive bonus** equals the mean of bonus relative to total compensation across all executives. **Executive non-equity incentive** equals the mean of non-equity incentives relative to total compensation across all executives. **CEO equity** equals the equity portion (stock + options) relative to total compensation for CEO.**T** is an indicator for a treatment group. **breach year and the following year** is an indicator variable with value one for the year breach has happened in it or the year after that. It is zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	**Executive equity**	**Executive equity**	**Executive salary**	**Executive bonus**	**Executive non-equity incentive**
Dit	0.01	0.00	-0.01	0.01	0.00
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
D2	0.02	0.01	-0.01	0.02	-0.00
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
D3	0.01	-0.01	0.01	0.02*	-0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
D4	-0.00	-0.01	-0.01	0.02	0.00
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
D5	0.00	0.02	-0.01	0.01	-0.00
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Constant	0.62***	0.61***	0.38***	0.22***	0.08
	(0.08)	(0.13)	(0.09)	(0.06)	(0.10)
Observations	1,614	1,614	1,614	1,614	1,532
R-squared	0.75	0.88	0.90	0.84	0.81
Control variables	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
Industry*Year FE	N	Y	Y	Y	Y
Industry FE	N	N	N	Y	Y

7.2 Data breach incidents as industry-wide shocks

Table 21: Composition of the treatment groups and the control group in the industry shock sample. The table shows the number of observations and the number of unique 4-digit SIC codes in each group.

	Observations	***4-digit Industries***
Control group	*962*	*70*
Treatment groups		
2006	27	1
2007	206	6
2008	889	2
2010	72	6
2011	78	7
2012	114	5
2013	91	5
2014	45	5
2015	136	6
2016	164	12
2017	136	11
2018	81	4
2019	185	5
Total	3,186	

Table 22: Effects of the first noticeable data breach incident in each 4-digit SIC industry on other firms within the same industry. This table presents results of Equation 3. The sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. In Panel A, **Active ratio** equals the number of active ESOP participants relative to the firm employee population. In Panel B, column (1) **log(ESOP/participants)** equals the logarithm of ESOP assets per participant. In column (2), **log(emp)** equals the logarithm of the firm's employee population. In column (3) **log(active employees)** equals the logarithm of the ESOP active participants. In column (4), **log(wage)** equals the logarithm of wage per employee. In column (5), **log(ESOP)** equals the logarithm of ESOP asset value. **T** is an indicator for a treatment group.**(2-5 years after the first breach in the industry)** is an indicator variable with a value of one for the period two to five years after the breach in the industry, and zero otherwise. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. **Breach year+11-15** means an indicator with a value of one for the period eleven to fifteen years after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	Panel A			
	Active ratio			
VARIABLES	*(1)*	*(2)*	*(3)*	*(4)*
T ×	0.03*	0.03**	0.03	0.03**
(2-5 years after the first breach in the industry)	(0.02)	(0.01)	(0.03)	(0.01)
T×Breach year	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
T×Breach year+1	0.02	0.01	0.00	0.01
	(0.02)	(0.02)	(0.04)	(0.02)
T×Breach year+6		0.01	0.00	
		(0.01)	(0.03)	
T×Breach year+7		0.00	-0.00	
		(0.02)	(0.02)	
T×Breach year+8		-0.00	-0.01	
		(0.02)	(0.02)	
T×Breach year+9		-0.02	-0.02	
		(0.02)	(0.02)	
T×Breach year+10		-0.01	-0.01	
		(0.02)	(0.02)	
T×Breach year+11-15			0.01	
			(0.04)	
Constant	0.48***	0.47***	0.48***	0.84***
	(0.09)	(0.07)	(0.07)	(0.11)
Observations	2,547	3,141	3,141	3,141
R-squared	0.94	0.93	0.93	0.93
Control variables	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y

	Panel B				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
	log(ESOP/ participant)	***log(emp)***	***log(active employees)***	***log(wage)***	***log(ESOP)***
T×Breach year	0.01	-0.02	0.01	0.01	0.04
	(0.06)	(0.04)	(0.03)	(0.01)	(0.07)
T×Breach year+1	-0.13**	-0.03	0.03	0.00	-0.07
	(0.06)	(0.04)	(0.04)	(0.02)	(0.06)
T×Breach year+2	-0.09*	-0.04	0.07*	0.01	-0.03
	(0.05)	(0.03)	(0.04)	(0.02)	(0.05)
T×Breach year+3	-0.11*	0.01	0.10***	-0.05*	-0.01
	(0.06)	(0.04)	(0.04)	(0.03)	(0.06)
T×Breach year+4	-0.10*	-0.02	0.07*	-0.03*	-0.02
	(0.05)	(0.03)	(0.04)	(0.02)	(0.05)
T×Breach year+5	-0.10**	-0.02	0.08**	-0.02	-0.02
	(0.05)	(0.03)	(0.04)	(0.02)	(0.04)
Constant	10.73***	3.45***	0.44	10.88***	17.68***

	(0.42)	(0.10)	(0.36)	(0.19)	(0.48)
Observations	3,141	3,182	3,141	3,119	3,139
R-squared	0.95	0.99	0.99	0.94	0.98
Control variables	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y

Table 23: Effects of the first noticeable data breach incident in each 4-digit SIC industry on peer firms' executives.
This table presents results of Equation 3. The sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. **Executive equity** equals the mean of the equity portion (stock + options) relative to total compensation across all executives. **Executive salary** equals the mean of salary relative to total compensation across all executives. **Executive bonus** equals the mean of bonus relative to total compensation across all executives. **Executive non-equity incentive** equals the mean of non-equity incentives relative to total compensation across all executives. **T** is an indicator for a treatment group. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number refers to an indicator with a value of one for the specific time after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. All control variables are defined in the Appendix. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	*(2)*	*(3)*	*(4)*
VARIABLES	***Executive equity***	***Executive salary***	***Executive bonus***	***Executive non-equity incentive***
T×Breach year	-0.01	0.01	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
T×Breach year+1	-0.01	0.03**	-0.00	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)
T×Breach year+2	-0.00	0.03**	0.00	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
T×Breach year+3	0.01	0.01	-0.00	-0.02*
	(0.02)	(0.01)	(0.01)	(0.01)
T×Breach year+4	-0.01	0.02	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
T×Breach year+5	0.00	0.02	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.77***	0.26***	0.13***	0.04
	(0.05)	(0.05)	(0.02)	(0.04)
Observations	3,117	3,117	3,117	2,950
R-squared	0.80	0.83	0.62	0.65
Control variables	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
State*Year FE	N	N	N	N

Table 24: Effects of the first noticeable data breach incident in each 4-digit SIC industry on other firms within the same industry. **Firms experiencing the shock in 2007 or 2008 are excluded from the sample.**
This table presents results of Equation 3. The primary sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. **Active ratio** equals the number of active ESOP participants relative to the firm employee population. Column (2) **log(ESOP/participants)** equals the logarithm of ESOP assets per participant. In column (3), **log(emp)** equals the logarithm of the firm's employee population. In column (4) **log(active employees)** equals the logarithm of the ESOP active participants. In column (5), **log(wage)** equals the logarithm of wage per employee. In column (6), **log(ESOP)** equals the logarithm of ESOP asset value. **T** is an indicator for a treatment group.**(2-5 years after the first breach in the industry)** is an indicator variable with a value of one for the period two to five years after the breach in the industry, and zero otherwise. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. **Breach year+11-15** means an indicator with a value of one for the period eleven to fifteen years after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Active ratio**	**log(ESOP/ participant)**	**log(employees)**	**log(ESOP employees)**	**log(wage)**	**log(ESOP)**
T×	0.04**					
(2-5 years after the first breach in the industry)	(0.02)					
T×Breach year	-0.00	-0.06	-0.02	0.04	0.00	-0.01
	(0.02)	(0.08)	(0.06)	(0.05)	(0.02)	(0.07)
T×Breach year+1	0.01	-0.08	-0.05	0.04	0.00	0.02
	(0.02)	(0.07)	(0.07)	(0.07)	(0.03)	(0.06)
T×Breach year+2		-0.03	-0.06	0.07	0.02	0.05
		(0.08)	(0.06)	(0.05)	(0.04)	(0.07)
T×Breach year+3		-0.04	0.03	0.12**	0.01	0.09
		(0.08)	(0.06)	(0.06)	(0.03)	(0.08)
T×Breach year+4		0.01	0.03	0.10*	-0.01	0.11
		(0.08)	(0.06)	(0.05)	(0.03)	(0.08)
T×Breach year+5		-0.03	0.01	0.11	0.03	0.08
		(0.06)	(0.06)	(0.08)	(0.03)	(0.10)
Constant	0.73***	6.66***	3.49***	2.40***	10.58***	14.77***
	(0.06)	(0.37)	(0.07)	(0.17)	(0.17)	(0.38)
Observations	2,052	2,052	2,089	2,052	2,031	2,050
R-squared	0.94	0.96	0.99	0.99	0.94	0.98
Control variables	Y	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y

Table 25: Effects of the first noticeable data breach incident in each 4-digit SIC industry on other firms within the same industry. **Financial firms are excluded from the sample.**
This table presents results of Equation 3. The primary sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. **Active ratio** equals the number of active ESOP participants relative to the firm employee population. Column (2) **log(ESOP/participants)** equals the logarithm of ESOP assets per participant. In column (3), **log(emp)** equals the logarithm of the firm's employee population. In column (4) **log(active employees)** equals the logarithm of the ESOP active participants. In column (5), **log(wage)** equals the logarithm of wage per employee. In column (6), **log(ESOP)** equals the logarithm of ESOP asset value. **T** is an indicator for a treatment group.**(2-5 years after the first breach in the industry)** is an indicator variable with a value of one for the period two to five years after the breach in the industry, and zero otherwise. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. **Breach year+11-15** means an indicator with a value of one for the period eleven to fifteen years after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Active ratio**	**log(ESOP/ participant)**	**log(employees)**	**log(ESOP employees)**	**log(wage)**	**log(ESOP)**
T×	0.03*					
(2-5 years after the first breach in the industry)	(0.01)					
T×Breach year	-0.01	-0.04	0.04	0.05	0.00	0.01
	(0.02)	(0.07)	(0.04)	(0.05)	(0.02)	(0.07)
T×Breach year+1	0.01	-0.02	0.01	0.10*	0.02	0.08
	(0.02)	(0.07)	(0.05)	(0.05)	(0.02)	(0.06)
T×Breach year+2		-0.00	-0.02	0.09*	0.02	0.10
		(0.07)	(0.05)	(0.05)	(0.02)	(0.06)
T×Breach year+3		-0.01	0.07	0.12**	0.02	0.11
		(0.07)	(0.05)	(0.05)	(0.02)	(0.07)
T×Breach year+4		0.03	0.04	0.09**	-0.01	0.12*
		(0.07)	(0.05)	(0.04)	(0.03)	(0.07)
T×Breach year+5		-0.01	0.02	0.08	0.00	0.07
		(0.06)	(0.04)	(0.06)	(0.03)	(0.08)
Constant	0.75***	6.50***	3.49***	2.49***	10.57***	14.63***
	(0.06)	(0.31)	(0.07)	(0.15)	(0.17)	(0.33)
Observations	2,049	2,049	2,084	2,049	2,027	2,047
R-squared	0.94	0.97	0.99	0.99	0.96	0.98
Control variables	Y	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y

Table 26: Effects of the first noticeable data breach incident in each 4-digit SIC industry on other firms within the same industry. **Financial firms and treatment groups of 2007 and 2008 are excluded from the sample.**
This table presents results of Equation 3. The primary sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. **Active ratio** equals the number of active ESOP participants relative to the firm employee population. Column (2) **log(ESOP/participants)** equals the logarithm of ESOP assets per participant. In column (3), **log(emp)** equals the logarithm of the firm's employee population. In column (4) **log(active employees)** equals the logarithm of the ESOP active participants. In column (5), **log(wage)** equals the logarithm of wage per employee. In column (6), **log(ESOP)** equals the logarithm of ESOP asset value. **T** is an indicator for a treatment group.**(2-5 years after the first breach in the industry)** is an indicator variable with a value of one for the period two to five years after the breach in the industry, and zero otherwise. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. **Breach year+11-15** means an indicator with a value of one for the period eleven to fifteen years after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Active ratio**	**log(ESOP/ participant)**	**log(employees)**	**log(ESOP employees)**	**log(wage)**	**log(ESOP)**
T×	0.03*					
(2-5 years after the first breach in the industry)	(0.02)					
T×Breach year	-0.01	-0.07	0.03	0.04	0.01	-0.02
	(0.02)	(0.08)	(0.04)	(0.05)	(0.02)	(0.08)
T×Breach year+1	0.01	-0.07	0.00	0.09	0.01	0.03
	(0.02)	(0.07)	(0.05)	(0.06)	(0.03)	(0.06)
T×Breach year+2		-0.03	-0.04	0.09	0.02	0.07
		(0.09)	(0.06)	(0.06)	(0.03)	(0.07)
T×Breach year+3		-0.05	0.05	0.14**	0.03	0.10
		(0.08)	(0.06)	(0.06)	(0.03)	(0.08)
T×Breach year+4		0.03	0.03	0.11*	0.01	0.14
		(0.09)	(0.06)	(0.06)	(0.03)	(0.08)
T×Breach year+5		-0.02	-0.00	0.13	0.04	0.11
		(0.07)	(0.05)	(0.08)	(0.03)	(0.11)
Constant	0.76***	6.42***	3.49***	2.46***	10.56***	14.51***
	(0.06)	(0.33)	(0.07)	(0.17)	(0.17)	(0.34)
Observations	1,873	1,873	1,908	1,873	1,852	1,871
R-squared	0.94	0.96	0.99	0.98	0.94	0.98
Control variables	Y	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y

Table 27: Effects of the first noticeable data breach incident in each 4-digit SIC industry on other firms within the same industry. **The sample of treatments includes only financial firms that experienced a treatment in 2007 or 2008.**
This table presents results of Equation 3. The primary sample consists of 3,186 firm-year observations. Of these, 962 observations come from industries with no noticeable data breach incident, covering 70 4-digit SIC codes. The remaining observations, constituting the treatment group, come from 75 4-digit SIC codes. **Active ratio** equals the number of active ESOP participants relative to the firm employee population. Column (2) **log(ESOP/participants)** equals the logarithm of ESOP assets per participant. In column (3), **log(emp)** equals the logarithm of the firm's employee population. In column (4) **log(active employees)** equals the logarithm of the ESOP active participants. In column (5), **log(wage)** equals the logarithm of wage per employee. In column (6), **log(ESOP)** equals the logarithm of ESOP asset value. **T** is an indicator for a treatment group.**(2-5 years after the first breach in the industry)** is an indicator variable with a value of one for the period two to five years after the breach in the industry, and zero otherwise. **Breach year** is an indicator with a value of one for the breach year in the industry and zero otherwise. Any other variable in the form of breach year plus a number means an indicator with a value of one for the specific time after the breach and zero otherwise. **Breach year+11-15** means an indicator with a value of one for the period eleven to fifteen years after the breach and zero otherwise. Control variables include: leverage, Q, log(Total assets), stock return volatility, ROA, and institutional block ownership. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level., and denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Active ratio**	**log(ESOP/ participant)**	**log(employees)**	**log(ESOP employees)**	**log(wage)**	**log(ESOP)**
T×	0.05*					
(2-5 years after the first breach in the industry)	(0.03)					
T×Breach year	0.03	0.06	-0.07	0.02	0.03	0.18
	(0.03)	(0.14)	(0.07)	(0.07)	(0.03)	(0.24)
T×Breach year+1	0.06	-0.42***	-0.06	0.10	0.01	-0.27
	(0.04)	(0.16)	(0.07)	(0.09)	(0.04)	(0.18)
T×Breach year+2		-0.40***	-0.07	0.13	0.02	-0.29**
		(0.11)	(0.06)	(0.09)	(0.05)	(0.14)
T×Breach year+3		-0.44***	-0.04	0.16*	-0.13**	-0.32**
		(0.11)	(0.06)	(0.09)	(0.06)	(0.13)
T×Breach year+4		-0.32***	-0.07	0.11	-0.08**	-0.17*
		(0.10)	(0.05)	(0.08)	(0.04)	(0.10)
T×Breach year+5		-0.23***	-0.03	0.11	-0.06*	-0.09
		(0.09)	(0.05)	(0.08)	(0.03)	(0.09)
Constant	0.77***	10.90***	3.47***	0.58	10.71***	17.70***
	(0.14)	(0.56)	(0.15)	(0.43)	(0.21)	(0.68)
Observations	1,873	1,873	1,879	1,873	1,858	1,873
R-squared	0.94	0.96	0.99	0.98	0.94	0.98
Control variables	Y	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
State*Year FE	Y	Y	Y	Y	Y	Y

7.3 Robustness check

Table 28: Robustness check- Sample: all reports of data breach.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach incident in a given year, and zero otherwise. The sample consists of 197 firm-year observations that reported a data breach incident in the following fiscal year and the remaining observations that did not experience a data breach incident over the period 2005 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

Dependent variable = Data breach incident (indicator)						
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Active ratio	-0.31**	-0.15*	-0.36**	-0.01	-0.35**	-0.35**
	(0.12)	(0.09)	(0.18)	(0.09)	(0.17)	(0.17)
log(ESOP/active participants)	-0.17***	-0.29***	-0.28***	-0.07	-0.06	-0.06
	(0.06)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Executive equity/Executive total pay	0.53*	1.04**	1.01**	1.08***	1.10***	1.10***
	(0.30)	(0.41)	(0.42)	(0.42)	(0.41)	(0.41)
Active ratio ×	0.54**		0.43		0.69**	0.69**
Executive equity/Executive total pay	(0.24)		(0.34)		(0.30)	(0.30)
log(wage)	0.23	0.72	0.71	0.59	0.58	0.58
	(0.29)	(0.45)	(0.47)	(0.38)	(0.39)	(0.39)
Risk committee (indicator)	0.15	-0.05	0.00	0.36	0.49	0.49
	(0.23)	(0.24)	(0.24)	(0.41)	(0.42)	(0.42)
Number of board committees	-0.01	0.01	0.01	-0.00	-0.01	-0.01
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Constant	-5.06	-7.33	-7.13	-7.32*	-7.16	-7.57*
	(3.18)	(4.98)	(5.16)	(4.38)	(4.51)	(4.40)
Observations	4,222	1,592	1,592	1,497	1,497	1,497
Independent variables	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.343	0.375	0.376	0.368	0.371	0.371
Industry*Year FE	N	Y	Y	N	N	N
State*Year FE	N	N	N	Y	Y	Y
State FE	N	N	N	N	N	Y

Table 29: Robustness check- Sample: the first report of a breach.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach incident in a given year, and zero otherwise. The sample consists of 86 firm-year observations that reported a data breach incident in the following fiscal year and the remaining observations that did not experience a data breach incidents over the period 2010 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

Dependent variable = Data breach incident (indicator)						
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Active ratio	-0.25**	-0.15**	-0.24	-0.05	-0.28*	-0.28*
	(0.10)	(0.07)	(0.16)	(0.09)	(0.17)	(0.17)
log(ESOP/participants)	-0.12**	-0.18**	-0.18**	-0.05	-0.05	-0.05
	(0.06)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Executive equity/Executive total pay	0.44	0.84**	0.85**	0.95**	1.01**	1.01**
	(0.30)	(0.39)	(0.39)	(0.44)	(0.43)	(0.43)
Active ratio ×	0.33		0.20		0.49	0.49
Executive equity/Executive total pay	(0.20)		(0.33)		(0.32)	(0.32)
log(wage)	0.19	0.22	0.19	0.57*	0.55	0.55
	(0.21)	(0.32)	(0.32)	(0.34)	(0.35)	(0.35)
Risk committee (indicator)	-0.25	-0.47	-0.46	-0.14	-0.08	-0.08
	(0.31)	(0.53)	(0.53)	(0.42)	(0.43)	(0.43)
Number of board committees	-0.06	-0.10*	-0.10*	-0.03	-0.03	-0.03
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Constant	-4.38*	-2.30	-2.04	-8.02**	-7.82*	-6.97
	(2.47)	(3.65)	(3.65)	(4.09)	(4.21)	(4.27)
Observations	3,983	1,143	1,143	859	859	859
Independent variables	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Pseudo R-squared	0.169	0.250	0.250	0.219	0.221	0.221
Industry×Year FE	N	Y	Y	N	N	N
State×Year FE	N	N	N	Y	Y	Y
State FE	N	N	N	N	N	Y

Table 30: Lags of dependent and key independent variables.
The table presents estimates of probit regressions in which the dependent variable is an indicator that takes the value one if a firm experiences a data breach incident in a given year, and zero otherwise. The sample consists of 125 firm-year observations that experienced a data breach in the following fiscal year (83 distinct ESOP firms) and the remaining firm-year observations that did not experience an incident during the period from 2005 to 2023. All explanatory variables are measured one year before the attack except for Tobin's q, which is measured two years before the attack. All control variables from Table 13 are included in estimation but not present in the table. The appendix provides detailed descriptions of the construction of the variables. Standard errors reported in parentheses are adjusted for heteroskedasticity and clustering at the firm level.***,** and * denote significance at the 1%, 5%, and 10% levels, respectively. Normalized values of the active ratio are used in this table.

Dependent variable = data breach incident (indicator)				
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**
Active ratio	-0.33***	-0.57**	-0.57**	-0.54**
	(0.11)	(0.24)	(0.24)	(0.24)
log(ESOP/active participants)	-0.11*	-0.15**	-0.15**	-0.15**
	(0.06)	(0.06)	(0.06)	(0.06)
Executive equity/Executive total pay	0.15	0.13	0.13	0.09
	(0.31)	(0.37)	(0.37)	(0.36)
Active ratio×	0.52**	0.55*	0.55*	0.51*
Executive equity/Executive total pay	(0.23)	(0.28)	(0.28)	(0.28)
L.breach	0.04		0.05	0.05
	(0.18)		(0.19)	(0.18)
L2.breach				0.27
				(0.20)
L.active ratio		0.44	0.44	0.43
		(0.27)	(0.27)	(0.27)
L2.active ratio		-0.28	-0.28	-0.28
		(0.27)	(0.27)	(0.28)
L3.active ratio		0.05	0.05	0.06
		(0.22)	(0.22)	(0.23)
Constant	-2.80	-3.26	-3.22	-3.20
	(2.93)	(3.75)	(3.69)	(3.59)
Observations	3,421	2,528	2,528	2,528
Independent variables	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Pseudo R-squared	0.245	0.242	0.242	0.245

Figures

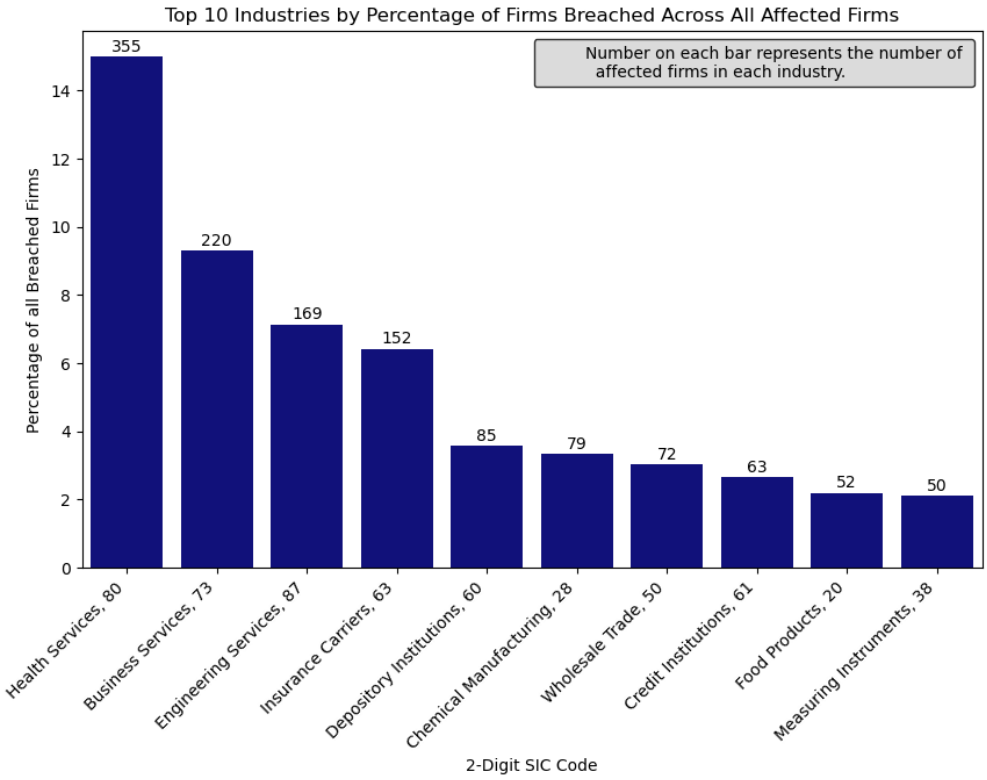


Figure 1: Percentage of Public and Private Firms with a Data Breach Incident in Each 2-Digit SIC Industry

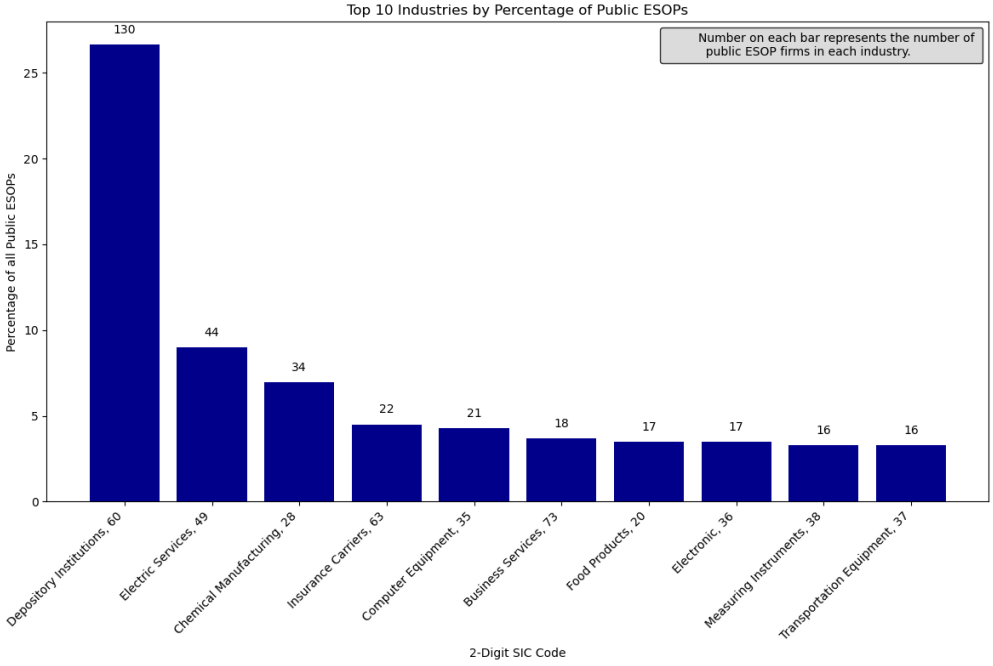


Figure 2: Percentage of Public ESOP firms in each industry

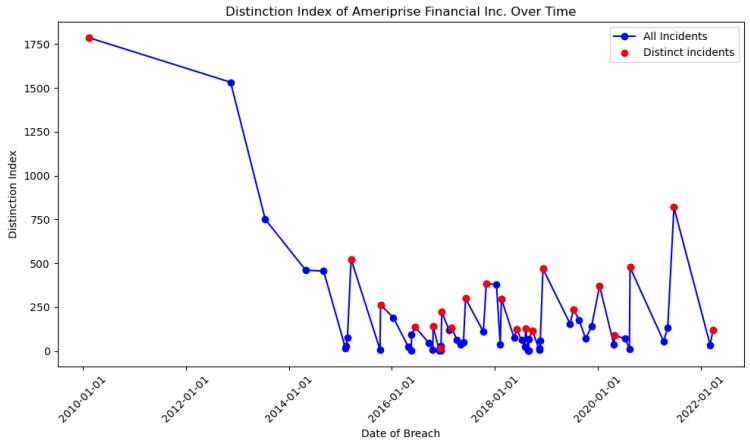
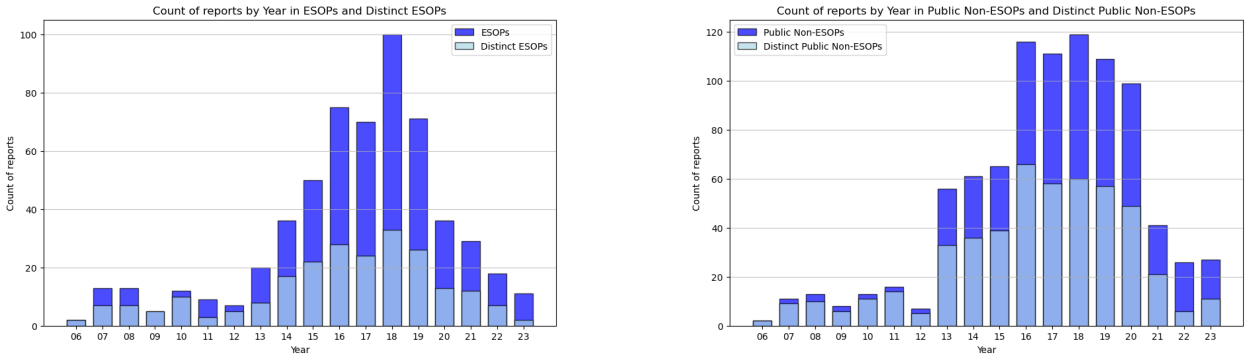


Figure 3: Ameriprise Financial Inc. incidents, including the distinct ones highlighted in red



((a)) Public ESOPs

((b)) Public Non-ESOPs

Figure 4: Distribution of reports in two samples of public ESOP firms and public non-ESOP firms

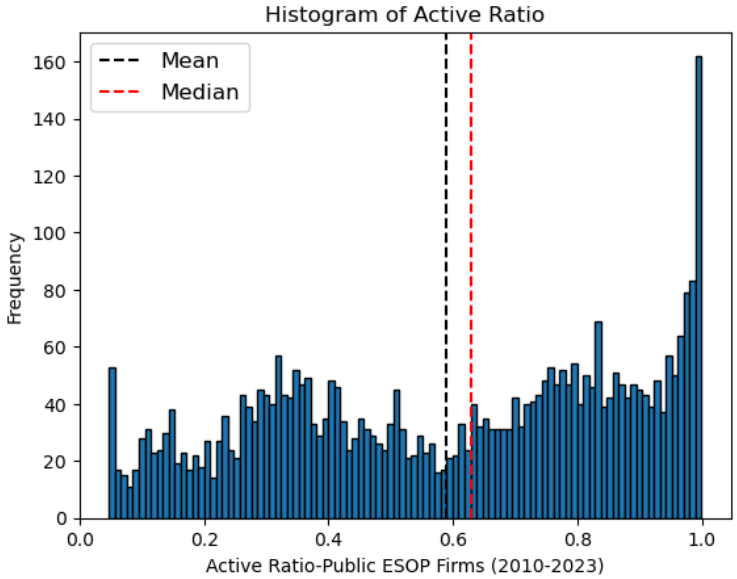


Figure 5: Distribution of active ratio

7.4 Data breach incidents as industry-wide shocks

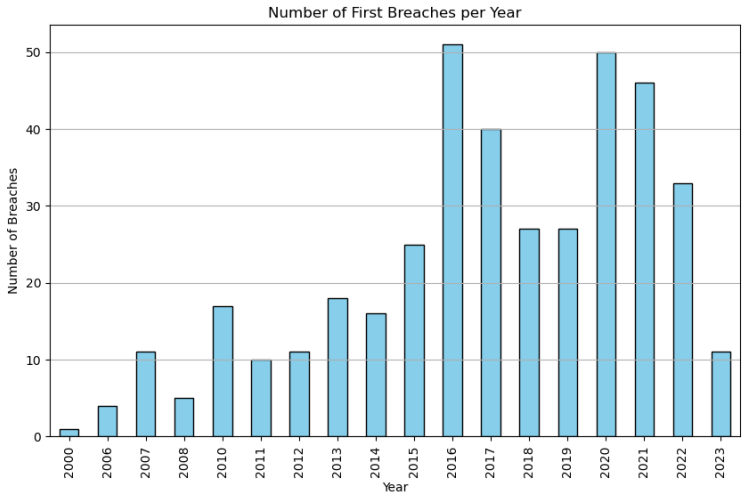
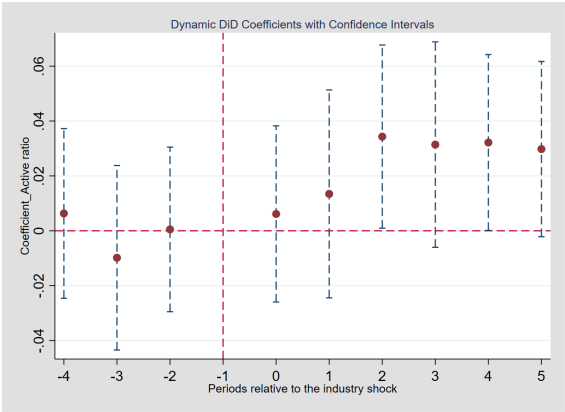
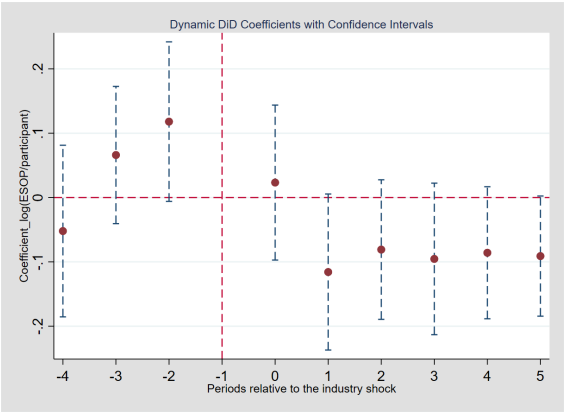


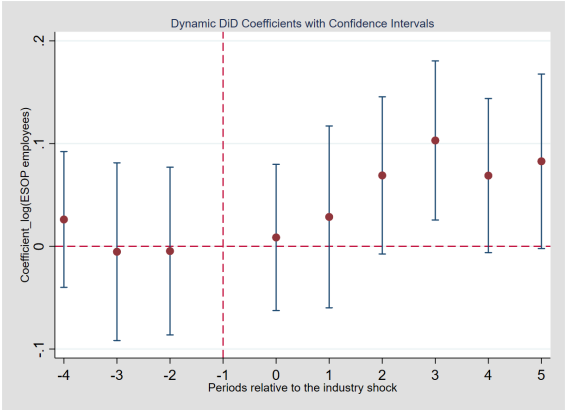
Figure 6: Distribution of industry shocks in the data breach dataset



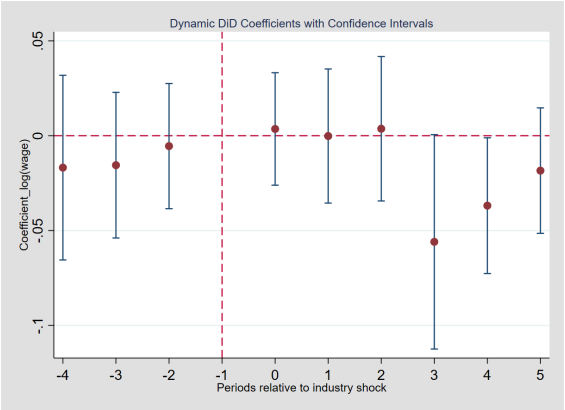
Active ratio



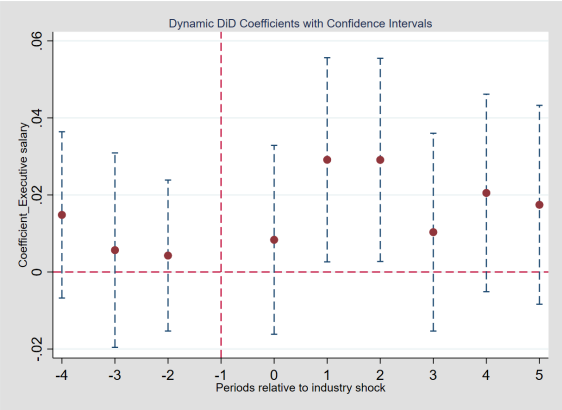
log(ESOP/participant)



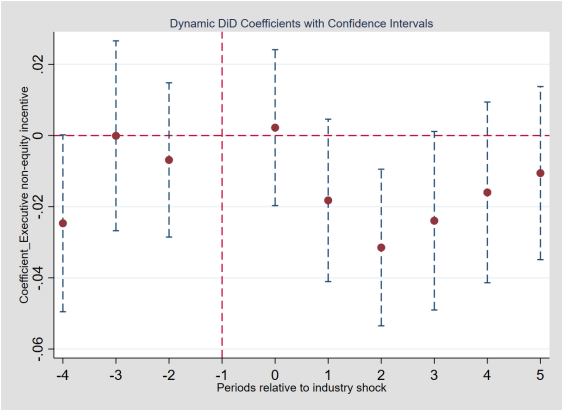
ESOP employees



log(wage)



Executive salary



Executive non-equity incentive

Figure 7: Investigating pre-shock trends for significant coefficients of the DID

Appendix A.

Table 32: Detailed descriptions of all the variables used in the tables.

Variable	Description	Source
Active participants/total employees	Active participants in the plan/ employees population	IRS Form 5500, Compustat
Assets in master trust	Assets in master trust (may include employer securities) (maseoy)	IRS Form 5500
Asset intangibility	1 –totalproperty, plant, and equipment (ppent)/ total assets (at)	Compustat, 10-K and 10-Q fillings
All executives salary	Total dollar amount of salary paid to all executives of a firm during a fiscal year	ExecuComp, DEF 14A
All executives total pay	Executives total pay including salary, option awards, stock units, bonus, etc.(tdc1)	ExecuComp, DEF 14A
CAPX/assets	Capital expenditures (capx)/total assets (at)	Compustat, 10-K and 10-Q fillings
CEO salary	Total dollar amount of salary paid to the CEO of a firm during a fiscal year	ExecuComp, DEF 14A
CEO total pay	CEO total pay including salary, option awards, stock units, bonus, etc.(tdc1)	ExecuComp, DEF 14A
CFO salary	Total dollar amount of salary paid to the CFO of a firm during a fiscal year	ExecuComp, DEF 14A
CFO total pay	CFO total pay including salary, option awards, stock units, bonus, etc.(tdc1)	ExecuComp, DEF 14A
ESOP assets per participant	Total ESOP assets/ active participants	IRS Form 5500
Employer securities in plan	Employer securities in plan (seceoy)	IRS Form 5500
Executive equity based compensation option awards for all executives	Sum of the grant-date value of stock awards and Execucomp	
Financial constraint (indicator)	The WW index derived by White and Wu (2006) based on an Euler equation approach from a structural model of investment. The WW index is a linear combination of six factors according to the following formula: WW = −0.091CF−0.062DIVPOS+0.021TLTD −0.044LNTA+0.102ISG−0.035SG, where CF is the ratio of cash flow to total assets; DIVPOS is a dummy variable that takes the value of one if the firm pays cash dividends; TLTD is the ratio of the long-term debt to total assets; LNTA is the natural log of total assets; ISG is the firm's three-digit industry sales growth; and SG is firm sales growth. Firms with a higher value of the WW index are more constrained. I rank firms based on the WW index and group the top (bottom) tertile into constrained (unconstrained) portfolios.	Compustat, 10-K and 10-Q fillings
Financial industry (indicator)	One for industries with SIC codes of 6000 and above and less than 7000, and zero otherwise	Compustat

Fortune 500 membership (indicator)	One if a firm is included in the list of Fortune 500 companies in a given year, and zero otherwise	Fortune.com, 50pros.com, Kaggle.com
Industry's Herfindahl index	index computed as the sum of squared market shares of firms' sales at the two-digit SIC industry level	Compustat
Industry's Tobin's q	Median Tobin's q of all firms in the same two-digit SIC code industries in a given year.	Compustat
Institutional block ownership	Max(0,Number of shares held by institutional shareholders that own more than 5% of a firm's equity scaled by the total number of shares outstanding)	Thompson 13F
Leverage	long term debt(dltt)+ debt in current liabeities(dlc)/ stock holder equity(seq)	Compustat, 10-K and 10-Q fillings
Max of ESOP assets per participant	Min of ESOP assets per participant+ assets in master trust	IRS Form 5500
Min of ESOP assets per participant	Max(Employer securities in plan, 51% * total plan assets)	IRS Form 5500
mean(executive equity/total pay)	Average equity portion of compensation across all executives	Execucomp and ED
Number of board committees	Number of board committees in a given fiscal year	Boardex
R&D/assets	Max (0, R&D expenditures (xrd))/total assets (at)	Compustat, 10-K and 10-Q fillings
Risk committee (indicator)	One if the name of a firm's board committee includes "risk," and zero otherwise	Boardex
ROA	Net income (ni)/total assets (at)	Compustat, 10-K and 10-Q fillings
Sales growth	$\text{Sales}_t / sales_{t-1}$	Compustat, 10-K and 10-Q fillings
Service industries	One for industries with SIC codes of 7000 and above and less than 9000, and zero otherwise	
Similarity measure	Result of the S-BERT model, which quantifies the similarity between the descriptions of two data breach incidents.	
Stock performance	Buy-and-hold return for the year net of the CRSP value-weighted index return	CRSP
Stock return volatility	Standard deviation of a firm's daily stock returns during a fiscal year	CRSP
Taxable income	(Federal tax + foreign tax)/top marginal corporate tax rate) - change in tax loss carry forward	Compustat, 10-K and 10-Q fillings
Tobin's q	(Total assets (at) – common/ordinary equity (ceq) + market value of equity (prcc_f × csho))/ total assets (at)	Compustat, 10-K and 10-Q fillings
Total ESOP assets	Total plan assets (taseoy)	IRS Form 5500
Transportation and communications	One for industries with SIC codes of 4000 and above and less than 4900, and zero otherwise	
Wage	Staff expense as reported in Compustat is divided by the total employee population. This item includes all elements of employee compensation, including salary, pension, etc. If it is missing, it is replaced by the median wage per employee in the corresponding 2-digit SIC code over the same fiscal year.	Compustat

Wholesale trade and retail trade (indicator)	One for industries with SIC codes of 5000 and above and less than 6000, and zero otherwise
Zero institutional ownership indicator	An indicator with value one if a firm has no fillings in Thompson 13F, and zero otherwise.
Zero R&D indicator	An indicator with value one if a missing R&D expense is replaced by zero, and zero otherwise

References

Acquisti, Alessandro, Allan Friedman, and Rahul Telang (2006) "Is there a cost to privacy breaches? An event study," *ICIS 2006 proceedings*, 94.

Akey, Pat, Stefan Lewellen, Inessa Liskovich, and Christoph Schiller (2021) "Hacking corporate reputations," *Rotman School of Management Working Paper* (3143740).

Aldatmaz, Serdar, Paige Ouimet, and Edward D Van Wesep (2018) "The option to quit: The effect of employee stock options on turnover," *Journal of Financial Economics*, 127 (1), 136–151.

Ashraf, Musaib (2022) "The role of peer events in corporate governance: Evidence from data breaches," *The Accounting Review*, 97 (2), 1–24.

Babenko, Ilona, Michael Lemmon, and Yuri Tserlukevich (2011) "Employee stock options and investment," *The Journal of Finance*, 66 (3), 981–1009.

Babenko, Ilona and Rik Sen (2016) "Do nonexecutive employees have valuable information? Evidence from employee stock purchase plans," *Management Science*, 62 (7), 1878–1898.

Babenko, Ilona and Yuri Tserlukevich (2009) "Analyzing the tax benefits from employee stock options," *The Journal of Finance*, 64 (4), 1797–1825.

Bana, Sarah, Erik Brynjolfsson, Wang Jin, Sebastian Steffen, and Xiupeng Wang (2022) "Human Capital Acquisition in Response to Data Breaches," *Available at SSRN 3806060*.

Banker, Rajiv D and Cecilia Feng (2019) "The impact of information security breach incidents on CIO turnover," *Journal of Information Systems*, 33 (3), 309–329.

Boasiako, Kwabena A and Michael O'Connor Keefe (2021) "Data breaches and corporate liquidity management," *European Financial Management*, 27 (3), 528–551.

Call, Andrew C, Simi Kedia, and Shivaram Rajgopal (2016) "Rank and file employees and the discovery of misreporting: The role of stock options," *Journal of Accounting and Economics*, 62 (2-3), 277–300.

Chang, Xin, Kangkang Fu, Angie Low, and Wenrui Zhang (2015) "Non-executive employee stock options and corporate innovation," *Journal of financial economics*, 115 (1), 168–188.

Core, John E and Wayne R Guay (2001) "Stock option plans for non-executive employees," *Journal of financial economics*, 61 (2), 253–287.

Ettredge, Michael, Feng Guo, and Yijun Li (2018) "Trade secrets and cyber security breaches," *Journal of Accounting and Public Policy*, 37 (6), 564–585.

Freeman, Richard, Douglas Kruse, and Joseph Blasi (2008) "Worker responses to shirking under shared capitalism,"Technical report, National Bureau of Economic Research.

Gatzlaff, Kevin M and Kathleen A McCullough (2010) "The effect of data breaches on shareholder wealth," *Risk Management and Insurance Review*, 13 (1), 61–83.

Hall, Brian J and Kevin J Murphy (2003) "The trouble with stock options," *Journal of economic perspectives*, 17 (3), 49–70.

Hannes, Sharon (2006) "Reverse monitoring: On the hidden role of employee stock-based compensation," *Mich. L. Rev.*, 105, 1421.

Hartmann, Caroline C and Jimmy Carmenate (2021) "Academic research on the role of corporate governance and IT expertise in addressing cybersecurity breaches: Implications for practice, policy, and research," *Current issues in auditing*, 15 (2), A9–A23.

He, Chris Zhijian, Tracie Frost, and Robert E Pinsker (2020) "The impact of reported cybersecurity breaches on firm innovation," *Journal of Information Systems*, 34 (2), 187–209.

Hochberg, Yael V and Laura Lindsey (2010) "Incentives, targeting, and firm performance: An analysis of

non-executive stock options," *The Review of Financial Studies*, 23 (11), 4148–4186.

Hsu, Carol and Tawei Wang (2014) "Exploring the association between board structure and information security breaches," *Asia pacific journal of information systems*, 24 (4), 531–557.

Huang, Henry He and Chong Wang (2021) "Do banks price firms' data breaches?" *The Accounting Review*, 96 (3), 261–286.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M Stulz (2021) "Risk management, firm reputation, and the impact of successful cyberattacks on target firms," *Journal of Financial Economics*, 139 (3), 719–749.

Kim, E Han and Paige Ouimet (2014) "Broad-based employee stock ownership: Motives and outcomes," *The Journal of Finance*, 69 (3), 1273–1319.

Kong, Dongmin, Jia Liu, Yanan Wang, and Ling Zhu (2023) "Employee stock ownership plans and corporate environmental engagement," *Journal of Business Ethics*, 1–23.

Kumar, Alok, Jeremy K Page, and Oliver G Spalt (2011) "Religious beliefs, gambling attitudes, and financial market outcomes," *Journal of financial economics*, 102 (3), 671–708.

Lending, Claire, Kristina Minnick, and Patrick J Schorno (2018) "Corporate governance, social responsibility, and data breaches," *Financial Review*, 53 (2), 413–455.

Lin, Lihua and Zhengyu Zhang (2022) "Interpreting the coefficients in dynamic two-way fixed effects regressions with time-varying covariates," *Economics Letters*, 216, 110604.

Masulis, Ronald W, Cong Wang, and Fei Xie (2020) "Employee-manager alliances and shareholder returns from acquisitions," *Journal of Financial and Quantitative Analysis*, 55 (2), 473–516.

Oyer, Paul (2004) "Why do firms use incentives that have no incentive effects?" *The Journal of Finance*, 59 (4), 1619–1650.

Oyer, Paul and Scott Schaefer (2005) "Why do some firms give stock options to all employees?: An empirical examination of alternative theories," *Journal of financial Economics*, 76 (1), 99–133.

Pagano, Marco and Paolo F Volpin (2005) "Managers, workers, and corporate control," *The journal of finance*, 60 (2), 841–868.

SUN, Yuan (2021) "Data security and M&A."

Wang, Qian, Eric WT Ngai, Daniel Pienta, and Jason Bennett Thatcher (2023) "Information Technology Innovativeness and Data-Breach Risk: A Longitudinal Study," *Journal of Management Information Systems*, 40 (4), 1139–1170.

Whited, Toni M and Guojun Wu (2006) "Financial constraints risk," *The review of financial studies*, 19 (2), 531–559.

Wooldridge, Jeffrey M (2023) "Simple approaches to nonlinear difference-in-differences with panel data," *The Econometrics Journal*, 26 (3), C31–C66.

Wu, Fang, June Cao, and Xiaosan Zhang (2023) "Do non-executive employees matter in curbing corporate financial fraud?" *Journal of Business Research*, 163, 113922.

Zhang, Haofei, Jin Peng, Juan Mao, and Shouhuai Xu (2024) "Repeated data breaches and executive compensation," *Applied Economics Letters*, 1–10.