# WHAT WORKS? QUASI-EXPERIMENTS IN CYBERSECURITY POLICY INTERVENTIONS

A Dissertation
Presented to
The Academic Faculty

By

Karl Grindal

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Georgia Institute of Technology
School of Public Policy

Georgia Institute of Technology

August 2021

# WHAT WORKS? QUASI-EXPERIMENTS IN CYBERSECURITY POLICY INTERVENTIONS

Thesis committee:

Dr. Milton Mueller, Advisor
School of Public Policy
*Georgia Institute of Technology*

Peter Swire
Scheller College of Business
*Georgia Institute of Technology*

Dr. Juan Rogers
School of Public Policy
*Georgia Institute of Technology*

Dr. Sasha Romanosky
*RAND Corporation*

Dr. Hans Klein
School of Public Policy
*Georgia Institute of Technology*

Date approved: June 2, 2021

Errors using inadequate data are much less than those using no data at all.

*Charles Babbage*

For my nieces, Eira and Kara, may your parent’s sharing of baby photos be your only loss of privacy.

# ACKNOWLEDGMENTS

I am grateful to have received throughout this dissertation process excellent academic guidance and been encouraged by understanding friends and family. First, I would like to thank my advisor, Professor Milton Mueller; his advice was invaluable in formulating the research question and focusing my effort. Next, I would like to thank the other members of my committee for their constructive engagement. I am also grateful for the support of the Science, Technology and Innovation Policy (STIP) Program and the advising of Dr. Jan Youtie at the Enterprise Innovation Institute. Their support in the Summer of 2019 allowed me to get an early start using breach data. As government employees compiled much of the data I worked with as part of open records requests, I would also like to thank these unknown bureaucrats. Their work has helped me to knit together a more comprehensive picture of the breach landscape. I cannot begin to express my thanks to Jason Healey. Before joining Georgia Tech, Jason served as my mentor and introduction to the field of cyber conflict; I would not be here without him.

Essential to doctoral studies is a strong cohort. These past five years at Georgia Tech, I have been fortunate to have close friends and colleagues in my fellow Ph.D. students. In particular, thanks to Sorawit Siangjaeo, Daniel Sanbeg, Seokkyun Woo, Karim Farhat, and Daniel Schiff. Thank you to my dog Woodrow whose affection and devotion keep me sane. Thanks to my brother Erik Grindal and his wife Shannon for their constant support and love. Thanks to my parents, Michele and Ted Grindal, they prepared me for the world and have always encouraged me to be my best. A special thanks to my partner, Beena Meena, for her patience and love. She has stood by me through COVID and the doldrums of dissertation work.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ACRONYMS

**ACF** autocorrelation function

**ARIMA** autoregressive integrated moving average

**ARRA** American Recovery and Reinvestment Act

**ASN** Autonomous System Number

**CAPs** corrective action plans

**CERT** Computer Emergency Response Team

**CFO** Chief Financial Officers Act

**CFR** Code of Federal Regulations

**CISA** Cybersecurity Information Sharing Act

**CISO** Chief Information Security Officer

**CMR** Code of Massachusetts Regulations

**CMS** Centers for Medicare and Medicaid Services

**COPPA** Children’s Online Privacy Protection Act

**CRS** Congressional Research Service

**DBIR** Data Breach Investigations Report

**DDoS** Distributed Denial of Service

**DUNS** Data Universal Numbering System

**ePHI** electronic public health information

**FAIR** Factor Analysis of Information Risk

**FCRA** Fair Credit Reporting Act

**FDA** Food and Drug Administration

**FIPS** Federal Information Processing Standards

**FISMA** Federal Information Security Management Act

**FOI** Freedom of Information

**FTC** Federal Trade Commission

**GDPR** General Data Protection Regulation

**GLBA** Gramm-Leach-Bliley Act

**HHS** Health and Human Services

**HIPAA** Health Insurance Portability and Accountability Act

**HIT** health information technology

**HITECH Act** Health Information Technology for Economic and Clinical Health Act

**IAPP** International Association of Privacy Professionals

**ITS** interrupted time series

**KPSS** Kwiatkowski-Phillips-Schmidt-Shin

**KS** Kolmogorov Smirnoff

**MDDI** Million Dollar Directory

**NAICS** North American Industry Classification System

**NCSL** National Conference of State Legislatures

**NVD** National Vulnerability Database

**NY DFS** NY Department of Financial Services

**NYCRR** New York Codes, Rules and Regulations

**OCABR** Office of Consumer Affairs and Business Regulation

**OCR** Office of Civil Rights

**OMB** Office of Management and Budget

**ONC** Office of the National Coordinator for Health Information Technology

**PACF** partial autocorrelation function

**PCI** Payment Card Industry

**PCI DSS** Payment Card Industry Data Security Standard

**PHI** protected health information

**PII** Personally Identifiable Information

**RFI** request for information

**SHIELD Act** Stop Hacks and Improve Electronic Data Security Act

**SOX** Sarbanes–Oxley Act

**VERIS** Vocabulary for Event Recording and Incident Sharing

**WISP** written information security program

# SUMMARY

Given the significance policymakers place on cybersecurity, how effective has a decade of policy interventions been at reducing social costs? This paper uses the limited regulations implemented by States and United States government agencies as quasi-experiments. This work measures regulatory efficacy by compiling mandatory state-level data breach reports to create novel breach incident data sets. A reduction in breach frequency serves as the kind of measurable outcome that regulators would intend cybersecurity policy interventions to address. To this end, I evaluate four cybersecurity regulations: the Massachusetts Data Security Law, the Health Information Technology for Economic and Clinical Health Act (HITECH Act), Federal Trade Commission (FTC) Section 5 enforcements against Wyndham Hotels, and the NY Department of Financial Services (NY DFS) cybersecurity regulations.

I assessed each regulatory intervention as a quasi-experiment, employing segmented time-series regressions to evaluate the relative change in reported data breaches. These quasi-experiments controlled for policy implementation phases and reporting requirements. As these policies have overlapping aims (creating information security programs), we can infer whether this meta-regulatory approach, the encouragement of self-regulation by industry with corresponding civil penalties, has been an effective regulatory strategy. An effectively regulatory system would sufficiently motivate the targeted population to improve their cyber posture, such that there was a reduction in breach reporting. Ultimately, three of the cases discussed did not show an impact. However, analysis of the NY DFS regulations suggests a meaningful decrease of approximately 27 breaches in the following year.

Comparing these regulations shows differences in scope, content, and penalties that may explain this disparate level of impact. Next, the efficacy of NY DFS regulations is

placed in context with a discussion of potential savings and the duration of the impact. While demonstrating that cybersecurity regulations can meaningfully reduce breaches, this work suggests that this effect is neither generalizable across diverse contexts nor a satisfactory solution to the complex and pervasive issues associated with identity theft, fraud, and cyber crime.

Overall, these findings suggest potential promise in this methodology for the policy evaluation of data security laws and regulations. Policymakers could improve these assessments by standardizing the reporting of mandatory breach notification data so that policy efficacy can be better measured. Because of its similarity to the NY DFS regulations, this finding may also provide preliminary empirical evidence for the Insurance Data Security Model Law propagated by the National Association of Insurance Commissioners. Drawing on this methodology, this model legislation and other data security and privacy regulatory interventions should now be the subject for future research. The first step for policy makers seeking to design rules to protect citizen's privacy and security is knowing what works?

# CHAPTER 1
# INTRODUCTION AND BACKGROUND

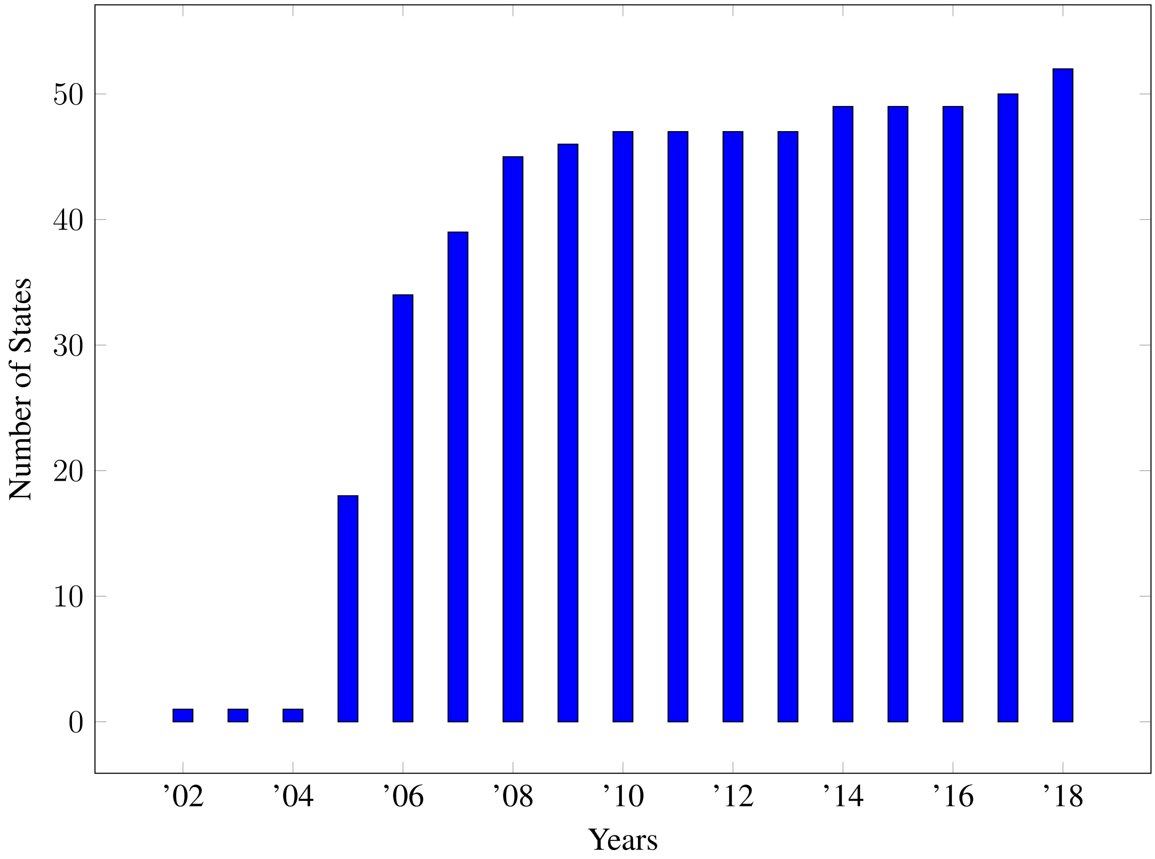
What works? For decades, United States legislators and regulators have sought to create a more secure cyberspace. The complexity of this subject area has led the United States to pursue a strategy of regulatory restraint in the belief that a rapidly evolving private sector would develop and implement best practices. Yet, within a limited number of states and nationwide industries, regulators have mandated private sector companies to adopt information security programs to protect consumer data. Despite state and federal experimentation, these interventions have not been subject to policy evaluation. Without assessing the effectiveness of these interventions, cybersecurity and privacy policies are operating in the dark.

However, because of its motley regulatory regimes, these US interventions create potential quasi-experiments. This work applies traditional policy evaluation methods to the novel domain of cyberspace. The four regulatory interventions selected are the Massachusetts Data Security Law, the HITECH Act, FTC Section 5 enforcement against Wyndham Hotels, and the NY DFS cybersecurity regulations. All four regulatory regimes share a policy logic of mandating the adoption of a specified cybersecurity program to protect Personally Identifiable Information (PII). These cybersecurity regulations share or expand on a mandate that companies should draft policies, appoint a manager to oversee the policies, and encrypt consumer data. Regulators enforce these rules with penalties for covered entities who are negligently or willfully non-compliant.

What metric can characterize the efficacy of these regulations? As mandatory state-level reporting of breaches precedes these policies' implementation, it creates a potential measure of effectiveness. Evaluating the frequency of breaches (often scaled for state population size) in the months preceding and following a regulatory intervention and comparing this to a control population produces a practical empirical test. So, what works? Ultimately, this research identifies the NY DFS cybersecurity regulations preventing approximately 27 breaches in the year following enactment. However, as the other three interventions did not significantly reduce breaches, the effectiveness of regulations can not be generalized.

Given that legislators and regulators prioritize cybersecurity issues, why has policy evaluation been ignored? This absence was primarily a result of the inadequacy of available data. Existing sources lacked state and industry variables from which to create control groups. This study overcame this obstacle, restructuring existing state-reported breach incident datasets enabling cross-state and cross-industry comparisons.

Figure 1.1: Number of States with Data Breach Notification Laws by Year



Data breach notification has a unique policy history, creating a mandatory reporting regime that at least in principle is comprehensive. Starting in 2002, California passed the nation’s first data breach notification legislation (*An Act to Amend, Renumber, and Add*

*Section 1798.82 of, and to Add Section 1798.29 to, the Civil Code, Relating to Personal Information.* 2002). This legislation required that companies notify the state residents whose records were lost. In 2005, California's legislation was put to the test when the Georgia-based company, ChoicePoint Inc, notified 35,000 California state residents of a data breach and voluntarily notified 111,000 customers located outside California (*United States of America v Choicepoint Inc.* 2010). This incident led to rapid policy diffusion as other states adopted similar legislation. Alabama was the last state to implement a data breach notification law for consumers in 2018 (Hosp and Drum 2018).

While this initial legislation required notification of consumers, many states subsequently amended this legislation to expand the notification requirements to include consumer reporting companies and government agencies. With the collection of breach notification records by state governments, this information was publicly available by 18 states as of January 2020 (*U.S. State Data Breach Lists*, 2020). Additional data breach notification records have been made available through state-level freedom of information requests. The following table shows state-based breach notification data availability over time as the collection percentage during a given year.

Why should we expect policy interventions to reduce breaches? The relationship between laws/regulations and data breach reporting is not a direct one. There is instead a long causal chain that links policy interventions to better security to fewer breaches. Nevertheless, Governors and regulators would often make this link themselves when enacting these policies. As Governor Cuomo said of the NY DFS regulations, they should "guarantee the financial services industry upholds its obligation to protect consumers and ensure that its systems are sufficiently constructed to prevent cyber-attacks to the fullest extent possible."

Consistent with this logic, Figure 1.2 shows a causal diagram modeling how data breach factors may relate. Arrows demonstrate the sequence of expected effects and the accompanying plus or minus signs indicate the expected direction of correlation. This diagram

Table 1.1: State Data - Percent of Collection for each Year

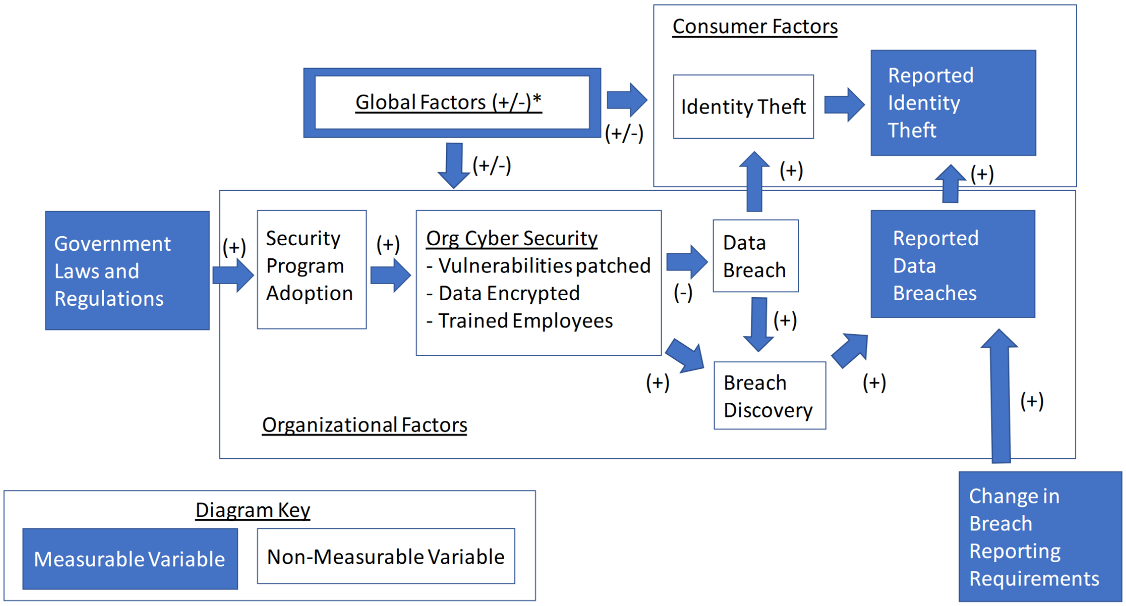
| **State** | **’05** | **’06** | **’07** | **’08** | **’09** | **’10** | **’11** | **’12** | **’13** | **’14** | **’15** | **’16** | **’17** | **’18** | **’19** | **’20** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| North Carolina | 85 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | |
| New Hampshire | | 53 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 83 |
| Hawaii | | | 47 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 97 | |
| Massachusetts | | | 46 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 42 |
| South Carolina | | | | 43 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 48 | |
| Maine | | | | 42 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 30 |
| California | | | | | | | | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 30 |
| Wisconsin | | | | | | | | 69 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 28 |
| Connecticut | | | | | | | | 25 | 100 | 100 | 100 | 100 | 100 | 100 | 77 | |
| Indiana | | | | | | | | | 72 | 100 | 100 | 100 | 100 | 100 | 100 | 26 |
| Maryland | | | | | | | | | | | 99 | 100 | 100 | 100 | 100 | 08 |
| Montana | | | | | | | | | | | 65 | 100 | 100 | 100 | 100 | 30 |
| Washington | | | | | | | | | | | 39 | 100 | 100 | 100 | 100 | 29 |
| Iowa | | | | | | | | | | | 28 | 100 | 100 | 100 | 100 | |
| Oregon | | | | | | | | | | | 17 | 100 | 100 | 100 | 100 | 31 |
| Nebraska | | | | | | | | | | | | | | 100 | 100 | |
| Vermont | | | | | | | | | | | | | 87 | 100 | 100 | 30 |
| Delaware | | | | | | | | | | | | | | 72 | 81 | |
| New Mexico | | | | | | | | | | | | | 51 | 100 | 75 | |
| North Dakota | | | | | | | | | | | | | | 99 | | |

shows measurable variables placed in blue boxes and non-measurable variables shown in white. While the regulations explored in this paper mandate particular firm behavior, broader global factors (i.e., the internet-connected population, cybersecurity literacy, new vulnerabilities, cyber hygiene) would also shape company behavior. Consequent to reducing data breaches, effective legislation would be expected to benefit consumers, as demonstrated by a reduction in reported identity theft.

This diagram demonstrates some of the presumptions underlying this policy analysis. One critical assumption is that increased organizational cybersecurity should decrease the number of data breaches more than it increases the number of breaches discovered. This supposition is not inconsequential, as an improvement to organizational cybersecurity maturity would reasonably improve threat detection.

Another essential premise is that various global factors would equally affect the treatment and non-treatment populations. Most of the relevant global factor data collected for

Figure 1.2: Diagram of Factor Relationships.



this research did not match the unit of analysis for state or industrial data used in the quasi-experiments. To the extent possible, the models used in this research control for global factors (checking, for example, that breach reporting requirements did not meaningfully change during the experimental period). While not ideal, these assumptions are necessary given the availability of historical data to make educated hypotheses.

Control populations could be identified in three of the four cases. For the Massachusetts Data Security Law, we compared reporting in Massachusetts with New Hampshire and North Carolina. With the HITECH Act, the health sector was compared with finance and all other sectors. For the NY DFS regulations, breaches affecting the New York finance sector were compared with breaches affecting financial firms outside the state of New York. The FTC Wyndham Hotel suit did not present a viable control. Instead, an interrupted time-series assesses the effect of two treatments, the initial FTC complaint and the subsequent Third Circuit District Court's decision through the settlement. Despite the inferential power of these quasi-experiments, only the NY DFS regulations showed a statistically significant measure of efficacy.

Why should we expect these laws to work? For one, the drafters of these regulations believed that they could, discussed in subsection 2.3.4. Further, the policy logic embedded in the regulations creates the prospect for real financial penalties costing companies millions of dollars. These laws frequently serve as models for future legislative and regulatory actions. While predating it by almost a decade, the Massachusetts Data Security Law, passed in 2010, resembles the Stop Hacks and Improve Electronic Data Security (SHIELD) Act in New York in 2019. The NY DFS regulations have served as a basis for model insurance legislation advanced by the National Association of Insurance Commissioners and passed by thirteen states as of April 2021. Policy makers should be aware of the context where these regulatory practices have worked and where they have not. These findings would allow for the limited time, resources, and agenda setting given to this topic to be focused.

In addition to core findings related to the quasi-experiments, supplementary descriptive findings related to spatial and temporal factors were also revealing. One finding shows an approximately 20% increase years-over-year in reported breaches by covered states. Additionally, preliminary evidence for a seasonal trend with a peak for breach reporting in the Spring and a drop in the Fall. Evidence also shows that approximately half of breaches are local in their effect, only reported by affected organizations in one of the covered states.

This document comprises chapters to summarize relevant literature (chapter 2), describe the selected cases (chapter 3, review the developed methodology (chapter 4), analyze findings (chapter 5), and explore their consequent implications (chapter 6). Given that the NY DFS finding showed a meaningfully beneficial effect, how did it differ from the other regulatory interventions? The implications chapter investigates differences in scope, content, implementation periods, and penalties between the legislation. It also attempts to estimate the monetary savings from the avoided breaches and hypothesizes on why this impact might not persist beyond the initial year of implementation. Given this context, questions regarding policy design and future work are addressed. Only then can we move to imagine a future where instead of asking “what works?” we ask “what works better?”.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Relevant Literature

Dean (2016) anticipated this research with the article “Natural and Quasi-Natural Experiments in Cybersecurity Policies.” Dean proposed an approximation of the quasi-experimental methodology used in this study, stating, “[u]sing a difference-in-differences methodology, one could conduct a quasi-natural experiment to determine the impact of mandatory data breach notification laws and regulations in the United States” (156). This article went on to highlight how a researcher might produce both inter-state and inter-firm sub-populations for comparison.[1]

Only a few researchers have made attempts to evaluate the efficacy of cybersecurity policies. Notable examples include Romanosky et al. (2011, 281), who connected data breach notification laws to a 6.1% reduction in identity theft. Kesari (2020, 18) noted that updates in 2016 to California data breach notification suggest “.1 fewer reports per 100,000 people” for reported medical identity theft. Liu (2020) found that state anti-phishing or credit freeze legislation did not impact annual identity theft reports. However, this literature has been slow to develop, despite governmental action and investments in cybersecurity remaining a governmental priority (Homan 2021).

While cybersecurity has not yet been the subject of rigorous legislative and regulatory policy evaluation, significant literature exists around the topic of data breaches. Much of this work has focused on descriptive efforts to characterize the domain. To ensure that the subsequent literature review was comprehensive, I created a propositional inventory of

[1] One specific challenge of Dean’s proposed study is its focus on “notification laws and regulations.” A difference-in-difference model could not accurately characterize cybersecurity efficacy if data breach reporting laws were either newly implemented or modified the requirements around reporting.

all articles on Scopus that referenced either “cybersecurity policy” or “data breach.” Additions to the literature review were less systemic, identifying relevant articles using search terms like “security breach,” “information breach,” “policy evaluation,” and “cybersecurity” within Google Scholar.

Topics of particular interest to the literature include: measuring the size and frequency of breaches (Wheatley, Maillart, and Sornette 2016; Edwards, Hofmeyr, and Forrest 2016) and estimating the costs of these incidents either looking at the effect on market value (Cavusoglu, Mishra, and Raghunathan 2004), or the cost of lawsuits against breached companies (Romanosky, Hoffman, and Acquisti 2014), or a model combining both (Romanosky et al. 2017). Given the corporate information associated with these data sets, researchers have also used breach data to predict the risk of breaches (Xu et al. 2018; Sarabi et al. 2016). Other work has attempted to see how respondents react to public breach disclosure by measuring its effect on consumer risk perceptions and behavior (Yixin Zou et al. 2018) or public reports of identity theft (Romanosky, Telang, and Acquisti 2011). As data breaches have increasingly become an assumed risk by businesses, researchers have started to tie corporate concerns about breaches to insurance policies. This relationship can provide insight into business perceptions of this risk by measuring the premiums they are willing to pay (Franke 2017) or the services and coverage gaps of these policies (Romanosky et al. 2017).

Quantitative findings relevant to the data breach literature have advanced on several fronts as data collection has improved and researchers have adopted more complex statistical methods. Eling and Wirfs (2019, 1112) identified a skewness in the loss distribution where “the mean loss (US$ 43.49 million) is higher than the median loss (US$ 1.54 million).” Eling and Wirfs (2019) model data breach size with a Pareto distribution (consistent with Wheatley, Maillart, and Sornette 2016) rather than with a log-normal distribution (Edwards, Hofmeyr, and Forrest 2016). Most critically for this work, Eling and Wirfs’ (2019) findings suggest an increase in breach frequency but a decrease in extreme losses over time.

Using the same dataset, Xu (2018, 2856) found a similar result, stating the "threat of cyber hacks is indeed getting worse in terms of their frequency, but not in terms of the magnitude of their damage." One hypothesis for this finding could be that certain larger companies complying with stricter regulatory regimes are becoming less susceptible to high-impact incidents. Relevant to the proposed research, an understanding that breach size and corporate losses are non-linear shapes how companies may perceive the risk of regulatory penalties.

Additional empirical findings related to data breaches measure the relationship between separate incidents. Xu (2018) recommends stochastic process models over distributions when modeling hacking breaches inter-arrival times and breach size due to issues with autocorrelation. Eling and Jungs (2018) look at cross-industry correlation and cross-breach correlation, finding cross-breach (i.e., risk factors) correlate more significantly than cross-industry. Concerns with autocorrelation and partial autocorrelation informed several methodological decisions discussed in chapter 4 and chapter 5.

As this work attempts to characterize how incentives shape company behavior, corresponding economistic literature have developed valuable models. To this end, Laube (2016, 38) notes that "a breach notification law with security audits and sanctions can incentivize firms to report breaches to authorities, regardless of accompanied disclosure costs." While subtle, this finding speaks to a particular concern with the proposed research; a stricter regulatory regime may be increasing reporting behavior. To disaggregate reporting from reduced breaches, one could test the effects of cybersecurity regulations on identity theft reporting while controlling for data breaches. However, in all four cases, either the dates of the interventions did not overlap with the identity theft data collected, or the identity theft data does not include intra-state industry-level reports. If reporting is overdetermined and not correlated with sanctions, this research shifts from a focus on corporate reporting to a measure of breach frequency. Gao (2015, 425) asserts that "security breach probability ... decreases with the efficiency of security investment and the related firm's monetary loss."

The incentives of regulatory penalties demonstrate additional potential monetary losses and would thus be hypothesized to drive cybersecurity investments.

## 2.2 Data Sources

In the literature described above, the researchers drew on pre-existing breach datasets. While this allowed these researchers to avoid working with primary source documents, secondary sources restrict what questions researchers can address using the available data.

- **Database 1** Advisen's Cyber Loss Data
- **Database 2** DataLossDB.org
- **Database 3** Hackmageddon
- **Database 4** HHS Breach Portal
- **Database 5** Privacy Rights Clearinghouse
- **Database 6** SAS® OpRisk Global Data
- **Database 7** VERIS Community Database (VCDB)
- **Database 8** Identity Theft Resource Center

Perhaps the most frequently employed source for data breach research has been the Privacy Rights Clearinghouse (Database 5). Because of its rich coverage and accessibility, Clearinghouse appeared to be the most popular platform for database research. Consequently, Eling and Jung (2018), Xu et al. (2018), Edwards et al. (2016), and Wheatley et al. (2016) all used this data source. At some point, Clearinghouse populated their dataset with records from the datalossdb.org project (Database 2). Database 2 is no longer being maintained but was used by researchers before the site owners shut it down. Romanosky (2014) used this data set to investigate the likelihood of lawsuits and settlements, finding that the "odds of a settlement are found to be ten times greater when the breach is caused by a cyber attack." Romanosky (76) also found that lawsuits were 3.5 times more likely when victims incurred financial losses and six times less likely when the breached entity provided free credit monitoring.

The Vocabulary for Event Recording and Incident Sharing (VERIS) Community Database (Database 7) provides a GitHub-based information-sharing resource that feeds into the annual Verizon Data Breach Investigations Report (DBIR). This data is compiled from information-sharing partners and provides extensive incident metadata; however, it appears to anonymize the names of the affected organizations. Because of its rich organizational and technical level metadata, VERIS has been used to predict the likelihood of a breach based on enterprise characteristics by Sarabi (2016) and Liu et al. (2015). The site hackmageddon.com (Database 3) is a comprehensive private sector data source that its creator Paolo Passeri has updated since 2014. Rarely used exclusively, Liu et al. (2015) and Fielder et al. (2014) combined hackmageddon.com data with Verizon data.

Health and Human Services (HHS) (Database 4) has mandated data breach reporting since 2009 for incidents affecting 500 or more individuals. Liu et al. (2015) attempted to fit a linear growth curve to incident frequency but did not find statistical significance but found confidence intervals on a range of additional factors. Bai et al. (2017) used HHS data to find a statistically significant relationship between the risk of data breaches and hospital size and status as a major teaching hospital. Related research findings of mine prepared for the Science, Technology and Innovation Policy (STIP) Program at Georgia Tech before this dissertation work matched state data to Compustat records for publicly listed firms and produced similar positive statistical correlations with firm size.

Increasingly, private sector firms have sought to develop incident datasets using proprietary monitoring strategies and Freedom of Information (FOI) requests (that proceed my own). While researchers have employed these data sources with large incident counts, they appear to be one-off arrangements. Example uses of this kind of data include the work by Eling and Wirfs (2019), which relied on SAS® OpRisk Global Data (Database 6), and Romanosky (2016), which used data from Advisen Ltd (Database 1). Romanosky's work demonstrated an interesting sectoral finding that companies in the information and retail industry experience the highest losses from a data breach and identify the average data breach as costing firms less than 200k.

## 2.3 Theoretical Framing

Organizational cybersecurity practices cannot be understood without being placed in the context of the economic incentives of information system operators. Literature on the economics of information security explains how concepts drawn from economics shape these operator incentives, expounding on concepts such as asymmetric information, externalities, and collective action problems. The following section explores prior work into the economics of information security literature to place the breach-specific literature in context and articulates the tacit theories of change held by legislators and regulators.

### 2.3.1 Economics of information security literature

The economics of information security literature builds on a broad economics literature that explores how incentives shape communications technologies developed in the 1990s and 2000s. This earlier literature covered a wide range of topics, including platform markets (Parker and Van Alstyne 2005), social production (Benkler 2002), digital convergence (Mueller 1999), and spectrum markets (Hazlett 1990). Not until communications technologies became ubiquitous and hacking incidents started to cause economic damages did the confidentiality and accessibility of data become an issue of significant concern to economists. Information security literature began in earnest in the mid-2000s as an extension of this analysis, exploring how economic factors and incentives shape the cybersecurity practices of the private sector.

Early research by Anderson and Moore (2006) identified a range of economic factors as playing a role in the economics of information security. This research identified misaligned incentives in the design and deployment of computer systems, noted the impact of externalities, described the market for software as “a market for lemons,” and foresaw the potential for insurance markets to distribute risk for low probability cyber events. Bauer and van Eaten (2009) go beyond merely summarizing potential economic issues shaping cyber behavior and propose a system where decentralized decisions create emergent patterns that reflect a mix of positive and negative externalities. They also suggest that to correct these misaligned incentives for diffused stakeholders, “[e]nhancing cybersecurity at a broader level will have to overcome this coordination and cooperation issue: it is a collective action problem” (Bauer and van Eeten 2009, 715). These collective action challenges are significant because of inter-dependency. Kunreuther and Heal (2003) explored the application of interdependent security, drawing on contagion models. They found that an “institution’s vulnerability depends not only on how it manages its risks but also on how other unrelated entities manage their risks” (232).

While initial work often characterized potential challenges with measuring hacking-related losses, later work would try to start and address these limitations. In particular, numerous efforts have sought to identify the costs associated with cybercrime. Anderson et al. (2013) proposed a framework that decomposed the costs of cybercrimes into three categories: direct costs, indirect losses, and defense costs. While the affected party experiences direct losses, indirect losses produce social costs like decreased trust in online services; defense cost measures the totality of security products, services, and public policy instruments used to prevent successful attacks. A thesis by Lenchik (2016) simulated another form of problem evolution. Even as cybersecurity measures may reduce the risk of identity theft to an online banking platform, the subsequent growth in the platform may make it a more attractive target for identity theft in the future. One particular challenge to achieving secure systems is the ability of hackers to adapt to changes in organizational defenses such that “tightened security may eventually lead to even more malicious forms of intrusion” (Asghari, van Eeten, and Bauer 2016, 281).

On the solution side, a Brookings Institute presentation by Allan Friedman (2012, 26) highlighted a Regulatory Spectrum (see Table 2.1) of potential action at a talk on The Economics of Cybersecurity. This document distinguished between the following:

Table 2.1: Regulatory Spectrum

| Regulation | Mandated Standards | Liability | Standards & Practices | Purchasing Power | Laissez Faire |
|---|---|---|---|---|---|

The proposed policies explored in the subsequent case studies operate across this continuum, adopting voluntary and mandated regulations. They serve to:

1. Mandate some standards (i.e., encryption of customer data)

2. Promote the adoption of open-ended policies without specifying their substance

3. While not establishing liability, these regulations open the door to civil lawsuits when the affected entity inadequately adopts government-recognized standards and best practices.

### 2.3.2 Metaregulation

Meta-Regulation is an ideal term to characterize the US-based cybersecurity regulatory framework, applicable across the four cases explored in this work. The concept can be defined as “those ways that outside regulators seek to induce regulated entities to develop their own self-regulatory response” (Coglianese and Mendelson 2012) or the “process in which the regulatory authority oversees a control or risk management system, rather than carries out regulation directly - it ‘steers, rather than rows’” (Baldwin, Cave, and Lodge 2012, 147). Its theoreticians contrast Meta-Regulation with the concept of ’Command and control’ regulation, “which refers to the prescriptive nature of the regulation (the command) supported by the imposition of some negative action by the regulator (the control)” (Simon 2016, 1). “If adequately enforced, command and control regulation is dependable; it can specify operational parameters and regulatory obligations with clarity and immediacy” (1).

> ‘**Meta-regulation**’ has been used to describe regulation for self-regulation in different ways. At its most basic, it relates to corporate self-audits and safety cases where businesses develop their own rules and reporting for the regulator to assess.
>
> -F.C. Simon, (2016, 2)

### 2.3.3 Tacit Theories of Change

Legislators and regulators implement policies based on implicit assumptions that take the form of shared theories of change. These theories of change fit within the framework of the economic security literature discussed above, even while it is highly doubtful that legislators would have read this literature. Having identified the need to create public policy to address a problem, legislators and regulators identify stakeholders, perceived wrongs, and propose remedies. In the breach space, legislators are frequently concerned about the impact on residents, while regulators are concerned about the effect on consumers. As legislators and regulators perceive resident/customer security as a positive externality, governments seek to create incentives that shift responsibility for security onto the affected company.

A form of epistemic humility constrains this desire to create incentives for companies to adopt security measures. Governments may justify their restrained action with the belief that “technological innovation outpaces the ability of laws and regulations to keep up” (Therier 2018). Legislators, perceiving the private sector as better equipped to describe these evolving best practices, place the private sector in a position of information asymmetry where corporations are well-positioned to operate as independent agents. Based on the Ambiguity/Conflict Matrix developed by Richard E. Matland (1995), cybersecurity policy can be defined as experimental implementation, low on conflict and high on ambiguity, a domain with general uncertainty in goals, technology, and tactics with a shared core value. Legislators integrate this ambiguity by creating broad technology-neutral regulations that enable the private sector to develop standards and practices more organically.

Across all four cases, regulators embed tacit theories of change in one form or another; however, the degree of restraint and epistemic humility has evolved over time and at different levels of authority. Consequently, the Massachusetts Data Security Law passed in 2007 is more constrained than the NY DFS regulations passed a decade later in 2017. In addition to growing confidence in best practices, the NY DFS regulations apply to a subset of the New York Financial sector that is advanced in its capability and a relatively homogeneous population, making it easier to govern.

### 2.3.4 Regulator and Legislative Intent

Beyond tacit theories of change, I review the explicit language used by legislators and regulators preceding the four cases. We can infer the particular intentions that characterized the cyber policy interventions within their historical context. Still, the passage with which a bill or regulation becomes adopted is not always orderly, and rarely is a singular argument put forward.

Before exploring the four cases, it is illuminating to look at the justification for California's original data breach notification regulation in 2002. Former California State Senator Joseph Simitian (2009), who drafted the legislation as part of Assembly Bill 700, characterized this process in an academic journal. Senator Simitian described how the idea for notification came two days before the legislative deadline during a small conference call. Senator Simitian had organized the call to discuss a privacy bill he was drafting. Deidre Mulligan, a lawyer with the University of California Berkeley Samuelson Clinic, proposed the idea during this call. Sen Simitian claims to have also considered notification but had previously dropped the proposal from the draft under discussion. This concurrent ideation gave him the courage to adopt the provision in the draft legislation. The bill's ultimate passage was a product of complex negotiations and instigated in part by the 2002 Teale Data Center breach, which included state employee information, including California legislators. When discussing the reasoning behind data breach notification, Senator Simitian emphasized consumer notification empowering individuals to make informed judgments about how to secure themselves. However, Senator Simitian also "hoped to provide an incentive to those responsible for public and private databases to improve their security (and thus reduce the risk for all of us)." (Simitian 2009, 1015)

The 2007 TJ Maxx data breach incident is partially responsible for instigating the Massachusetts Data Security Act. Representative Castillo, who drafted the initial legislation, linked the legislation to the national incident in interviews in the preceding months, saying, "Folks like TJX, which are multi-million dollar companies, should be seriously investing in secure systems."(Arnold and Costello 2007) However, preventing breaches was not the only purpose of the legislation. Understanding the purpose behind the law requires separating the security regulations and the breach notification requirement. In the case of breach notification, the intention was "to help Massachusetts consumers protect personal information, fight fraud" (Massachusetts Office of the Governor 2007, 181). In part, this emphasized, like with the California legislation, consumer's ability to take steps to protect themselves. In contrast, OCABR was tasked to "set regulations for how businesses and government agencies must protect consumers' information to prevent data breaches." The objective of the regulation itself is to "protect against unauthorized access to or use of such information that may result in substantial harm or inconvenience to any consumer." (Office of Consumer Affairs and Business Regulation 2017)

In a transcript of Governor Deval Patrick's comments on August 2, 2007, the day he signed the legislation, he starts by describing the problem of identity theft and projects growing and more destructive cyber criminal behavior. He contrasts this with his and the legislators' efforts to enact "critical new safeguards to help you protect your credit and your good name." (Governor Deval Patrick, 2021) He continues, "It all begins with prevention. The new identity theft law sets clear standards requiring businesses and other organizations to protect your valuable personal information" (2). This statement explicitly links data protection with identity theft.

Insights into the origin of the HITECH Act come from Representative Peter Stark, who drafted a commentary in the American Journal of Managed Care describing the process for the HITECH Acts passage (Stark 2011). Representative Stark had drafted HR 6898 - Health-e Information Technology Act was ultimately integrated by Congress with HR 6357 - Protecting Records, Optimizing Treatment, and Easing Communication through Healthcare Technology Act (Dingell 2008) and S.1693 - Wired for Health Care Quality Act (Kennedy 2007). Ultimately they were integrated into the HITECH Act under the American Recovery and Reinvestment Act (ARRA). The core focus of these earlier bills was to increase the adoption of health information technology with incentives. Privacy and security issues were seen as a core component of these incentives to increase provider and consumer confidence in new technologies.

Two weeks before Representative Obey introduced the ARRA in the House, the Senate Committee on Health, Education, Labor, and Pensions held a hearing called “Investing in Health IT: A Stimulus for A Healthier America” (Kennedy et al. 2009). The event included speakers from Kaiser Permanente, the Healthcare Leadership Council, and the Government Accountability Office that spoke to the privacy issues central to this expansion in Health IT. Their words describe the justification for congressional action:

- “We believe that HIPAA should remain the basis of new privacy rules. However, privacy policy also must cover personal health data consistently, regardless of what entity holds the records. Privacy requirements can achieve better protection for consumers without adding to the cost of HIT, changing the practice of medicine, or creating medical liability issues. There are good models in State law for guarding against security breaches [...] In our experience, California law provides a model for breach notification that is clear and consistent across all types of entities, events, and circumstances.” - Kaiser Permanente (10)
- “We recognize that, as we move towards widespread use of HIT, some aspects of

the HIPAA Privacy Rule will need to be updated to meet these emerging privacy and security concerns. For example, meaningful notification of privacy or security breaches is an important improvement necessary to protect individuals whose identifiable health information has been compromised." - Mary Grealy, President, Healthcare Leadership Council (Kennedy et al. 2009, 29)

- "As the use of electronic health information exchange increases, so does the need to protect personal health information from inappropriate disclosure. The capacity of health information exchange organizations to store and manage a large amount of electronic health information increases the risk that a breach in security could expose the personal health information of numerous individuals. Addressing and mitigating this risk is essential to encourage public acceptance of the increased use of health IT and electronic medical records." - Statement of Valerie Melvin, Director, Information Technology, The Government Accountability Office (37)

These words were echoed with appreciation by members of congress and explicitly tied to the pending Recovery Act legislation, with Senator Mikulski stating, " I want to work with our President and really then move health IT in the stimulus." (45) Senator Enzi while more skeptical, said the following "Greater adoption of health IT also presents an opportunity to increase the privacy and security of patient records [...] In some of these instances it may be necessary to take a fresh look at the current privacy and security rules, but I urge my colleagues to proceed with caution." (55) Ultimately, HR 6357 included language on breach notification and expanded HIPPA to covered entities. A notable difference between this prior legislation and the HITECH Act was the substitution of "unencrypted" for "unsecured" protected health information.

The Congressional Research Service (CRS) stated that the "HITECH Act is intended to promote the widespread adoption of health information technology (HIT) to support the electronic sharing of clinical data among hospitals, physicians, and other health care stakeholders" (Redhead 2009, 1). CRS continued, "the legislation strengthens enforcement of the Health Insurance Portability and Accountability Act (HIPAA) privacy rule and creates a right to be notified in the event of a breach of identifiable health information" (Redhead 2009, 2). In a statement on Privacy and Security by HHS and OCR, they identified more narrowly the intentionality behind the privacy and security regulatory actions associated with the HITECH Act, claiming they would "strengthen the privacy and security of health information" and that this was an "integral piece of the Administration's efforts to broaden the use of health information technology." (Blumenthal and Verdugo, 2021, 1) Concerning the notification provision, the statement identified this rule "as an incentive to the health care industry to improve privacy and security." (1)

Before the passage of the NY DFS cybersecurity regulations, the department implemented a series of three studies on the banking and insurance sectors and third-party service providers. These three studies intended to take a "holistic view" of the problem and inform "flexible" policies where examiners could work with companies to implement solutions. The studies identified "[b]olstering cyber security" as "a high priority for the Department" (*Report on Cyber Security in the Insurance Sector* 2015, 14) and sought to "foster smarter, stronger cyber security programs" (*Report on Cyber Security in the Banking Sector* 2014, 12).

Upon the initial pronouncement of the regulations, Governor Cuomo explicitly stated that these regulations should "guarantee the financial services industry upholds its obligation to protect consumers and ensure that its systems are sufficiently constructed to prevent cyber-attacks to the fullest extent possible." (*Press Release* 2016) Superintendent Maria Vullo was somewhat more circumspect, stating that "Consumers must be confident that their sensitive nonpublic information is being protected and handled appropriately" and that the effect of regulations on covered entities would "work to reduce vulnerabilities in their existing cybersecurity programs" (2016).

Themes from the NY DFS regulations consequently touch on the three themes of consumer confidence, improving organizations' cyber posture, and consumer protection through reduced breaches. The theory of change proposed by the regulation is to "establish and maintain a cybersecurity program designed to protect consumers' private data" (*Press Release* 2017).

Overall, evaluating the original intentions behind these regulations shows underlying themes. These themes include 1) a focus on consumer confidence or trust, 2) a focus on improving the cybersecurity of regulated organizations, and 3) link the regulations with consumer privacy protections. A separate justification is used for breach disclosure, which empowers the public to respond to a data breach. This research work focuses on this third theme, the assumption that new regulations would reduce data breaches. However, the mechanism this paper assumes will reduce breaches is the improved cyber maturity of regulated organizations. This more ambitious aim, not just to change corporate practice, but also to protect consumer privacy was echoed by both Governors Deval Patrick and Andrew Cuomo. Regulators appeared to be somewhat more circumspect and less ambitious in their hope for the policies.

## 2.4 Summary

While breaches have served as a subject of inquiry in the cybersecurity empirical literature, their use in cybersecurity policy evaluation has been under-explored. This chapter has identified how several data limitations resulted in this data source being overlooked,[2] and general dearth of empirical policy literature in the cybersecurity domain. Theoretical progress has been made in the economics literature, which has sought to model incentives and externalities related to cybersecurity resource allocation. US legislators and regulators have tacitly understood these findings. However, they have generally been hesitant to implement cyber security regulations. In the few cases explored in this research work, regulators opted to enact metaregulatory policies, which promote self-regulation. Chapter 7 will test the efficacy of these policies as measured by a reduction in data breaches. We may as-

[2] Methods to overcome these challenges will be shown in subsection 4.1.2.

sume this serves as an adequate measure of success because legislators and regulators have expressed as much, and the economics literature brings attention to an under-investment in cybersecurity that the incentives in these regulations could overcome. The policies explored in this work are varied in their targeted population and effect. The next chapter will explore the history and scope of these experiments in US cyber regulations.

# CHAPTER 3

# CASE SELECTION

The interventions explored in this research are the 2009 HITECH Act, the 2010 Massachusetts Data Security Standard, FTC Section 5 enforcement against Wyndham Hotels in 2012, and the NY DFS cybersecurity regulations passed in 2017. I selected these four cases because they shared a policy logic that contrasts with a traditional laissez-faire approach pursued across the United States. In all four cases, regulators sought to shape company incentives through regulatory requirements backed with sanctions for non-compliance. This multiple case study design focused on a shared policy logic that allows for a discussion of the more generalized finding, do cyber regulatory policy interventions work?

Concerning my case selection, many promising cases lacked an embedded measure of success. Congress enacted three of the nation's most significant cybersecurity laws before states started collecting data breach reports, including the 1996 HIPAA, the 1999 Gramm-Leach-Bliley Act, and the 2002 Federal Information Security Management Act (FISMA). Alternative policy actions, including the 2015 U.S.-China Cybercrime and Related Issues High Level Joint Dialogue and Cybersecurity Information Sharing Act (CISA), were also considered and ultimately rejected as neither changed the national regulatory environment applicable to US corporations.

A crucial dimension of case selection was seeking to identify quasi-experimental interventions with an identifiable control group. Without a control group, past efforts at policy evaluation have been subject to criticisms that variability in external factors could easily explain variations in pre and post-intervention effects (this relationship is explored in Figure 1.2).

To avoid this issue with internal validity, three of the four cases identify a comparable state or industry that did not experience the intervention. Consequently, these regulations are mapped in Table 3.1 onto a 2x2 targeting a sub-population based on state or industry.

Table 3.1: Case Selection Categories.

| | **State-Level** | **National-Level** |
|---|---|---|
| **Industry Level** | NY DFS cybersecurity regulation (March 1, 2017) | HITECH Act, part of the American Recovery and Reinvestment Act (ARRA) (February 17, 2009) |
| **All Industries** | Massachusetts Data Security Standard - 201 C.M.R. 17 (March 1, 2010) | FTC Section 5: Unfair or Deceptive Acts or Practices (Enforcement 2005-2020) |

What follows is an in-depth overview of the four policies, including early enforcement actions related to the regulation, relevant context, and administrative efforts at policy evaluation. While this chapter discusses these cases independently, chapter 6 compares the cases' regulatory content, implementation periods, and penalties.

## 3.1 The Massachusetts Data Security Law

The Massachusetts Data Security Standard applies to "persons who own or license personal information about a resident of the Commonwealth of Massachusetts" (*201 CMR 17.00: Standards for the Protection of Personal Information of Residents of the Commonwealth*, 2020, 1). The Office of Consumer Affairs and Business Regulation (OCABR) finalized the policy on September 22, 2008. The initial regulations would have required compliance on January 1, 2009; however, OCABR pushed back the enforcement date on three separate occasions, with enforcement formally starting on March 1, 2010. The legislatures passed this regulatory initiative as Title XV, Chapter 93 H, Section 2 of the Massachusetts Law. Wherein it instructed "[t]he department of consumer affairs and business regulation to adopt regulations ... to safeguard the personal information of residents" (*Regulations to Safeguard Personal Information of Commonwealth Residents*, 2020). The legislature adopted this language as *An Act Relative to Security Freezes and Notification of Data*

*Breaches* (2007), after which Governor Deval Patrick signed it on August 2, 2007 (Judy, Towle, and Mahoney 2007).

Chapter 93 H, Section 3(a) of the Act required that affected organizations “shall provide notice ... to the attorney general, the director of consumer affairs and business regulation and to such resident” (*An Act Relative to Security Freezes and Notification of Data Breaches* 2007). Massachusetts was the fourth state to adopt this requirement, following North Carolina in 2005 and New Hampshire and Hawaii in 2006. Consequently, Hewlett Packard Company was the first reported breach to OCABR on August 16, 2007 (*Data Breach Notification Report, 2007* 2019). As the reporting requirement proceeded the enactment of new cybersecurity regulations, OCABR reviewed notification letters for approximately 300 breaches. This review informed the drafting of the Massachusetts Data Security Standard (*The New Massachusetts Mandatory Security Regulations and Guidelines* 2008).

Both OCABR and the Attorney General receive copies of template letters from affected breaches. These differ from breach reports to the public as they need not entail the nature of the breach. Public data posted by the Attorney General identifies whether the incident involved the loss of electronic or paper records and which incidents involve lost equipment. Requests for specific notification letters are likely common, as the web page where they post data breach statistics and meta-data on reports contains language indicating that one “can request a copy of the notification via a public records request” (*Massachusetts Data Breach Notification Report*, 2020).

The regulatory requirements of 201 Code of Massachusetts Regulations (CMR) 17 mandate that covered organizations create a written information security program (WISP). They must further assign an employee to oversee this program, punish employees who violate the program, monitor access, encrypt personal information, and train employees. However, the phrase “to the extent technically feasible” limits many of the specific technical requirements (*201 CMR 17.00: Standards for the Protection of Personal Information of Residents of the Commonwealth*, 2020).

Consistent with Massachusetts Law, OCABR had to produce a small business impact statement (OCABR 2009). An OCABR memo (*The New Massachusetts Mandatory Security Regulations and Guidelines* 2008) provides a fiscal statement wherein it imagines a hypothetical 10-employee business that requires three laptops and a network server supporting seven desktops. They further assume that the company has a computer consultant available. Under these assumptions, they hypothesize an upfront cost of $3000 and a monthly cost of no more than $500 per month. The bulk of these costs are assumed to be associated with the encryption compliance requirement, which they project would be $2000 for two days of consultant assistance at $125 per hour. Given that a company already maintains this computer assistance, OCABR assumes a negligible cost for compliance.

Massachusetts General Law Title XV: Regulation of Trade, chapter 93A, section 4 (,2020) establishes limits on the penalties OCABR might assign for non-compliance. These are set at $5000 per violation and would be required to pay the investigation and litigation costs. These penalties do not take the place of a civil action, so the public is still free to file a separate suit for negligence.

OCABR quickly leveraged these new authorities settling with the Briar Group, a Boston-based restaurant group, for $110,000 on March 28, 2011. The complaint related to a breach in April 2009 instigated by hackers installing "malcode" on the Briar Group's system, which they failed to remove until December 2009. This complaint alleged that Briar Group employed:

> "default usernames and passwords on its point-of-sale computer system; allowed multiple employees to share commons usernames and passwords; failed to properly secure its remote access utilities and wireless network; and continued to accept credit and debit cards from consumers after Briar knew of the data breach" *Massachusetts Attorney General* (2011).

In this press release, Attorney General Martha Coakley committed to "continue to take action against companies that fail to implement basic security measures" (2011).

This act remained the law of the land in Massachusetts from 2007 through January 10, 2019, when Governor Baker signed HB 4806, *An Act Relative to Consumer Protection from Security Breaches* (2018). This law notably expands on reporting requirements in Section 8; however, it primarily focuses on consumers' rights concerning Consumer Reporting Agencies. A search for "data breach" and "data breaches" on the Massachusetts Legislature's website failed to identify any additional amending legislation (*Search Results for: Data Breach*, 2020). Further, the National Conference of State Legislatures (NCSL) tracks Security Breach Notification Legislation, and no new laws were recorded from 2010-2017 (see Table A.1). The only addition identified by NCSL law passed in 2018 was H.B. 5094 related to consumer credit freezes (*2018 Security Breach Legislation* 2019).

## 3.2 HITECH Act

Congress passed the HITECH Act on February 17, 2009, as part of the ARRA. While it primarily allocated billions of dollars to encourage the adoption of electronic medical records, the HITECH Act also contained several provisions intended to improve data security in the health sector (45 CFR Parts 160 and 164). On April 27, 2009, the HHS posted a request for information (RFI) for Section 13402 of the HITECH Act (*HITECH Act Breach Notification Guidance and Request for Public Comment* 2009). Promulgated by HHS, regulatory guidance was devised as "a joint effort by the Office of Civil Rights (OCR), the Office of the National Coordinator for Health Information Technology (ONC), and the Centers for Medicare and Medicaid Services (CMS)" (2009). These rules were released as interim final regulations applicable to breaches 30 days after posting, which CMS did not revise after receiving public comment.

The HITECH Act amendments to 45 CFR Parts 160 and 164 required the collection of breach data when health records for 500 or more individuals were affected,[1] and mandate that protected health information (PHI) would be "rendered unusable, unreadable, or

[1] 500 affected individuals is also the cut off for mandatory breach notification in California, Washington, Iowa, and Delaware.

indecipherable to unauthorized individuals" (*HITECH Act Breach Notification Guidance and Request for Public Comment* 2009). The latter provision was interpreted in the regulations to require encryption both at rest and in motion consistent with Federal Information Processing Standards (FIPS) 140–2. Notably, with § 164.314, the HITECH Act also extended Security Rule regulations to business associates of covered health care entities.[2] The HIPPA Security Rule was previously finalized on February 20, 2003, and required covered entities to comply by April 20, 2005 (*Federal Register* 2003, 8380). The Security Rule included administrative, physical, and technical safeguards that established an advanced security standard across the health care industry.

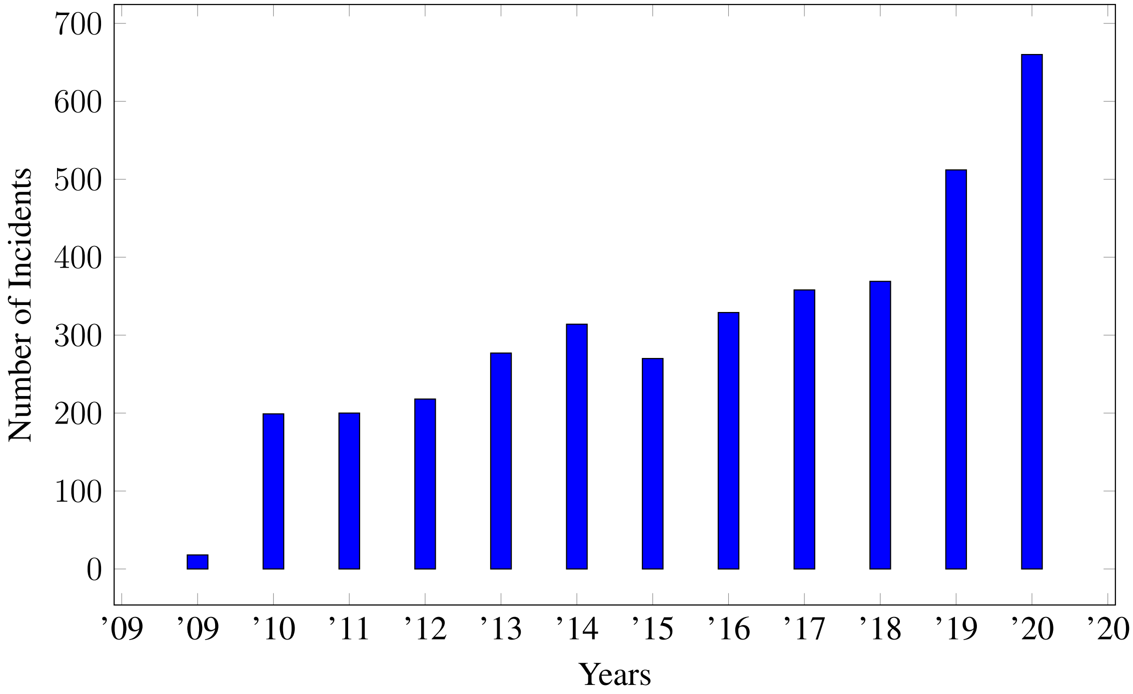
Since the passage of HIPAA, HHS has received complaints about potential violations. From April 20, 2005, through December 31, 2010, HHS reported receiving 803 complaints alleging violations of the Security Rule. These violations are dwarfed by the 57,375 complaints alleging violations of the Privacy Rule from 2003 to 2010. Unfortunately, complaints directed at violations of the HIPAA Security Rule were only categorized separately until 2010. Further, even before 2010, HHS only ever released aggregate information about these complaints, making it exceedingly difficult to measure policy efficacy based on complaint data.

However, breach incidents remain a compelling measurement of efficacy. The HITECH Act produced a national database of healthcare breach incidents. The first incident in the dataset was posted on October 21, 2009, by Brooke Army Medical Center. Bai et al. (2017) used this dataset to compare breached Hospitals with those not breached, and Liu et al. (2015) looked at predictive factors. As shown in Figure 3.1, the growth of incidents reported to HHS showed a relatively slow trajectory of incidents through 2018, followed by a relatively rapid increase in reporting more recently in 2019 and 2020.

This trend is of potential interest for future researchers seeking to understand the dis-

[2] The symbol "§" corresponds to the "section symbol" and helps with referencing the corresponding section of a document divided into sections. Consequently, "§ 160.404" refers to Section 314 of Part 164 of Title 45 in the Code of Federal Regulations.

Figure 3.1: Number of Breach Incidents Reported to HHS by Year



tribution of incidents nationally. However, it would require inferring broader population trends based on the health sector. As this state data has not previously been disaggregated, national population-based estimates have not been possible. Consequently, this data source could provide a rough approximation of the amount of breach activity in those states who do not collect breach data.

“§ 160.404 - Amount of a civil money penalty” defines both maximum and minimum amounts per violation depending on the date of commission, negligence, and correction. If a subsequent identical violation occurs within the calendar year, the regulation caps penalties at 1.5 million dollars per violation. § 160.408 further specifies what factors are considered when assessing a penalty.

Since the implementation of the regulations, additional information about the efficacy of the rules can be accessed through mandatory HIPAA Audits. The first of which was piloted and completed in December 2012. These audits were focused on the HIPAA Privacy, Security, and Breach Notification Rules. As part of the audit, HHS selects a limited number

Table 3.2: Penalties for Violation of 45 CFR Parts 160 and 164 Committed after February 18, 2009.

| Provision | Condition | Corrected | Per violation penalty |
|---|---|---|---|
| § 160.40(2)i | | | $100 to $50,000 per violation |
| § 160.40(2)ii | Reasonable cause | | $1,000 to $50,000 per violation |
| § 160.40(2)iii | Willful neglect | Corrected within 30-day | $10,000 to $50,000 per violation |
| § 160.40(2)iv | Willful neglect | Corrected within 30-day | Less than $50,000 per violation |

of covered entities for evaluation. However, the pilot exempted these selected organizations from enforcement action based on findings discovered through the audit. OCR contracted out to Booze Allen Hamilton to identify the covered entities and KPMG to design and conduct the audit protocol (Sanches, 2021). Ultimately, the 2012 audit surveying 115 covered companies. Two-thirds of the covered entities lacked a comprehensive security risk assessment (Sanches and Rinker 2013), 27% of entities in non-compliance reported: “they were unaware of the requirement” (*Annual Report to Congress on Breaches of Unsecured Protected Health Information: For Calendar Years 2011 and 2012* 2015, 25). Ultimately, OCR engaged Price Waterhouse Cooper (PWC) to evaluate the audit program (*Annual Report to Congress on HIPAA Privacy, Security, and Breach Notification Rule Compliance*, 2021, 23). OCR repeated the audit in 2016.

As for enforcement, OCR investigated 458 breach incidents that affected more than 500 individuals between the years 2011-2012. These resolutions agreements represented the first enforcement actions by OCR after investigating data breaches, with several of the incidents dating to 2009. Of these 458 investigations, OCR entered into resolution agreements with seven covered entities:

- Blue Cross Blue Shield of Tennessee ($1,500,000)
- Alaska Department of Health and Social Services ($1,700,000)

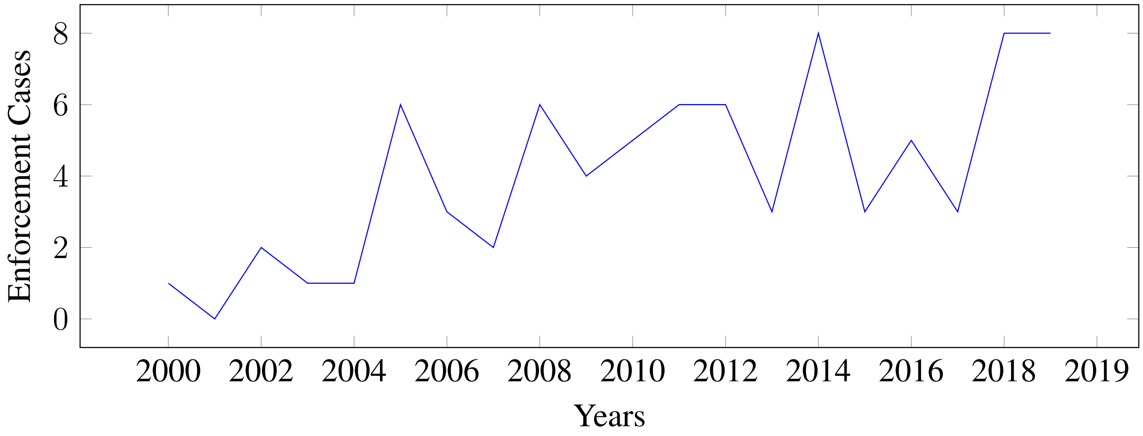
- Massachusetts Eye and Ear Infirmary and Massachusetts Eye and Ear Associates Inc ($1,500,000)
- Hospice of North Idaho ($50,000)
- Idaho State University ($400,000)
- WellPoint Inc ($1.7 million)
- Affinity Health Plan Inc ($1,215,780)

In addition to the settlement payments, shown above in parenthesis, firms also consented to corrective action plans (CAPs) that entail additional Privacy Rule and Security Rule compliance and training (*Annual Report to Congress on Breaches of Unsecured Protected Health Information: For Calendar Years 2011 and 2012* 2015, 20). Four of the seven resolution agreements related to the theft of physical devices, while Idaho State University and WellPoint experienced unauthorized access of electronic public health information (ePHI) over the Internet. OCR included additional information about these early enforcement actions in their *Annual Report to Congress on HIPAA Privacy, Security, and Breach Notification Rule Compliance*, 12-18.

## 3.3 Federal Trade Commission (FTC) Enforcement

The Federal Trade Commission (FTC) is an independent agency created in 1914 whose mission covers consumer protection and civil antitrust law. A five-member bipartisan commission oversees the FTC. Commissioners serve up to seven-year terms and are nominated by the president and confirmed by the Senate (FTC.gov 2021). While the FTC is not the only Federal Agency with regulatory authority related to data breaches, it has been the most active. In the past two decades, the FCC has made 84 enforcement actions related to "Data Security" (*Cases Tagged with Data Security*, 2020). Figure 3.2 shows the frequency of these data security actions since 2000. While cases grew in frequency from 2000-2010, they have seemingly plateaued since.

Figure 3.2: FTC Data Security Enforcement Cases per Year



FTC authority has extended to other areas of privacy and security beyond data security. This authority includes rules related to the Children's Online Privacy Protection Act (COPPA), financial regulations (including the Gramm-Leach-Bliley Act (GLBA) and the Fair Credit Reporting Act (FCRA)), and misrepresentations of compliance with Privacy Shield or the US-European Union Safe Harbor (Section 5(a), deceptive acts or practices). As they relate to consumers, data breaches are addressed within the FTC by the Bureau of Consumer Protection.

### 3.3.1 FTC Section 5 Authorities

The FTC has used its Section 5 authorities against unfair or deceptive trade practices to push companies into adopting cybersecurity measures. One of the first such incidents involving an FTC complaint regarding online privacy was against Eli Lilly in 2002. In this instance, the company, which primarily served as an online pharmacy selling Prozac, sent an email to 669 subscribers with their contact information in the 'To' portion of the email. As Eli Lilly claimed to uphold a "privacy code" that did not represent internal practices, they were accused of misrepresentation (*United States of America Federal Trade Commission In the Matter of ELI LILLY and COMPANY, a Corporation. COMPLAINT*, 2020). While this breach was inadvertent, it set a precedent as the first enforcement of data security standards. Eli Lilly agreed to create a security program through a consent agreement. This first agreement carried no penalty; however, violations of the consent agreement would entail a civil liability of $11,000 per violation (*United States of America Federal Trade Commission In the Matter of ELI LILLY and COMPANY, a Corporation. COMPLAINT*, 2020).

The FTC would subsequently employ consent decrees and enforce penalties for inadequate cybersecurity. For ten years, the FTC's authority in this space went unchallenged as companies settled rather than face costly legal battles. However, in 2012 this changed when the FTC filed a complaint against Wyndham Worldwide, a hotel chain, the first of two critical court challenges on the FTC's authority. The decision in Federal Trade Commission v. Wyndham Worldwide Corporation, et al. culminated in a settlement in the District Court of New Jersey. The settlement voided the FTC's penalties but retained a consent decree to adopt security practices (*Federal Trade Commission v. Wyndham Worldwide Corporation, et al., Stipulated Order for Injunction*, 2020; *Wyndham Settles FTC Charges It Unfairly Placed Consumers' Payment Card Information At Risk* 2015). A second court case challenging the FTC's Section 5 authority on data breach cases occurred with LabMD, whom the FTC filed a complaint against in 2013. Ultimately, the court sided with LabMD on June 6th, 2018. In 2019, the 11th Circuit ordered the FTC to pay legal fees amounting to over $843,000 to LabMD. However, as this decision was based on the vagueness of the proposed remedy, it may not limit the FTC's authority regarding data breaches. More recent consent agreements have attempted to define security program requirements with greater specificity.

### 3.3.2 Other FTC Relevant Authorities

Three additional cybersecurity regulations play a significant role in the FTC's cyber authorities. These include the Children's Online Privacy Protection Act (COPPA) (2013), *Standards for Safeguarding Customer Information* (2002), and the FTC "Red Flag Rule"

(,2020).

Since COPPA passed in 2012, the FTC has cited this law in filings against companies whose breach affected minors. COPPA regulations (16 CFR Part 312) require that "the operator must establish and maintain reasonable procedures to protect the confidentiality, security, and integrity of personal information collected from children" (*Part 312—Children's Online Privacy Protection Rule* 2013).

The *Standards for Safeguarding Customer Information* (2002), also known as the Safeguard Rules - 16 CFR Part 314, are FTC requirements first released on May 23, 2002, which apply consumer information protection rules to financial institutions. These regulations were required by section 501(b) of the GLBA (1999), which instructs agencies to promulgate rules for financial institutions "to protect against any anticipated threats or hazards to the security or integrity of such [consumer] records." As this rule proceeds breach data collection (concurrent with the year California implements breach notification), we can not test GLBA's efficacy.

The FTC "Red Flag Rule" (,2020) is a component of the FCRA. While Congress passed the FCRA in 1970, the part 681 Identity Theft Rules were created under the authority of the Fair and Accurate Credit Transactions Act of 2003 (Public Law 108–159, sec. 114; 15 U.S.C. 1681m(e).) They appear in the Code of Federal Regulations as "Detection, Prevention, and Mitigation of Identity Theft" (*Part 681—Identity Theft Rules* 2007).

### 3.3.3 FTC v. Wyndham Worldwide Corporation, et al.

One of the most meaningful data security actions by the FTC was its 2012 complaint against Wyndham Hotels. Wyndham challenged the action, believing that the FTC lacked the "authority to assert an unfairness claim in the data-security context" (*Federal Trade Commission v. Wyndham Worldwide Corporation, et al., Stipulated Order for Injunction* 2014, 2). This case ultimately confirmed the FTC's authority, expanded its scope to regulate franchisees, and with Wyndham's settlement mandated Payment Card Industry Data Security

Standard (PCI DSS) compliance and certification.

A cybersecurity incident at the Phoenix data center of Wyndham Worldwide Corporation resulted in Russian hackers accessing half a million customers' records between 2008 and 2010 (Sperry 2012). Wyndham, one of the world's largest hotel chains, operates Ramada, Super 8, and Days Inn. The FTC's complaint against Wyndham suggests that they "unreasonably and unnecessarily exposed consumers' personal data" due to a lack of "reasonable and appropriate security" (2012, 10). This lack of cybersecurity preparedness ultimately resulted in three separate breaches. A thorough discussion of these incidents is included in the Complaint for Injunctive and Other Equitable Relief (Tom et al. 2012).

- In the first incident, on April 2008, a hacker leveraged a local Wyndham-affiliated hotel's network to access the broader Hotels and Resorts network. In May of 2018, the hacker then attempted a brute force attack on an administrator account. The hacker was ultimately successful and gained administrator privileges. In the process, however, 212 user accounts were locked out. This alerted the company to a problem, but they failed to identify the breach until four months later. The hacker installed "memory-scraping malware" to collect temporary credit card data on the server and found payment card data on the company server in plain text.

- A second breach occurred in March 2009; hackers accessed the account with a "service provider's administrator account." With this new access, hackers continued using the memory-scraping malware and changed the hotel's software outputting credit card account data in clear text files. Also, in May 2009, Wyndham Hotel, based on reports of fraudulent charges by customers, scanned the network and found the memory-scraping malware installed at 30 Wyndham branded hotels.

- The final breach occurred in late 2009; the hackers were again able to gain access to an administrator account. They reinstalled the memory-scraping malware. The compromise extended to 28 hotels. Eight of which were directly managed by Wyndham

Hotel Management, while 20 were franchises of Hotel and Resort.

Ultimately, the FTC attributed these three incidents to the lack of firewalls between Wyndham Hotel servers, local networks, and the Internet. The injunction suggests that 619,000 payment accounts were stolen, leading to 10.6 million dollars in fraudulent losses. The FTC identified the actions as violating Section 5(a) of the FTC Act.

Wyndham's challenge of this complaint identified three issues to justify a motion to dismiss the unfairness claim (*Federal Trade Commission v. Wyndham Worldwide Corporation, et al., Stipulated Order for Injunction* 2014).

- The first issue the challenge identified was a lack of explicit FTC authority, analogizing the case to FDA v. Brown & Williamson Tobacco Corp., 529 U.S. 120 (2000). In this analogous case, the courts denied the Food and Drug Administration (FDA)'s authority to regulate tobacco products.

- The second issue identified by Wyndham was that the FTC had not promulgated regulations before it made its complaint. This perceived absence was identified as violating the fair notice principles.

- Finally, Wyndham claimed that the FTC did not adequately demonstrate a claim for either unfairness or deception. They suggested that the complaint was not consistent with federal pleading requirements.

The District Court of New Jersey found in favor of the FTC, rejecting the three claims by Wyndham. Wyndham subsequently appealed to the Third Circuit, Philadelphia appeals court. In August of 2015, the Third Circuit affirmed the District Court decision in favor of the FTC, setting a new precedent (*Federal Trade Commission v. Wyndham Worldwide Corporation, et Al.* 2015). In December 2015, Wyndham would finally settle with the FTC (*Federal Trade Commission v. Wyndham Worldwide Corporation, et Al.* 2015). The settlement would require the adoption of a comprehensive information security program.

### 3.4 NY Department of Financial Services Regulations

The NY DFS enacted comprehensive cybersecurity regulations on March 1, 2017. These regulations were formally promulgated as Part 500 of Title 23 of the Official Compilation of Codes, Rules and Regulations of the State of New York (23 New York Codes, Rules and Regulations (NYCRR) 500), referred to throughout this work as the "NY DFS regulations" (*In the Matter of Tuition Options LLC and Edvantage LLC Consent Order*, 2020).

The state created the NY DFS when the New York State Banking Department and the New York State Insurance Department merged on October 3, 2011, under the Financial Services Law. This change allowed the Department to "oversee a broader array of financial products and services." Appointed by the governor, a Superintendent of Financial Services leads the department. Maria T. Vullo served as Superintendent from 2016-2019. Her tenure covered the periods preceding the passage of the regulations and its implementation phase.

The provisions of the NY DFS regulations require that covered entities: write a cybersecurity policy (§ 500.03), appoint a Chief Information Security Officer (CISO) (§ 500.04), perform regular penetration testing and vulnerability assessments (§ 500.05), generate audit trails (§ 500.06), limit user access privileges (§ 500.07), create procedures for evaluating, assessing, or testing applications (§ 500.08), conduct risk assessments (§ 500.09), provide cybersecurity training (§ 500.10), create policies for third-party service providers (§ 500.11), implement multi-factor authentication (500.12), limit data retention (§ 500.13), monitor authorized use (§ 500.14), encrypt non-public information where feasible (§ 500.15), and develop an incident response plan (§ 500.16) (*Cybersecurity Requirements for Financial Services Companies*, 2020).

These policies entered into force 180 days from the effective date with annual reporting requirements to the NY DFS starting on February 15, 2018. Covered entities including Banks and Trust Companies, Cash Checkers, Credit Unions, Health Insurers, and Mortgage Bankers. Exemptions to these regulations were created for companies with fewer than ten employees, less than 5,000,000 in annual revenue, or 10,000,000 in total year-end assets (*Cybersecurity Requirements for Financial Services Companies*, 2020). However, regulated businesses must contact NY DFS to notify them that they are requesting an exemption. This requirement creates a reasonably high standard for compliance where NY DFS knows specifically who is covered and what they have done to comply.

Enforcement action penalties are set by the superintendent and limited under the New York Banking Law (Section 44 on Violations and Penalties) (*New York Banking Law, Sec. 44 Violations; Penalties*, 2020). These penalties have a maximum limit per day, including fines of $2,500, $15,000, and $75,000, depending on the recklessness, intention, and effect of the violator's action. The first enforcement action by NY DFS was against First American Title Insurance Company on July 22, 2020 (*Department of Financial Services Announces Cybersecurity Charges Against A Leading Title Insurance Provider For Exposing Millions of Documents with Consumers' Personal Information* 2020). This action was unusual for NY DFS as it was "one of only two instances we are aware of in which the DFS issued a Statement of Charges against a financial institution, rather than a Consent Order or Settlement Agreement" (Dembosky et al. 2020). On March 3, 2021, the NY DFS made another enforcement action against Residential Mortgage Services Inc. (REM) for violations of 23 NYCRR 500. REM agreed in its consent agreement to pay 1.5 Million dollars and submit a Cyber Security Incident Response Plan, Cybersecurity Risk Assessment, and evidence of Training and Monitoring to NY DFS (*New York State Department of Financial Services, In the Matter of Residential Mortgage Inc.* 2021). Another significant consent order was signed with National Securities Corporation on April 21, 2021, with an agreed payment of three million (*Press Release* 2021).

Implementation Schedule for NY DFS Regulations:

**June 2018** - NY DFS expand cybersecurity regulations to "cover consumer credit reporting agencies that reported on 1,000 or more New York consumers" (New York State Department of Financial Services 2018). The initial announcement by Governor

Cuomo was made in September 2017 and was identified as a response to the Equifax hack. This announcement entailed a "phased in schedule of compliance, starting April 4, 2018" (*Governor Cuomo Announces New Actions to Protect New Yorkers' Personal Information in Wake of Equifax Security Breach* 2017).

**March 2018** - First CISO reporting deadline to NY DFS.

**March 2019** - Requirement for multi-factor authentication comes into effect.

The NY DFS regulations have, since their initial proposal, been generally well-received. In 2017, the Ponemon Institute prepared an industry survey to measure preparedness for New York State's cybersecurity regulations. One of the survey findings was that 60% of participants thought that compliance was more challenging than GLBA, HIPAA, or Sarbanes–Oxley Act (SOX), and 65% anticipated compliance would improve their cybersecurity posture (*NYC NYDFS 23 NYCRR 500 Cybersecurity Event A Big Success* 2017). Ultimately, the NY DFS regulations informed the development of model legislation by the National Association of Insurance Commissioners in December 2018 (*Insurance Data Security Model Law* 2017). Several states subsequently adopted this model legislation, including Louisiana, Mississippi, Alabama, South Carolina, Virginia, Indiana, Michigan, Ohio, Virginia, Delaware, and New Hampshire (11 states as of June 16, 2020) (Weatherford, McAdam, and Bradstreet 2020).

### 3.4.1 Other New York Regulatory Actions

Since the implementation of NY DFS regulations, the New York legislature and NY DFS have taken several actions to improve cybersecurity in New York. Governor Cuomo signed S.5575B, the Stop Hacks and Improve Electronic Data Security Act (SHIELD Act), into law on July 25, 2019 (*NY State Senate Bill S5575B* 2019). This legislation functionally resembles the Massachusetts Data Security Law, requiring that firms implement "reasonable data security" and "provides standards tailored to the size of a business" to protect the PII of New York residents. These new SHIELD Act regulations included creating an information security program, designating employees to oversee the program, training employees to meet standards, and implementing a wide range of risk management strategies and technical safeguards. The regulation gave businesses given 240 days to come into compliance with SHIELD Act on March 21, 2020.

The SHIELD Act served as the first significant amendment to New York's 2005 Information Security Breach and Notification Act (*S05827 Summary:* 2021). This 2005 law was part of the first wave of data breach reporting requirements established in 2005. The law was ahead of its time, requiring reporting to several state agencies, including the Attorney General, the Consumer Protection Board, and the State Office of Cyber Security and Critical Infrastructure Coordination ("OCSCIC") (Reinke 2006). New York was even early to provide a form for affected parties to identify details about the breach. Yet, an initial effort at an open record request with the New York Attorney General's office produced 2,888 pages of records on November 21, 2019, covering incidents around 2019; however, these records did not date back to the NY DFS cybersecurity regulations' implementation.

In addition to the SHIELD Act, New York, starting in 2018, passed several pieces of cybersecurity legislation, beginning with S.B. 6886, which removed fees from consumers freezing their credit report following a data breach. In 2019, the SHIELD Act was passed on the same day; the governor also signed AB 2374, which required credit reporting agencies to provide identity theft prevention in the event of a breach of the agency's system. This law was a response to the 2017 Equifax breach. Governor Cuomo would ultimately settle with Equifax for 19.2 million dollars. North American Industry Classification System (NAICS) identified no other breach-related legislation from 2010-2017 (see Table A.1). In addition to these statewide changes, on May 22, 2019, NY DFS Superintendent Linda Lacewell created a new Cybersecurity Division to oversee these regulations and appointed Justin Herring (former Chief of the Cybercrimes Unit of the US Attorney's Office for the District of New Jersey) as the Executive Deputy Superintendent.

# CHAPTER 4

# METHODOLOGY

As the United States has not yet developed a comprehensive regulatory cybersecurity regime (FTC enforcements notwithstanding), these piecemeal initiatives create quasi-experimental opportunities for nonequivalent control group research using segmented regression. A nonequivalent control group design is defined as "an experimental group and a control group both given a pretest and posttest, but in which the control group and the experimental group do not have pre-experimental sampling equivalence" (Campbell and Stanley 1963, 47-50). By subdividing the regulatory regimes by affected states or industries, one can compare breach frequency trends before and after initiating the policy with a control group. As breach frequency is stochastic and variable from month to month, an improved design is the "Comparative Design Interrupted Time-Series" (55). Table 4.1 demonstrates how multiple observations (shown as Os in the table below) may be made before and after the intervention (shown as an X). This approach is considered a highly effective quasi-experimental design that addresses core internal validity issues related to selection and maturation.

Table 4.1: Comparative Design Interrupted Time-Series

| **Group** | **Pre-Test** | **Treatment** | **Post-Test** |
|---|---|---|---|
| Experimental Group | $O_1 \, O_2 \, O_3 \, O_4 \, O_5 \, O_6$ | X | $O_7 \, O_8 \, O_9 \, O_{10} \, O_{11} \, O_{12}$ |
| Control Group | $O_{13} \, O_{14} \, O_{15} \, O_{16} \, O_{17} \, O_{18}$ | | $O_{19} \, O_{20} \, O_{21} \, O_{22} \, O_{23} \, O_{24}$ |

Source: *Campbell and Stanley* (1963, 55)

A benefit to exploring regulatory interventions with quasi-experiments is that the enactment of regulation and its ultimate enforcement have clearly defined dates. These dates provide for a period of adoption, after which we would expect to see the effects of the policy in place. Linking disparate policy experiments with a shared policy logic can be integrated into a more extensive research question.

**Research Question:** *Have regulatory cyber policy interventions effectively reduced the frequency of data breach incidents, ceteris paribus?*

I selected four cases to operationalize this research question (see chapter 3) to support a claim of external validity. While quantitative rather than qualitative in the research design approach, I intended case selection to follow a replication logic. As "replication logic [in case studies] is directly analogous to that used in multiple experiments" (Yin 2017, 55). In this research, the selected cases are quasi-experiments with a shared policy logic. The cases are discussed separately in chapter 3 and chapter 5 and comparatively in chapter 6. This comparative approach resembles qualitative comparative case analysis, a methodological approach whose use in public policy was explored by Agranoff and Radin (1991).

While the literature on cybersecurity policy evaluation research is surveyed briefly in chapter 2, this research has been comparatively underdeveloped in large part because of perceived data limitations. This research is an ambitious attempt to bring to the cybersecurity domain an objective and metrics-based approach for evaluating state and national policies. This process requires overcoming a range of data and statistical concerns.

Most notably, many datasets used in the breach notification literature suffer from sampling bias, as many datasets rely on publicly reported incidents (i.e., news reports) rather than mandatory reporting. Another statistical challenge is overcoming critiques that the cyber environment has changed between time 1 and time 2. These global factors may relate to changes in technology, vulnerabilities, organizational policies, or underlying criminal enterprise. While it limits which policy interventions may be explored, the quasi-experimental comparative research design overcomes several statistical challenges. Based on this technique, one can assume that similar states or industries would be subject to similar global factors. Consequently, effort must be taken with non-equivalent control groups to justify that they are similar enough to the treatment subject to be similarly affected by global factors. Accomplishing this requires matching or thoughtful selection (i.e., bordering jurisdictions). State-level breach notification counts can not be directly compared as states have varying populations and resident requirements for notification. To achieve cross-comparable incident populations, I employed a simple per capita measure.

Average Breach Incidents per Month per Million Residents

$$= \frac{\bar{b}_{MA}}{p\bar{o}p_{statex}}$$

- $\bar{b}_{freq}$ = Simple Count of Breaches in a given Month for State X
- $pop_{state}$ = Estimate of Population of State X in Millions on a given Month

To calculate state populations in a given month, I downloaded US Census *Annual Population Estimates* (2019) and applied a simple linear model to impute the population change each month (see Table D.1). As the number of incidents is frequently smaller than a state's population, I re-scaled population values to be in the millions. Matching this state population estimate with the incident frequency, I could then calculate the number of incidents per million residents to produce a new value labeled 'incident_permil.'

Incident frequency or incident per million serves as time-series variables that can be subset for eight months before and following the policy's implementation. According to Penfold and Zhang (2013), eight intervals for pre and post-tests serve as a minimum requirement for ITS. The minimal period of eight months was used across all cases to limit conflicts with other relevant events that could present an alternative hypothesis (except for a robustness check that used 12 months for the NY DFS case).

This work seeks to generalize the finding of the four individual cases. Each case measures the effect of a form of regulation that mandates affected parties adopt information security programs. The four regulations are then compared across the scope of their coverage, contents of the regulation, dates of implementation, and the threat of sanction for non-compliance.

The methods discussed below will address data processing, methodological choices unique to each case, and the comparative interrupted time-series analysis. Finally, as part of this broader dissertation, there is a brief discussion of additional empirical data sources that could be used in future research opportunities but could not be incorporated into the quasi-experimental models used in this research.

## 4.1 Data Processing

Data collection and cleaning were a particular challenge to this effort. While the sources and data cleaning methodology are discussed at a higher level in the subsequent sections, interested readers can learn more about the coding effort and data sources in Appendix D.

### 4.1.1 Data Collection

To avoid issues with selection bias, this research effort leveraged mandatory state breach notification reports. To identify reporting states, I initially identified data from the "U.S. State Data Breach Lists" resource maintained by the International Association of Privacy Professionals (IAPP) (*U.S. State Data Breach Lists*, 2020). The IAPP keeps this list updated with links to "U.S. state agencies that publish lists of reported data breaches in their respective state." In addition to this list, I systematically surveyed state notification legislation to see which states were required to report their information to a Department of Consumer Affairs or the Attorney General. For those states that did not publicly list their data, I filed open records requests. This data collection effort far exceeded the needs of the four quasi-experiments that form the core of this research effort. However, this comprehensive approach creates resources (datasets and code) that can be leveraged for future research work that builds on this initial effort. A list of the corresponding sources collected for each state is included here:

List of States and Sources:

- **California:** https://www.oag.ca.gov/privacy/databreach/list
- **Connecticut:** Received incidents through open records request filed with the state
- **Delaware:** https://attorneygeneral.delaware.gov/fraud/cpu/securitybreachnotification/database/
- **Hawaii:** http://cca.hawaii.gov/ocp/security-breach-notices/
- **Indiana:** https://www.in.gov/attorneygeneral/2874.htm
- **Iowa:** https://www.iowaattorneygeneral.gov/for-consumers/security-breach-notifications/2018-security-breach-notifications
  https://www.iowaattorneygeneral.gov/for-consumers/security-breach-notifications/2017
  https://www.iowaattorneygeneral.gov/for-consumers/security-breach-notifications/2016-security-breach-notifications
- **Maine:** https://www.maine.gov/ag/docs/Data-Breach-Spreadsheet.xlsx
  https://www.maine.gov/ag/docs/Maine_Attorney_General_Reporting_Form_Data.xlsx
- **Maryland:** http://www.marylandattorneygeneral.gov/Pages/IdentityTheft/breachnotices.aspx
- **Massachusetts:** https://www.mass.gov/lists/data-breach-notification-reports
- **Montana:** https://dojmt.gov/consumer/databreach/
- **New Hampshire:** https://www.doj.nh.gov/consumer/security-breaches/a.htm
- **New Jersey:** https://www.doj.nh.gov/consumer/security-breaches/a.htm
- **North Carolina:** https://iapp.org/media/pdf/resource_center/North_Carolina_State_Data_Breaches.pdf
- **North Dakota:** https://iapp.org/media/pdf/resource_center/North_Dakota_Data_Breaches_2018.pdf
- **Oregon:** https://justice.oregon.gov/consumer/DataBreach/
- **Rhode Island:** Received incidents through open records request filed with the state
- **South Carolina:** Received incidents through open records request filed with the state
- **Vermont:** https://ago.vermont.gov/data-security-breaches/
- **Virginia:** Received incidents through open records request filed with the state, records from 2012-2018 were collected by IAPP and posted at https://iapp.org/resources/article/u-s-state-data-breach-lists/

- **Washington:** https://www.atg.wa.gov/data-breach-notifications
- **Wisconsin:** https://datcp.wi.gov/Pages/Programs_Services/DataBreaches.aspx

I compiled state-level breach data from a variety of file formats. For breaches collected in Portable Document Format, i.e., PDFs (Indiana, Massachusetts, North Carolina, North Dakota, Virginia), these files were opened using Adobe Acrobat Pro software and exported into Microsoft Excel (a spreadsheet software application). For most cases thought, I manually extracted the data from state websites (California, Delaware, Hawaii, Iowa, Maryland, Montana, New Hampshire, Oregon, Vermont, Washington, Wisconsin). In the future, it would be helpful to have web scraping scripts to automate this work.

### 4.1.2 Data Cleaning

The initial process of data cleaning could, in many cases, not be automated and required significant manual labor in the form of data entry and manipulation. Where needed, I used Excel for manual data entry and manipulation. In only two cases (Maine and Connecticut) was the data already made available in a usable machine-readable format. To the extent possible, data cleaning code was written in R, "a language and environment for statistical computing" (R Core Team 2020). Ultimately, I saved state breach data in a comma separate value (.csv) format that closely approximates the original data file.

As states collect different meta-data from breached companies, the initial column titles were kept in the raw data. The first order of data cleaning was to remove excess information from organization names. Columns with multiple dates listing the 'date of breach' and 'end of breach' were split. Columns were given common variable names. Values that had been assigned "yes" or "no" were replaced with "1" and "0," respectively. A column was also created naming the state of collection for every row. Data sets were then merged with the function `rbind`.

All breach data contained a date corresponding to its collection by a state agency. Dates were converted into a consistent format using the R package `lubridate` (Grolemund and

Wickham 2011). In 557 cases, the collected data could not be consistently formatted and these incidents were dropped. All incidents could then be sorted by their state collection date.

Name cleaning was then applied manually across 32,753 unique incident names to create common names for incidents referenced by multiple states. A new column with the clean name was added to the data frame. The dataset could then be sorted alphabetically by ‘org name’ and then by ‘date of breach.’ New variables were created identifying a date 14-days before and 14-days after the reporting date for every incident. These pre and post-dates were used to create an interval for the specified time range. A for loop then identified incidents as a ‘truematch’ for rows where the org name matches the previous row, and reporting dates occur within 28-days of each other. Unique incidents were then numbered based on an algorithm (Listing Line 4.1.2) ordered first by earliest date and then alphabetically.

Listing 4.1: Code Used to Number Matching Incidents

```
j = 0
for (i in 1:nrow(AllState1)){
  if(AllState1$truematch[i] == ”TRUE”){AllState1$incident_id[i] = j}
  if(AllState1$truematch[i] == ”FALSE”){AllState1$incident_id[i] = j+1}
  j = AllState1$incident_id[i]
  }
```

The first step to transform the vertical to a horizontal data structure was to use the dcast function of the `rshape` package (Wickham 2007). This function created new columns for incidents’ corresponding states, designated with a dummy variable. Incident numbering subsequently provided the basis for combining these rows using the aggregate function to create an incident database combining like columns.

Listing 4.2: Code Used to Combine Incidents

```
AllStateClean <- aggregate(BreachID[,NamesDDPLY],
    by=BreachID['incident_id'],
    function(x) c(paste(unique(x[!is.na(x)]),
    collapse=”, ”))
)
```

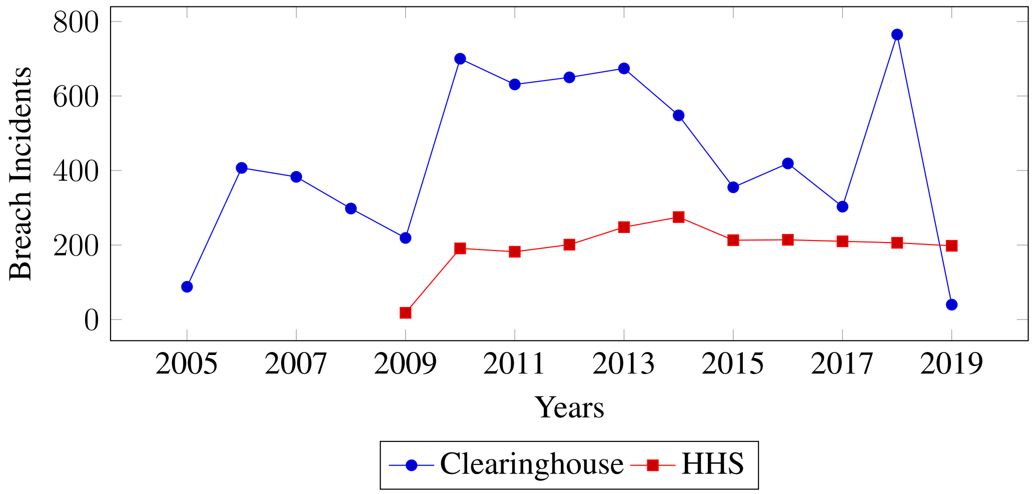
This script converted a dataset of state-reported incidents into a database of 19590 unique incidents, with state reference columns. Merged incidents were then sorted by the earliest reporting date. Consequently, all these disparate state data sets were merged into one organized file for analysis. For additional information about how to access the code used for data cleaning, see Cleaning_RawBreachData.Rmd in Table D.1.

Table 4.2: Descriptive Statistics of Captured Incidents

| **Statistics** | **Measure** |
|---|---|
| Number of captured incidents | 54,338 |
| Incidents dropped for No Reported Date | 559 |
| Incidents dropped for Amended Submission | 45 |
| Incidents dropped for Unclear Org Name | 14 |
| Incidents remaining after drops | 53,720 |
| Breaches after incident matching | 19,590 |

Additional state breach incidents can be incorporated from other public data sources, such as those maintained by the Clearinghouse and HHS (see AllSourceBreachData.Rmd in Table D.1). Incorporating additional data sources requires a duplication of a number of the steps described above. To maintain the ability to limit the dataset to mandatory reporting, events were coded with a separate reference variable, and a new variable was created to indicate when incidents were reported by one of the collecting states. While useful for its metadata, the frequency of incidents in Figure 4.1 suggests the problems with existing data sources. Clearinghouse shows significant yearly variability in reporting as it uses media reporting as a source. In contrast, HHS shows fairly flat yearly reporting suggesting limited sectoral developments over time.

Figure 4.1: Data Breach Incidents by Year in Public Datasets



## 4.2 Methodological Approach for each Case

Each of the four cases requires unique methodological decisions to design an appropriate quasi-experiment. These cases may contain either one or several tests of efficacy through the application of quasi-experimental methods.

This research employs a comparative ITS design for three cases: the Massachusetts Data Security Standard, the HITECH Act, and the NY DFS cybersecurity regulations. The FTC Wyndham actions lack a control population but are evaluated with the interrupted time-series research design. Each of these regulatory interventions is explored with two to three variants. What follows are descriptions of the methodological decisions made for each case.

Table 4.3: Case Analysis and Methodology

| | Massachusetts Data Security Law | HITECH Act | New York DFS Regulations | FTC v. Wyndham Hotel |
|---|---|---|---|---|
| Comparison population | Matching State | Matching Industry | Financial Industry Outside New York | NA |
| Pre-Test Date | Sept 2008 | Feb 2009 | March 2017 | NA |
| Post-Test Date | March 2010 | May 2009 | Sept 2018 | NA |

### 4.2.1 Case 1: Massachusetts Data Security Law

**Hypothesis 1A:** *The Massachusetts Data Security Law reduced the growth rate of data breach reports for companies doing business in the state of Massachusetts.*

To test the effect of the Massachusetts Data Security Standard, I compare the effect on Massachusetts with that on New Hampshire and North Carolina. North Carolina started collecting data in 2005, Massachusetts and New Hampshire in 2007. While the impact of the Massachusetts Standard technically has national implications for business, I identify the subset of incidents not reported in both states to confirm that a subset of incidents is unique to each state. Those companies with breaches that did not report customers from the state of Massachusetts would not need to comply with Massachusetts law. Consequently, we might reasonably suspect that their behavior would remain independent of this legislation.

#### New Hampshire

New Hampshire's legislation resembles Massachusetts in that it covers incidents affecting even a single resident. However, the proximity between Massachusetts and New Hampshire means New Hampshire-based companies might reasonably consider working to comply with Massachusetts law.

#### North Carolina

North Carolina's data breach notification law only required notification of breaches affecting 1000+ residents through 2009. Because the Massachusetts Data Security Law was implemented on September 22, 2008, comparing the two states requires using a count of only those Massachusetts incidents where 1000+ residents were reported affected.

Of potential concern with this specific case is the risk that the treatment might impact the control group. However, given that we have evidence of which breach incidents affected both states, we can infer that a sizable sample of enterprises is non-overlapping, thereby indicating that they either a) do not do business with residents in the other state or b) intentionally or unknowingly are flouting reporting requirements.

### 4.2.2 Case 2: HITECH Act

**Hypothesis 1B:** *The HITECH Act reduced the growth rate of data breach reports in the healthcare sector.* To test the effect of the HITECH Act, I coded the industry data of affected firms during the period before and after implementation. For industry coding, I used the database Mergent Intellect by FTSE Russell to export industry identifiers, including Data Universal Numbering System (DUNS) and NAICS codes for all collected incidents from 2005-2010 with an org name that could be matched to a record in the database. The Healthcare sector (NAICS 62) is then compared with all non-healthcare sector industries as well as a close sector comparison, Finance (NAICS 52).

As the HITECH Act was national legislation affecting a specific industry, the appropriate treatment group would be the healthcare sector across collecting states. This quasi-experiment consequently used the combined incidents reported by six states (North Carolina, New Hampshire, Hawaii, Massachusetts, South Carolina, and Maine). While these states have different reporting standards, these standards did not change during the quasi-experimental period. The implementation phase for the HITECH Act was fairly short as the policy was enacted on February 17, 2009, and came into enforcement fairly quickly on May 27, 2009.

### 4.2.3 Case 3: NY DFS Regulations

**Hypothesis 1D:** *The NY Department of Financial Services regulations have reduced the growth rate of data breaches in the New York Financial Sector.*
NY DFS regulations implemented in 2017 were the most recent regulatory intervention I looked at, presenting the opportunity to leverage more extensive data collection and public reporting by states. Unfortunately, I was still unable to draw on data directly from New

York state. An open records request resulted in the receipt of a PDF document several thousands of pages long that contained individual letters mailed to the state by breached entities. As these documents did not cover the targeted quasi-experimental period for NY DFS cybersecurity regulations, I did not code these documents. Instead, I identified that two reporting states (Maine and Connecticut) included information on the location of firms' headquarters that allowed me to identify New York businesses.

Companies regulated by the New York Department of Financial Services and the NAICS finance and insurance sector are not a perfect match. Nevertheless, this research design assumes a sufficient overlap, such that NAICS code 52 serves as an adequate proxy for the 1,800 insurance companies and 1,500 banking and other financial service corporations regulated by NY DFS. I could not identify a list of the firms regulated under NY DFS authority; otherwise, this would have served as an even better designation than industry.

To designate industry, I used two different approaches for data reporting from Maine and Connecticut. Connecticut's data included an industry designation that the filer could complete. I reviewed these self-designated industry identifiers and assigned an appropriate NAICS code using the 'NAICS & SIC Identification Tools' service at www.NAICS.com/search. In contrast, Maine's data only added an industry identifier starting in late 2018. Consequently, I used the database Mergent Intellect by FTSE Russell, which includes access to the D&B Million Dollar Directory (MDDI) and allows for the generation of DUNS and NAICS codes.

The NY DFS regulations, unlike the Massachusetts Data Security Standard, apply to a specific subset of in-state companies. They are not applying regulations that would require compliance by out-of-state companies. This significantly reduces the risk of the treatment affecting the control population.

### 4.2.4 Case 4: FTC Section 5

**Hypothesis 1C:** *FTC Section 5 enforcement authorization expansion through Wyndham Hotels reduces the likelihood of national breaches.* As FTC Section 5 regulations pre-date the data source, rather than judging a law or regulation for the FTC case, I focus on a significant expansion of FTC enforcement authority with the 2012 Wyndham Hotel lawsuit. The FTC Complaint against Wyndham Hotels contains two essential interventions. The first is the publicity around the initial complaint, which received significant media attention, including a New York Times article (Wyatt 2012). A second intervention was the decision by the Third Circuit to affirm the regulatory authority of the FTC, which had been put in doubt by the multi-year case instigated by Wyndham Hotel's challenge to the complaint. Both interventions presented a significant expansion of cybersecurity liability from franchisors to franchisees that would have ramifications across the economy and government recognition of PCI DSS certification. This impact is consequently measured with two ITS quasi-experiments. This analysis used the combined incidents reported by seven states (North Carolina, New Hampshire, Hawaii, Massachusetts, South Carolina, Maine, and Iowa).

## 4.3 Comparative Interrupted Time-Series

Regression equation:

$$y = \alpha + \beta_1 T + \beta_2 X + \beta_3 XT + \beta_4 Z + \beta_5 ZT + \beta_6 ZX + \beta_7 ZXT + \epsilon$$

Where:

Z = treatment as 1 or control as 0,

ZT = time for treatment and 0 for control,

ZX = study phase for treatment and 0 for control, ZXT = time after interruption for treatment and 0 for control

The above formula and labels were directly pulled from Caswell (2019, 3). While Caswell (2018) describes a SAS macro to produce comparative ITS, my research was completed using R. Consequently, I created R code to produce an identical data frame to that used as part of the proposed regression and imputed the "Rate" values with summative month incident counts (sometimes scaled for population).

The data frame required for the regression is made by appending the treatment and the control. An example data frame used for the comparative ITS analysis for the Massachusetts Data Security Law is included in the Appendix as Table B.1. With this data frame, I ran an ordinary least square (OLS) regression that produced estimates, standard errors, and probabilities for each of the eight parameters. The eight parameters measure different aspects of the time-series lines shown in subsequent figures.

- $\beta_1$: Control Pre-Trend
- $\beta_2$: Control Post-Level Change
- $\beta_3$: Control Post-Trend Change
- $\beta_4$: Treatment/Control Pre-Level Difference
- $\beta_5$: Treatment/Control Pre-Trend Difference
- $\beta_6$: Treatment/Control Post-Level Difference
- $\beta_7$: Treatment/Control Change in Slope Difference Pre- to Post-

## 4.4 Other Variables for Future Research

The following material was collected in the process of seeking to identify prospective supplementary analyses that would leverage additional control variables. Ultimately I did not pursue this approach as I could not find a way to integrate these variables into the quasi-experimental design of this research. Time-series regression models would ideally be paired with control values measured every month for the corresponding treatment and control groups.

### 4.4.1 Data Breach Letters

Despite not being integrated into this research, I began coding metadata from template letters collected from the 17 states where I compiled notification letters. These states include California, Delaware, Florida, Hawaii, Idaho, Iowa, Maryland, Missouri, Montana, Nebraska, New Hampshire, New Mexico, New York, North Dakota, Vermont, Washington. To manage these individual documents, I used the software Tropy developed by the Center for History and New Media at George Mason University. The Tropy software is designed for archives to manage letters and images and is appropriate to a project that requires organizing letters. A novel metadata template (i.e., ontology) was created to make the Tropy tool more suitable for tracking data breach notification letters. This ontology creates rich textual properties such as ‘date of breach,’ ‘residents affected,’ ‘hack type,’ etc., that can be filled out for each letter.

This coding project would ultimately serve several purposes 1) it populates new state data collected from open records requests 2) these letters contain descriptions of the nature of the breach that could be useful if attempting to identify hacking-related incidents 3) recording notification date and discovery date would improve cross-state incident matching. However, based on this initial experience, 2-3 research assistants would likely be needed to make any meaningful progress.

### 4.4.2 Reports of Identity Theft

While not essential to this research, I submitted an FOI to the FTC for monthly state reports of identity theft. This data was also collected by Romanosky et al. (2011) for the years 2002-2009 through an FOI request. The FTC has continued to collect this data and publish it annually in the Consumer Sentinel Network Data Book (2013), in a summarized form.

I ultimately received a response to my request with data from 2015-2020. Unfortunately, this did not go back far enough to use it as a robustness check against any of the quasi-experiments explored in this work. However, it could help to confirm the findings through an alternative dependent variable. The diagram in chapter 1 modeled how identity theft is downstream of data breaches. In theory, one could control for reported data breaches to see if there is a tangible effect independent of increased breach discovery and reporting. This research is a project for future work.

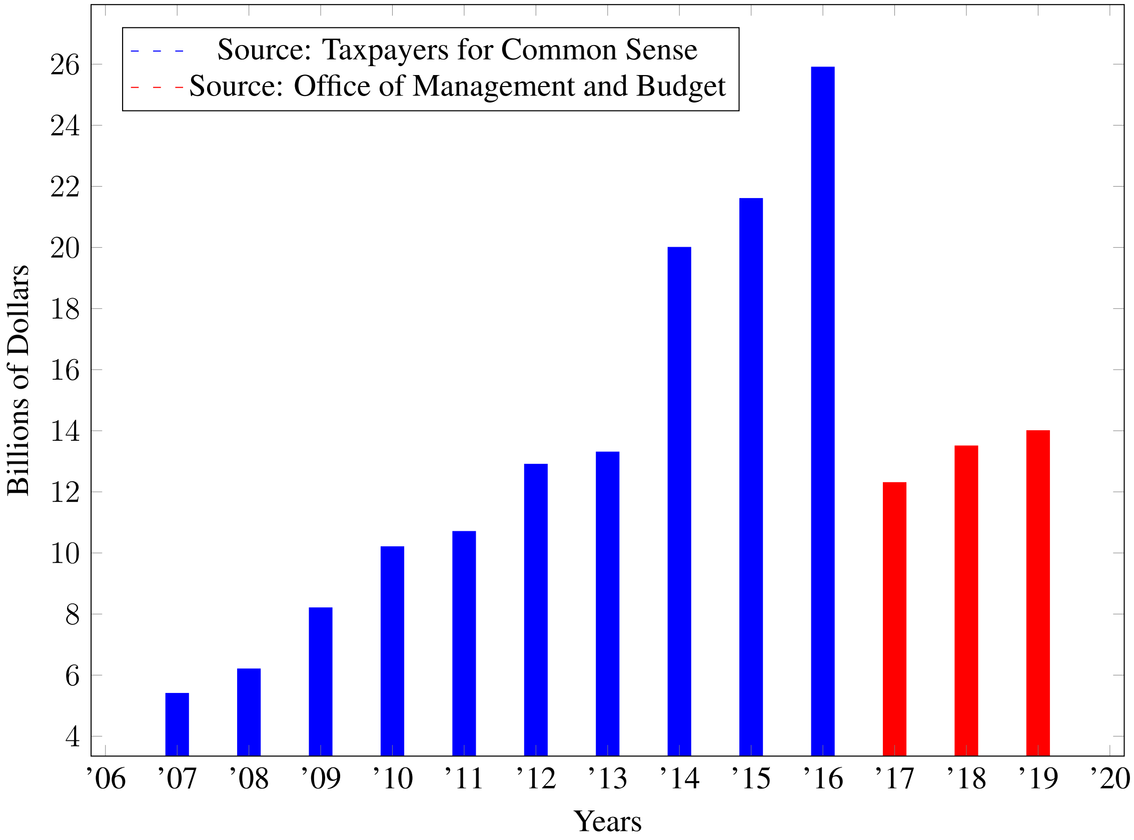
### 4.4.3 Federal Spending on Cybersecurity

As my research project focused on nationwide legislation, one of the factors I saw as potentially relevant to this effort was a measure of government capacity. I sought out data on government spending as it related to cybersecurity initiatives and compiled the available data. Chief Financial Officers Act (CFO) Agency[1] spending was tracked by the organization Taxpayers for Common Sense (“Taxpayers”) between 2008 and 2016. Funded by the Hewlett Foundation, this initiative claims to have “searched publicly available federal budget submissions to Congress and budget justification documents to identify programs” Taxpayers then “analyzed those documents to identify individual budget lines that contain programs that the government acknowledges relate to cyber spending” (*Database of Unclassified Cyber Spending*, 2020).

After this initiative had completed its work, the White House, through the Office of

[1] Under the Chief Financial Officers Act, 24 federal departments and agencies are defined as CFO Agencies. With an appointment of a Chief Financial Officer, they have additional budgetary reporting requirements

Figure 4.2: Cybersecurity Spending by the Federal Government in Billions of Dollars



Management and Budget (OMB), began releasing their own CFO Agency spending estimates (*Efficient, Effective, Accountable An American Budget, Fiscal Year 2019*, 2020; *A Budget for A Better America, Fiscal Year 2020*, 2020). Taxpayers and the government applied very different standards that make these two sources not cross-comparable, as demonstrated in Figure 4.2. However, one can identify from this data collection effort a linear growth rate of 2.2 additional billion dollars per year up to 2016 and 850 million dollars per year from 2017 to 2019.

### 4.4.4 Cyber Vulnerabilities and Hygiene

Two potential future sources of vulnerability could serve as global factors for the proposed models, National Vulnerability Database (NVD) Vulnerability Severity (,n.d.) or Cyber-Green (2020). One prospective way to use the NVD Vulnerability Severity data as a metric for broader cyber vulnerability would be to identify a count of newly posted vulnerabilities in the NVD within a specific window of time (defined as a standard patch window). This count would then be weighted for the severity of the vulnerabilities.

Of potential interest for control variables would be to control for global cyber hygiene. Initially sponsored by Japan's Computer Emergency Response Team (CERT), the CyberGreen (2020) project was developed by Dan Geer to measure how national cyber hygiene creates the potential for exploitation by adversaries. The index measures a count of commonly misconfigured protocols by Autonomous System Number (ASN) and weighs Distributed Denial of Service (DDoS) potential as the risk that a nation's internet infrastructure will be leveraged to host attacks on third-party countries. I pulled this data to explore its potential, summarizing it below by year and risk type to measure the scope.[2]

Table 4.4: CyberGreen Risk Types by Year.

| | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** |
|---|---:|---:|---:|---:|---:|---:|
| Risk Type 1 | 94711 | 1293147 | 1678042 | 1188379 | 2329170 | 3029327 |
| Risk Type 2 | 1192755 | 1322560 | 1290507 | 1074492 | 2231511 | 2298575 |
| Risk Type 4 | 199838 | 882042 | 794122 | 736749 | 1118179 | 1389725 |
| Risk Type 5 | 0 | 0 | 494565 | 378711 | 452614 | 441294 |
| Risk Type 6 | 0 | 0 | 102595 | 248608 | 0 | 0 |
| Risk Type 7 | 0 | 0 | 0 | 0 | 67487 | 73416 |

[2] Oddly, the CyberGreen's Metrics document does not specify the meaning of the different risk types. To leverage this data appropriately would require additional reference materials or a discussion with the developers.

# CHAPTER 5
# FINDINGS

## 5.1 Overview

Across four targeted cases, only the NY DFS regulations showed a statistically significant decrease in the level of breaches after a policy intervention. This post-treatment difference in breaches held up when replicated for a second collecting state. I observed this trend comparing breaches from the New York financial sector with breaches affecting the finance sector elsewhere. However, this finding was not consistent when comparing breach counts for the New York financial sector to all other New York sectors, as New York state saw a generalized decrease in breaches during the relevant dates. I did not find a significant drop in data breach incident reporting following the regulatory intervention in the other three cases. This general lack of post-treatment level significance could result from unspecified confounding effects or the interventions' weakness. One explanation for this difference could be that the requirements embedded in the NY DFS regulations are more extensive than the requirements in the other three cases. Also, the penalties for non-compliance appeared to be steeper (and were denominated by day rather than by violation).

Consequently, there are substantive reasons to assume that the NY DFS intervention may have been more effective than the other three cases. However, the NY DFS intervention may suffer from a regression towards the mean, as the New York financial sector reported significantly more incidents in 2020.

As part of this research project, some descriptive findings drawing on a decade of breach data seem to be meaningful. These findings are addressed in section 5.2.

- Across all states, there is a slow but persistent rate of growth in breach incidents at approximately 20% per year.

- The number of individuals affected by a breach has an exponential distribution with a $\lambda$ of 1.471129e-09.

Spatial and temporal findings also present themselves as insightful. The method for deriving these findings is included in section 5.3, and their implications are discussed in section 6.1.

- There is a similar number of breaches per-capita in states with similar reporting requirements.
- The consistent seasonal variation observed in data breach reporting increased in the Spring and decreased in the Fall.

An analysis of the NY DFS case will follow in section 5.4. The remaining three regulatory cases without a significant intervention are then presented in chronological order. Analysis for all four cases will include an overview of case-specific data methods, followed by trends in the underlying time-series data, and a discussion of regressions results. Finally, there is a brief discussion of statistical tests. The implications of these findings is discussed in greater depth in chapter 6.

## 5.2 Descriptive Statistics

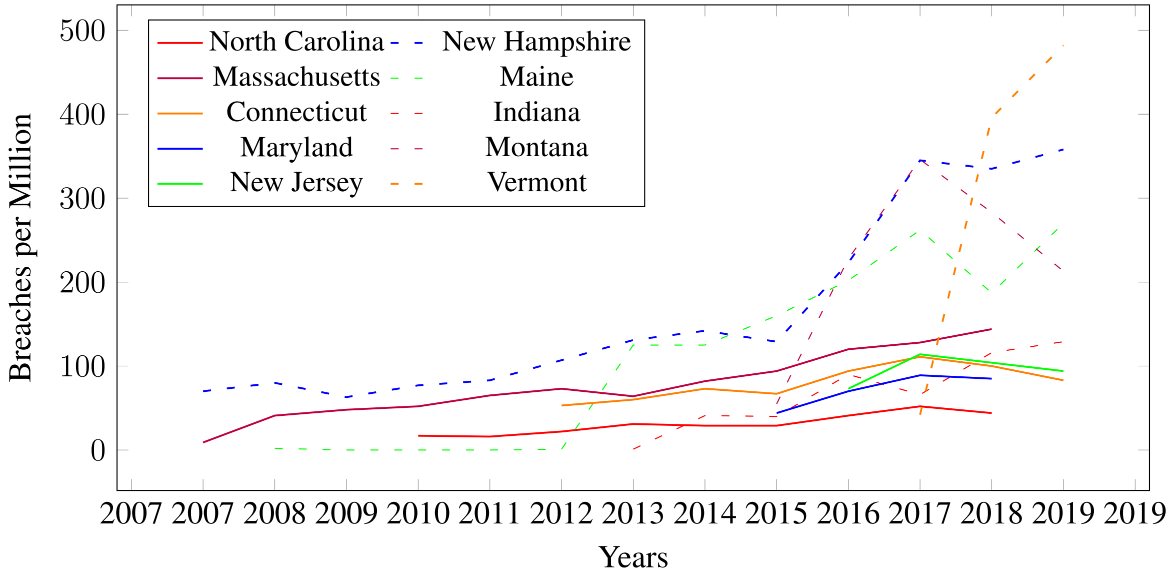
Drawing on the state-level data described in the prior data collection section, descriptive statistics provide valuable findings regarding the magnitude, character, and direction of data breaches over the past decade. There were a total of 54,340 reports of data breaches captured from 21 reporting states. After data cleaning and incident matching, this left 19,592 overall breach incidents. Table 5.1 describes the procedures for dropping data. The methods for name matching and merging state-level data sets are explored in more detail in chapter 4.

Table 5.1: Descriptive Statistics of Captured Incidents

| **Statistics** | **Measure** |
|---|---|
| Number of captured incidents | 54,340 |
| Incidents dropped for No Reported Date | 559 |
| Incidents dropped for Amended Submission | 45 |
| Incidents dropped for Unclear Org Name | 14 |
| Incidents remaining after drops | 53,722 |
| Breaches after incident matching | 19,592 |

Graphing the yearly frequencies of state breach reports while normalizing for state population reveals a close relationship between state incident frequencies over time. Figure 5.1 shows frequencies for states without reporting thresholds. This figure demonstrates that for states without a residency limit, the frequency of breach reports clusters around 100 breaches per million state residents per year in 2018. Some smaller states reported significantly more incidents, including New Hampshire, Vermont, Maine, and Montana. Based on this figure, we may assume that smaller population states show a higher number of breach reports per million residents systematically and not merely a product of more significant variance.

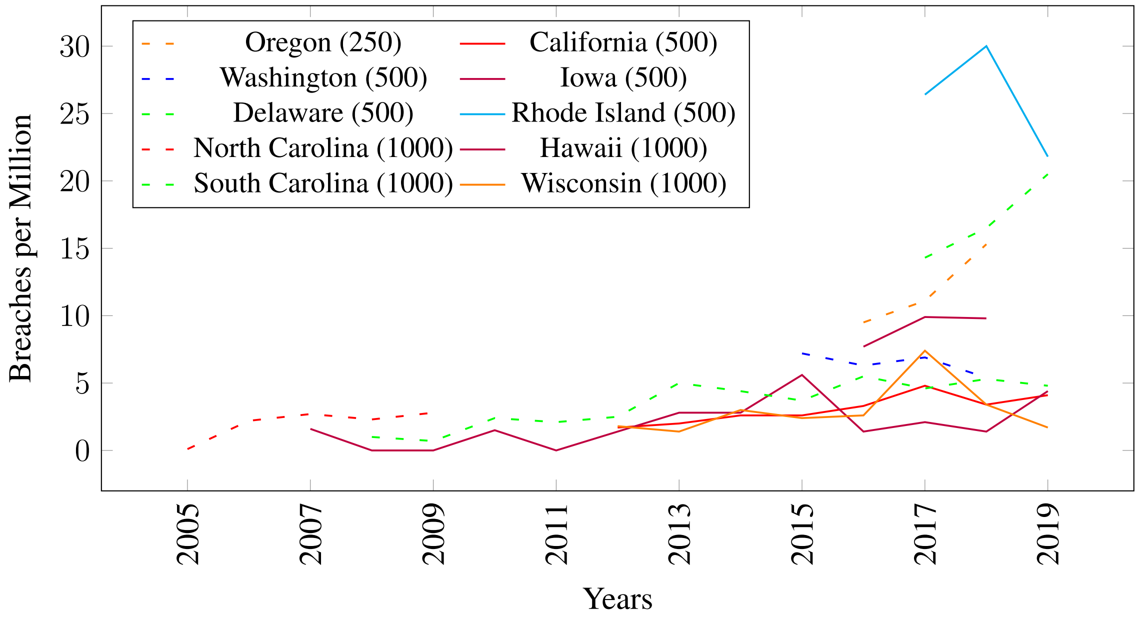
Figure 5.1: Breaches per Million with no Resident Limits



One can also look at those states that only require reporting if more than 250, 500 or

1000 residents are affected. As seen in Figure 5.2, the frequency of breaches per year for those states that require reporting when 1000+ residents are affected cluster around five incidents per million residents per year in 2019. States with more frequent breach reports per million residents per year include Oregon, Delaware, and Rhode Island. Notably, their reporting requirements are triggered when a smaller number of residents are affected. Oregon's breach reporting requirement has a residency requirement of 250 people, while Delaware and Rhode Island have a residency requirement of 500. Both Washington and California seem to cluster with those states that have a higher reporting requirement. State population size plays a role in reported breach frequency per million, with Washington and California showing a lower number than expected.

Figure 5.2: Breaches per Million with Resident Limits

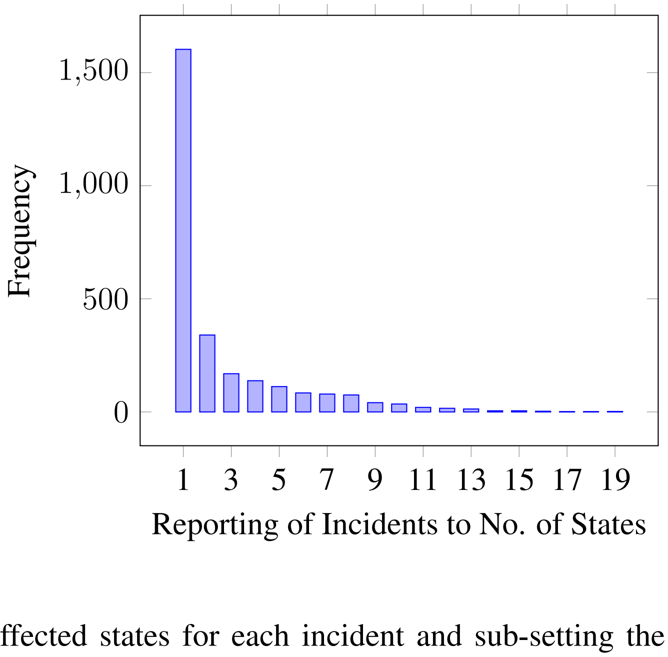


The trendline across states is also consistent once we account for rapid growth between the initial and second year of adoption and ignore the last collection year. This modified trendline avoids comparing a full year of data with a partial year of data. Calculating the Average Growth Rate for these full years for all of the states in Figure 5.1 and Figure 5.2 shows a fairly consistent growth rate. Averaging the Average Growth Rate across all states and years yields a 20.4% growth rate with a standard error of 3.5%. Alternatively, if we first calculate the mean Average Growth Rate for each state and then average these state values, it produces a 21.1% growth rate with a standard error of 2.9%. The magnitude of this linear growth trajectory may reflect some general underlying growth trends. For example, US population growth between 2006 and 2016 falls below 1%. During this period, the growth rate for adult Internet users was approximately 2.4% per year, and the growth rate for home broadband use was 7.7% per year. While these growth trends would reduce the magnitude of a 20% growth rate, we would still see consistent double-digit growth for reported breach incidents. In recent years, this consistent trend saw a significant upswing in the growth rate from 2016 (34.4%) and 2017 (45.5%), followed by a plateau in 2018 and 2019 as incident frequency shrunk by 2%.

## 5.3 Spatial and Temporal Factors

While media reporting centers national breaches, the bulk of breach reporting appears to be localized. Figure 5.3 demonstrates this tendency for local reporting using 2018 data.

Figure 5.3: Number of Incidents Reported Across Different States in 2018



Summing the affected states for each incident and sub-setting the data to 2018 (the year with data collected from the most states) showed that 1603 of 2742 incidents or 59%

of incidents in that year were only registered in one state (see Figure 5.3). The incidents reported to the most states in 2018 were Delta Airlines and Task Rabbit, both of which submitted reports in 19 of the 21 collecting states.

To create the dataset for this analysis, I used the data cleaning methods described in subsection 4.1.2. This method entailed combining state-level reports by merging incidents reported by the same organization within a 28-day window of each other. This combination was necessary as sometimes organizations would choose to provide duplicative submissions updating state agencies in rapid succession. State-level reports were thus integrated into national incidents (i.e., converting the data from long-form to wide-form).

This integration process is imperfect, as manual data cleaning may have missed associated incidents if variable organizational names were cited. Also, a 28-day window, the maximum range between incidents for matching, was insufficient to capture all breach reports linked to a given national incident. Even with spotty national coverage, Figure 5.3 provides preliminary evidence that breach reporting frequently demonstrates a local effect.

We can also process the data to assess regional similarities in reporting. Table 5.2 shows the frequency of co-occurrence of incident reports across five national regions for the 2,742 incidents collected in 2018. The frequency of the diagonal shows the total regional incident account. New England has the highest frequency of reported incidents and shows the highest co-occurrence with the other four. Variance in regional magnitude is explained by the particular circumstances of collecting states and the presence or absence of a reporting requirement.

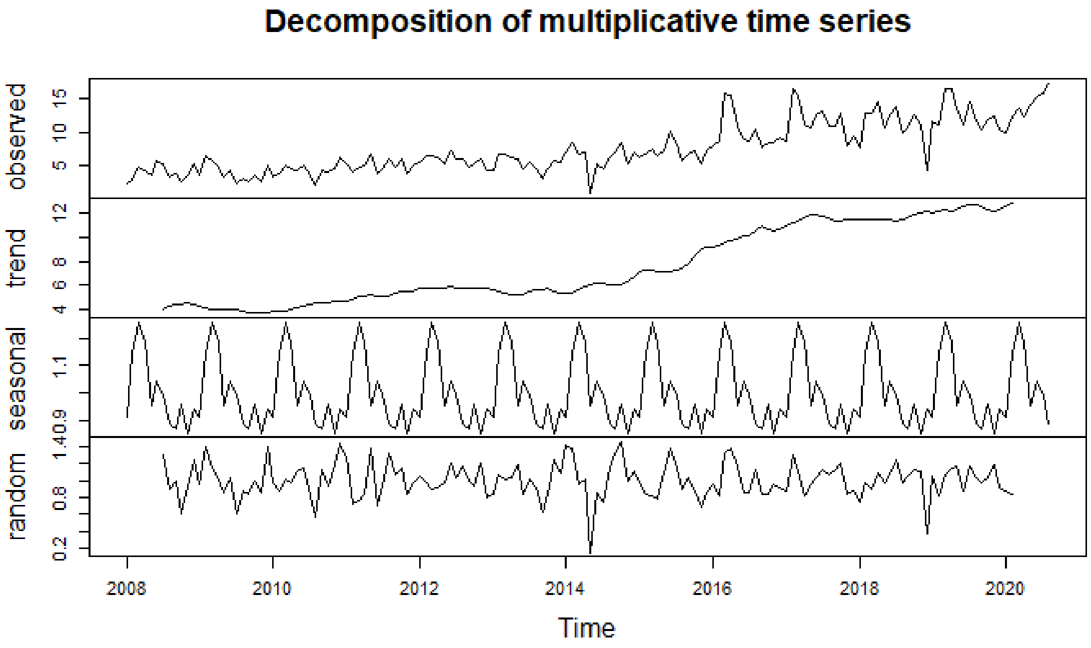
Table 5.2: Frequency of Reported Breaches Across Regions

| | Midatlantic | Midwest | New England | South | West |
|---|---:|---:|---:|---:|---:|
| Midatlantic | 879 | 420 | 591 | 561 | 329 |
| Midwest | 420 | 820 | 469 | 446 | 287 |
| New England | 591 | 469 | 1643 | 657 | 389 |
| South | 561 | 446 | 657 | 1142 | 315 |
| West | 329 | 287 | 389 | 315 | 599 |

### 5.3.1 Seasonal Decomposition

I decomposed monthly time-series data for several states to identify trend, seasonality, and randomness components. The `decompose` function is a default function in the R programming language. An example decomposition of monthly data from Massachusetts between 2008 to 2020 is included in Figure 5.4. Observing the underlying trend from Figure 5.4 shows slow but relatively consistent growth between 2008 and 2016. Growth rates appear to accelerate over the next year and a half. From mid-2017 to 2020, there seems to be a plateau in the number of incidents.

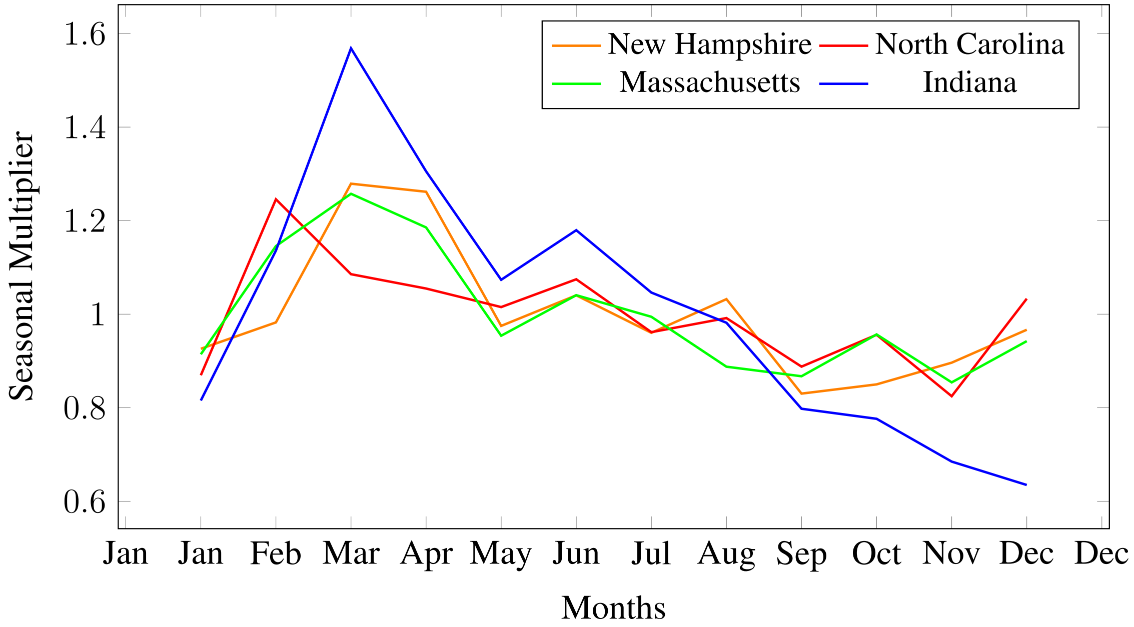
Figure 5.4: Sample Decomposition of Additive Time-Series for Massachusetts



Looking just at the season trend, we can segment the data to just a year and compare it with additional states. I chose the states of Massachusetts, New Hampshire, North Carolina, and Indiana to graph this seasonal variation in Figure 5.5 as they contained five or more years of data and did not have reporting thresholds.

As incident frequency and variance seem to increase with time, I assumed the decomposition would be multiplicative, which means that as the incident frequency grows over

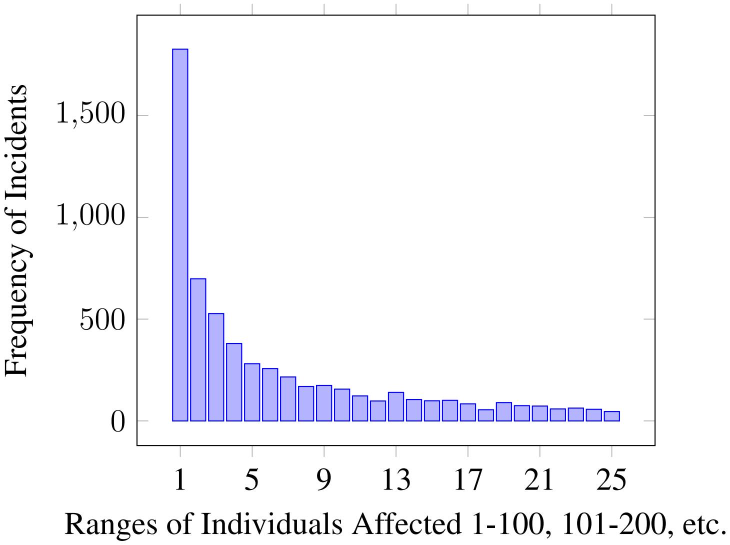
Figure 5.5: Seasonal Variation in Breaches per Million



time, seasonal variation would as well (i.e., Time-series = Trend * Seasonal * Randomness). Using a seasonal multiplier (rather than an additive decomposition) also creates a better scale for comparing various states.

Examining seasonal trends from sampled states (Figure 5.5) shows that the time-series increases in the Spring and decrease in the Fall. A potential hypothesis to explain this behavior might be a seasonal increase leading up to April 15th Tax Day or the delayed effect of reporting for incidents during the holiday shopping season but reported months later. The seasonal trends of Indiana reflect those of the other states but do seem somewhat magnified. The decomposition of Indiana only covered six years compared to the 12 available years for Massachusetts. This difference in the length of the time-series may partially explain the more significant seasonal variation. The other three states appear to cluster with an approximately 20% increase in breaches reported in March and an approximately 15% decrease in November and December. The impact of these findings is briefly explored in section 6.1.

Figure 5.6: Histogram of Total Individuals Affected



### 5.3.2 Size of Incidents

The most minor incidents recorded in the data affected a single user (1825 incidents). The three most significant reported incidents were: Marriott with 500 million affected in 2018, Equifax with 143 million affected in 2017, and Anthem with 80 million affected in 2015. Initially, the largest incidents in the dataset included Deli Management in 2018 and a combined submission for 'Cvs, allgenheny [*sic*], Johnson and Johnson, Microsoft and Pizza Hut' in 2015. The first case presents a user-submitted error where an incident affecting 2 million was listed as 2 billion, while the second combined case could not be verified. These two changes were fixed; however, submission errors of this type in the reports present an ongoing challenge to using these data sources.

Sub-setting the data before matching to look at the total number of individuals affected provides records on 10,067 incidents, slightly less than a fifth of the total. The variable 'total affected' indicates those affected regardless of residence and is collected and reported by Indiana, North Carolina, North Dakota, and Wisconsin. Maine began collecting 'total affected' in 2018. Figure 5.6 shows a histogram of the number of breaches affecting bin sizes of 100 (i.e., the frequency of incidents that affected 1-100 people, 101-200 people, etc.). From this chart, it becomes clear that there is a diminishing frequency of breaches that subsequently affect larger bins of 100. The median incident size from this sample was 490 individuals affected (approximately the first five bars). The trend line seems to follow a power-law distribution; however, subsequent testing shows that it is better described as an exponential distribution.

Statistical goodness of fit tests were run according to Wiley (2016), who recommends applying the Kolmogorov Smirnoff (KS) test, comparing the empirical findings to a large number of synthetic distributions. Alternative distributions were tested, including a power law, exponential with the xmin, exponential without the xmin, log normal with the xmin, and log normal without the xmin. Of these, the exponential functions without an xmin performed the best. The lambda of the exponential function was 1.471129e-09 showing the best goodness of fit with 2498 of 2500 KS tests failing to reject the null or 99.92% of the total.

## 5.4 NY Department of Financial Services Regulations

When the NY DFS regulations were announced on March 1, 2017, Richard Clarke stated: "With this regulation, DFS is leading the nation in promulgating strong minimum standards to protect regulated entities and the consumers they serve." Governor Cuomo described the regulations as "strong, first-in-the-nation protections." The 23 NYCRR Part 500 cybersecurity regulation is highly targeted and extensive in its compliance requirements. The regulation provided 180 days for firms to come into compliance, as discussed in section 3.4, and gave them through the year to submit a certification of compliance. The evidence shows a statistically significant finding for the efficacy of this intervention. The NY DFS regulation is shown to reduce the number of breaches compared to the financial industry located outside the state. This finding is validated by this research when it was replicated using data from a second state.

While I filed an open record request with the state of New York on October 2019, the documents I ultimately relieved were copies of letters and did not go far enough back in time to cover the pre-test or implementation phase of the NY DFS regulations. Consequently, for these quasi-experiments, I relied on records collected from the state of Maine and Connecticut.

### 5.4.1 Comparing New York Finance to Other Industries with Maine Data

I initially looked to the state of Maine, which has since December 2018 incorporated a highly detailed "Electronic Maine Security Breach Reporting Form" that contains such useful information as the state of the reporting entity and primary industry identifiers for finance, health care, education, and government. The state of Maine requires reporting if a single resident has been affected. However, as the cybersecurity regulation went into effect on February 15, 2018, I had to rely on earlier reporting to the state that did not designate whether the incident was a hack or identify the industry of the affected entity. As with the HITECH Act case, I employ the database Mergent Intellect by FTSE Russell, which includes access to the D&B MDDI that allows for the generation of DUNS and NAICS codes. In this case, industry codes were exclusively used to identify NAICS code 52, which corresponds to the Finance and Insurance sector. Having coded incidents during the reporting period, I ultimately was able to identify two-digit NAICS codes for 582 cases of the 1233 incidents reported in Maine during the eight months before and after the regulatory implementation period; this accounts for 47% of the incidents. While 75.9% of the incidents from this period were identified with their DUNS numbers, not all incidents were exportable by Mergent Intellect with corresponding NAICS numbers.

The time-series data in Figure 5.7 uses a dotted line to represent the implementation period, and the solid lines show the eight months pre and post this implementation period. This convention with dotted and solid lines is used throughout the chapter. In the case of the treatment, New York financial service firms did not report incidents to Maine during this period except for a single incident reported in March 2018 by the private wealth

Figure 5.7: Comparing New York Finance to New York Not-Finance

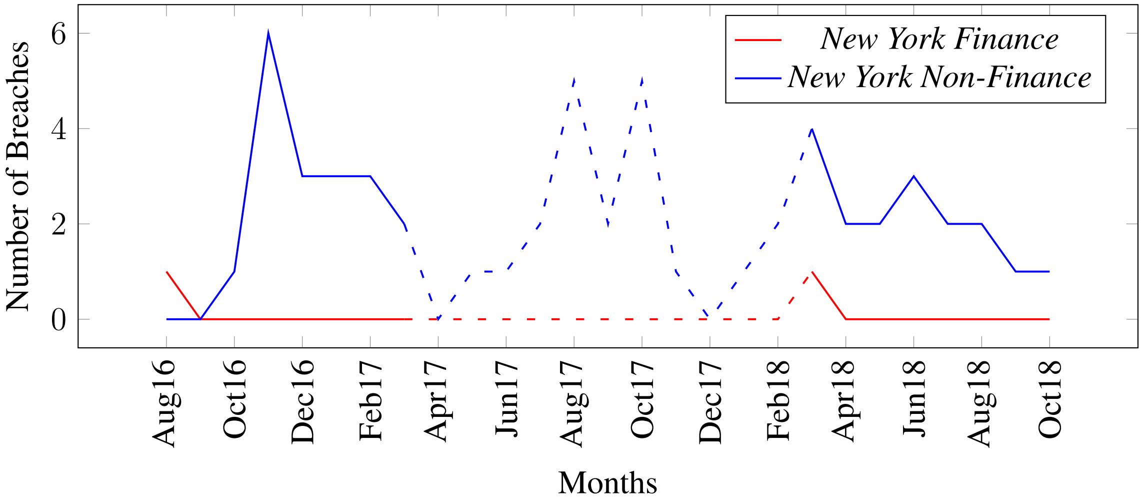


Table 5.3: Comparative ITS NY DFS Regulations (NY Finance Compared to NY Not-Finance)

| **Parameter** | **Interpretation** | **Estimate** | **Std Error** | **Probability** |
|---|---|---|---|---|
| $\alpha$ | Intercept | 0.54 | 0.79 | 0.50 |
| $\beta_1$ | Control Pre-Trend | 0.38 | 0.16 | 0.02 * |
| $\beta_2$ | Control Post-Level Change | –0.01 | 1.02 | 0.99 |
| $\beta_3$ | Control Post-Trend Change | –0.70 | 0.22 | 0.00 ** |
| $\beta_4$ | Treatment/Control Pre-Level Difference | –0.54 | 1.11 | 0.63 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | –0.38 | 0.22 | 0.10 . |
| $\beta_6$ | Treatment/Control Post-Level Difference | 0.51 | 1.45 | 0.73 |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | 0.62 | 0.31 | 0.06 . |

management firm Clarfeld Financial Advisors. The effect of zero incidents over multiple months is a flat trend line for the treatment. Whether this edge case is meaningful relates at least in part to whether we would expect New York financial incidents to report in Maine. Having comprehensive data from 2020 with industry self-identification showed ten unique financial data breaches reported in that year.

Meanwhile, all other data breaches originating from the state of New York, the control populations, showed a positive and statistically significant pre-intervention trend $\beta_1$, as given in Table 5.3. This trend changes following the intervention with a negative slope

$\beta_3$. There is also a modest 90% statistically significant change in slope produced by this time-series. During the time period of the quasi-experiment, the only data breach security legislation passed in the state of New York related to consumer credit freezes, which would not be expected to affect corporate data breach reporting. These regression results do not provide evidence for the efficacy of the intervention as $\beta_6$, which indicates a post-treatment level change is not statistically significant.

### 5.4.2 Comparing New York Finance to Finance Elsewhere with Maine Data

The more useful of the two quasi-experiments compares New York finance to all other finance incidents. Looking at Figure 5.8, we see that financial incidents went up significantly from late 2016 to early 2018.[1]

As a result of this precipitous increase in the number of reported financial incidents nationally and the comparative flatness of reports from the state of New York, there is a statistically significant level change between the treatment and control populations shown in Table 5.4. The model estimates this difference to be 3.55 incidents per month post-intervention compared to 1.11 incidents per month pre-intervention. This level change in the model $\beta_6$ works in conjunction with $\beta_2$, which shows statistical significance (0.001 probability) that the level of the control rises by four incidents per month.

### 5.4.3 Comparing New York Finance to Finance Elsewhere with Connecticut Data

There are some concerns with using Maine data to estimate New York regulatory efficacy. Of particular concern is that the lack of reported incidents might indicate that Maine residents lack close customer relationships with major New York financial firms. Ultimately,

[1] Given its significant role in finance, what proportion of financial breaches might we expect New York to occupy? If proportional to labor, and employment can serve as a proxy, the New York Department of Labor (,2021) estimated 510.4 thousand New Yorkers were employed in Finance and Insurance (NAICS 52) in contrast, the United States Bureau of Labor estimates 6.28 Million were employed nationally in January of 2018 (*Employment by Major Industry Sector* 2020). So at a minimum, one might expect 8% of incidents to be reported out of New York, which would have entailed two incidents in 2017. This, of course, ignores other important proxies like geographic proximity and the unique role played by the New York financial sector.

Figure 5.8: Comparing New York Finance to Not-New York Finance with Maine Data

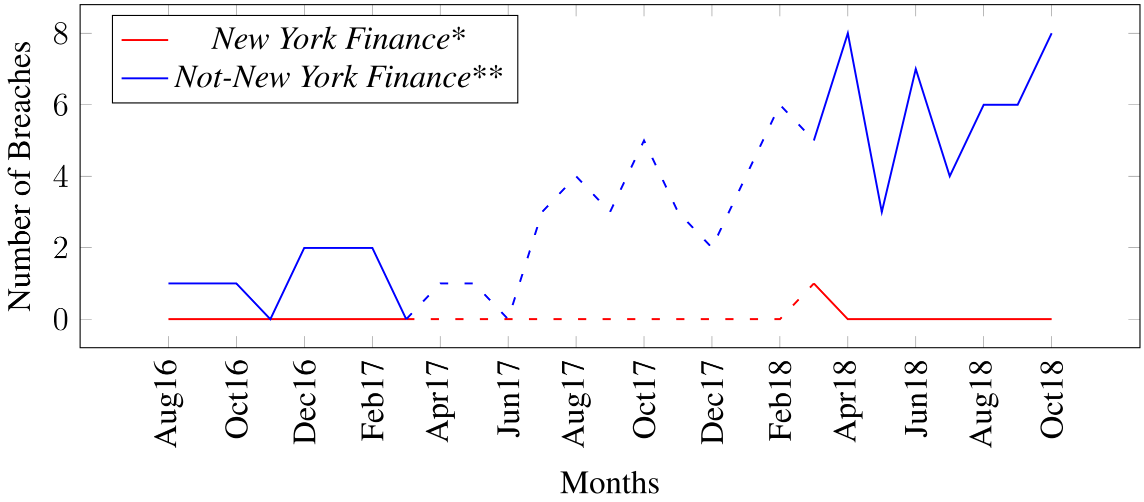


Table 5.4: Comparative ITS NY DFS Regulations (NY Finance Compared to Non-NY Finance) with Maine Data

| Parameter | Interpretation | Estimate | Std Error | Probability |
|---|---|---|---|---|
| $\alpha$ | Intercept | 1.11 | 0.81 | 0.18 |
| $\beta_1$ | Control Pre-Trend | −0.02 | 0.16 | 0.88 |
| $\beta_2$ | Control Post-Level Change | 4.05 | 1.05 | 0.00 *** |
| $\beta_3$ | Control Post-Trend Change | 0.23 | 0.23 | 0.34 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | −1.11 | 1.14 | 0.34 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | 0.02 | 0.23 | 0.92 |
| $\beta_6$ | Treatment/Control Post-Level Difference | −3.55 | 1.48 | 0.02 * |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | −0.31 | 0.31 | 0.34 |

Maine was selected because it included detailed address information for most companies affected, which allowed me to positively verify whether companies were headquartered in New York, and their associated industry. Connecticut is the only other reporting state that provided information on the location of the affected company with a category for 'State' and requests a description of the industry of the affected party. Of the 4369 incidents reported to Connecticut in the period between 2016 and 2019, 2708 include an industry identifier (61% of reported incidents). Consequently, rather than classifying the industry based on a query of the firm's name, I used the self-designated industry to search for a corresponding NAICS number. This coding exercise identified 872 of the 914 unique names or 95.4% of them. Ultimately, this left me with a sample of 2,050 industry collected incidents from the state of Connecticut.

Figure 5.9: Comparing New York Finance to Not-New York Finance with Connecticut Data

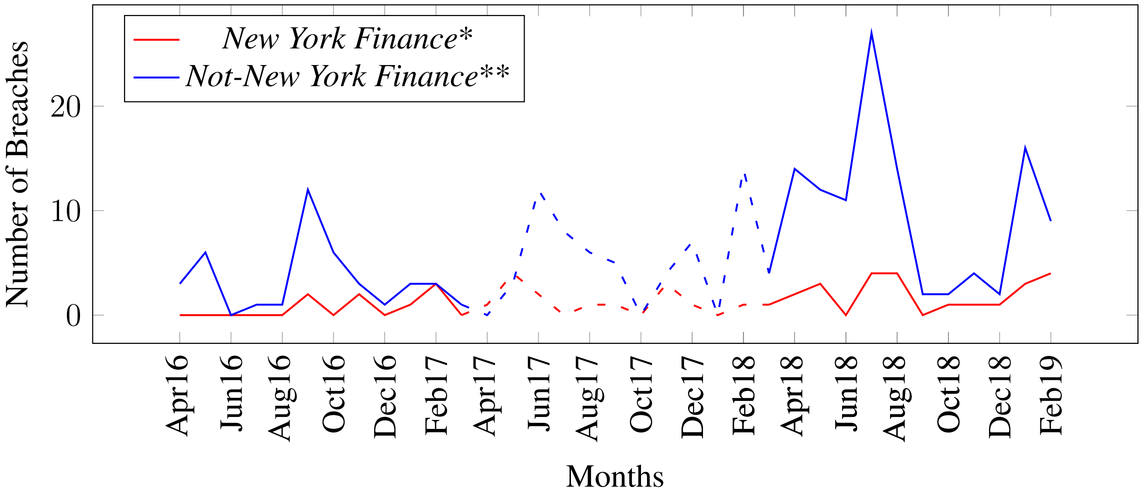


Table 5.5: Comparative ITS NY DFS Regulations (NY Finance Compared to NY Not-Finance with Connecticut Data)

| Parameter | Interpretation | Estimate | Std Error | Probability |
|---|---|---|---|---|
| $\alpha$ | Intercept | 3.97 | 2.68 | 0.15 |
| $\beta_1$ | Control Pre-Trend | –0.10 | 0.36 | 0.79 |
| $\beta_2$ | Control Post-Level Change | 9.66 | 3.57 | 0.01 * |
| $\beta_3$ | Control Post-Trend Change | –0.32 | 0.51 | 0.54 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | –4.17 | 3.79 | 0.28 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | 0.23 | 0.51 | 0.66 |
| $\beta_6$ | Treatment/Control Post-Level Difference | –9.51 | 5.05 | 0.07 . |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | 0.26 | 0.73 | 0.73 |

Figure 5.9 shows a similar trend to that observed in Figure 5.8. New York financial incidents remain relatively flat with a modest increase between the pre and post-intervention levels. In contrast, with the control, national financial incidents spiked in July of 2018, rising on average by approximately ten incidents per month from pre-treatment levels. As

there were no corresponding restrictions, and to increase the statistical power of the model, the Connecticut analysis used 12 months pre and post-treatment to evaluate efficacy. Looking at the regression results in Table 5.5 shows statistical significance for both $\beta_2$ and $\beta_6$, as was observed previously in Table 5.4. However, the strength of this statistical significance is weakened, as $\beta_2$, the post-level change of control, shows only a .05 probability. While, $\beta_6$, the post-level difference between treatment and control is weakly significant with .10 probability. Further the magnitude of this change is even more significant than in the prior model. The estimate of $\beta_6$ shows a difference of 9.51 incidents per month post-treatment. This level difference between $\beta_6$ and $\beta 4$ of 5.34 incidents per month is the effect of the regulation. The magnitude of this difference would suggest that even a smaller effect size would be a notable outcome for the regulation.

Unlike with the prior Maine quasi-experiment, the use of Connecticut data avoids the issues associated with a lack of geographical proximity or an insufficient number of cases pre and post-intervention to characterize the policy's effect adequately. While weakened, the statistical significance of this second finding still serves to validate this prior finding. The implications of the finding is discussed in greater depth in the next chapter.

### 5.4.4 Robustness Checks

There are three variants of robustness checks that were attempted on the Connecticut data. These included 1) switching the incident dating to the earliest reported date (most frequently the date of the breach), 2) changing the binning frequency from months to weeks, and 3) changing the period of pre and post-collection from 12 months to eight months. The first of these three robustness checks confirmed the initial finding, while the other two did not replicate the initial finding. This mixed replication suggests a sensitivity in the analysis that may or may not relate to the objective findings. Instead, the mixed performance of the robustness checks could be a product of the changes in measurement not being appropriate. An alternative explanation for the change in binning could be a product of too few observable incidents. When the data is limited to just reporting from one state and just within one sector, binning by weeks produces more variable data, and consequently, more observations show zero incidents. As demonstrated by Figure 5.9, the time-series shows a drop in national financial incidents observed in September thru December 2018 and a spike in January and February of 2019. When an eight-month rather than 12 month period is observed, the latter spike is excluded, and the significance of the model is lost.

Figure 5.10: Comparing New York Finance to Not-New York Finance with Connecticut Data (First Date of Breach)

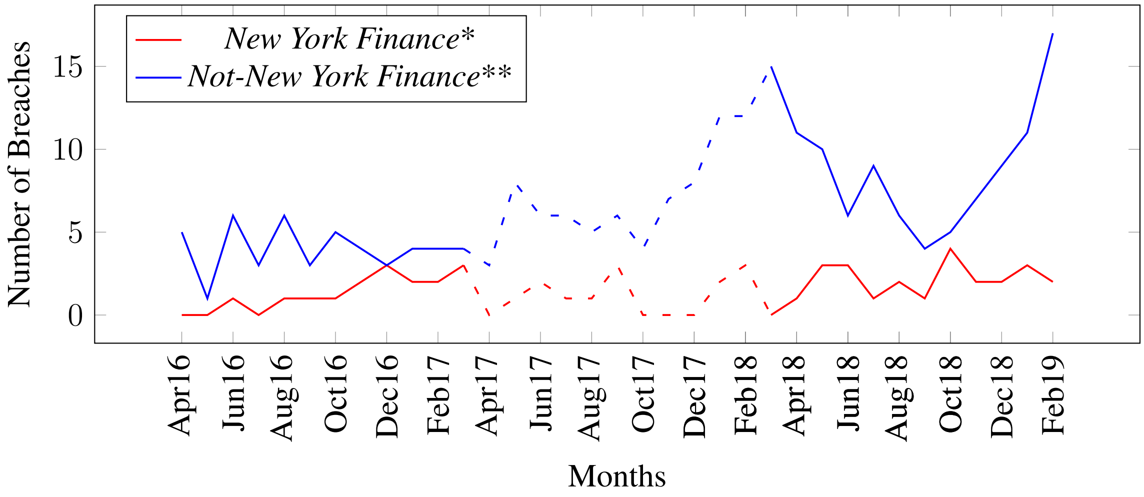


Table 5.6: Comparative ITS NY DFS Regulations (NY Finance Compared to NY Not-Finance, Connecticut Robustness Check)

| Parameter | Interpretation | Estimate | Std Error | Probability |
|---|---|---|---|---|
| $\alpha$ | Intercept | 4.05 | 1.40 | 0.01 ** |
| $\beta_1$ | Control Pre-Trend | 0.00 | 0.19 | 0.97 |
| $\beta_2$ | Control Post-Level Change | 5.07 | 1.87 | 0.01 ** |
| $\beta_3$ | Control Post-Trend Change | 0.03 | 0.27 | 0.92 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | −4.44 | 1.98 | 0.03 * |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | 0.27 | 0.27 | 0.31 |
| $\beta_6$ | Treatment/Control Post-Level Difference | −6.68 | 2.65 | 0.02 * |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | −0.16 | 0.38 | 0.66 |

However, the initial statistical findings in Connecticut were improved on by selecting the first reported date associated with each incident. For approximately 80% of the Connecticut data, this earliest date would be the initial date of the breach. If, as modeled earlier, we assume that the efficacy of the regulations would be to reduce the initial occurrence of a breach, this date would serve as a more effective proxy for the security of covered organizations. Unfortunately, the date of the breach is not required for submission by Connecticut and is rarely collected by other states. Additional dates preceding the reported date include the submission date (Connecticut distinguishes between when a breach is submitted into their system and the date the report was submitted to the Office) and the date the breach ended. By switching to the first available date, the finding is replicated for the year before and after implementation. These new findings shown in Table 5.6 demonstrate meaningful statistical significance of the post-treatment level change at the 95% confidence level, however a weaker estimate of $\beta_6$ at -6.68. If we subtract $\beta_4$ the pre-level difference from $\beta_6$ the post-level difference, this provides an effect size of 2.24 incidents per month.

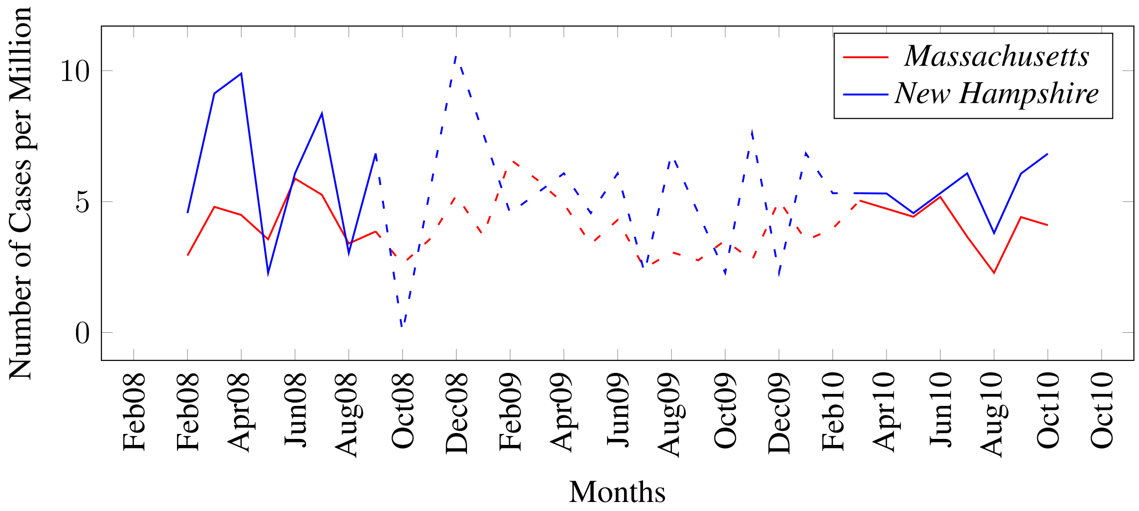
## 5.5 Massachusetts Data Security Law

In this second case, we assess the efficacy of the Massachusetts Data Security Law, promulgated as "201 CMR 17: Standards for the protection of personal information of residents of the Commonwealth." With a November 2009 implementation date, the Massachusetts Data Security Law was the earliest state-level cybersecurity regulation to be implemented.There are only two promising quasi-experimental state comparisons with this early implementation: New Hampshire and North Carolina. New Hampshire required reporting all incidents, regardless of residents affected, in the eight months prior and post to 201 CMR 17's implementation period. North Carolina, in contrast, only required reporting of incidents where 1000 residents were affected at the beginning of the implementation period and subsequently changed their data breach security law to require more extensive reporting.

### 5.5.1 Comparison with New Hampshire

Having normalized the number of reported breaches per million for Massachusetts and New Hampshire, we can compare the eight months before implementation (April - Sept 2008) with the eight months after implementation (March - August 2010). From Figure 5.11, we can see a remarkable amount of overlap in the incident frequency between the two states.

Figure 5.11: Comparing Massachusetts and New Hampshire Breaches



Comparing the incident frequency of Massachusetts and New Hampshire, the two states seem to track with each other before, during, and after implementing the regulations (this is the period between the regulations enactment and their enforcement). Visual observation seems to suggest a slight drop in the number of breach incidents during this period. While $\beta_1$ shows a negative coefficient, this is not statistically significant. Neither are any of the other parameters specified in Table 5.7. Breaches per million residents between the two states average to approximately five incidents per month. The mean of the incidents during the selected period for Massachusetts is 4.48, while the mean for New Hampshire is 5.6. Massachusetts exhibits less variation with a standard deviation of 1.20, while New Hampshire exhibits more with a standard deviation of 2.52. This increased variation may be a result of New Hampshire's smaller underlying population size presenting a smaller sample.

An important caveat in this particular case study is that the proximity of New Hampshire to the state of Massachusetts would increase the likelihood that corporations would have Massachusetts residents as customers. Statistically, this would result in contamination of the non-equivalent control group through procedural confounding. We can partially measure the potential risk posed for New Hampshire firms complying with Massachusetts regulations by looking at the overlap of incidents reported by both states. Having performed incident matching using corporate names, we can measure the degree of overlap shown in Figure 5.12. While this co-occurrence is reasonably small for Massachusetts at 13.3% of that state's incidents during this period, it is relatively large for New Hampshire, with 53.8% of its incidents overlapping.

Figure 5.12: Co-occurrence of Breach Incidents in Massachusetts and New Hampshire

| | Not in New Hampshire | Yes, in New Hampshire |
|---|---|---|
| Not in Massachusetts | 188 | 103 |
| Yes, in Massachusetts | 684 | 103 |

Intuitions about the lack of statistical difference made by the Massachusetts Data Security Law policy are confirmed by a comparative interrupted time-series result. Of specific interest is $\beta_7$, which marks a change in pre and post-implementation.

Table 5.7: Comparative ITS Massachusetts Data Security Law vs. New Hampshire

| **Parameter** | **Interpretation** | **Estimate** | **Std Error** | **Probability** |
|---|---|---|---|---|
| $\alpha$ | Intercept | 7.09 | 1.34 | 0.00*** |
| $\beta_1$ | Control Pre-Trend | −0.18 | 0.26 | 0.50 |
| $\beta_2$ | Control Post-Level Change | −0.91 | 1.73 | 0.60 |
| $\beta_3$ | Control Post-Trend Change | 0.33 | 0.37 | 0.38 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | −3.03 | 1.89 | 0.12 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | 0.23 | 0.37 | 0.55 |
| $\beta_6$ | Treatment/Control Post-Level Difference | 1.57 | 2.45 | 0.53 |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | −0.57 | 0.53 | 0.29 |

### 5.5.2 Comparison with North Carolina

Transitioning to the second quasi-experiment that can be run on Massachusetts employs the breach data reported by North Carolina. As previously mentioned, to employ this quasi-experiment required that the two states be subset to just look at incidents affecting 1000+ residents. This scoping allows us to effectively ignore the change of North Carolina's amendment to its data breach reporting legislation. This legislation titled either Session Law 2009-355 or Senate Bill 1017 amended Section 2. G.S. §75-65 on Protection from security breaches. This section was only amended by 1) changing the amount of information that those affected by data breaches were required to share and 2) the removal of the threshold for reporting (*An Act to Enhance Protections Against Identity Theft and to Protect the Credit of Crime Victims Compensation Fund Applications and Appeals*, 2021).

When focusing on just prominent incidents, those affecting over 1000 residents, there is no population effect. That suggests that we would not expect the size of the population to serve a significant role in the number of reported incidents. This can be demonstrated by looking at the raw incident counts in Figure 5.13, which show significant overlap between the two states despite the 2010 population of North Carolina at 9.5 million being almost 46% larger than the 6.5 million population of Massachusetts.

Figure 5.13: Comparing Massachusetts and North Carolina Breaches (1000+ residents)

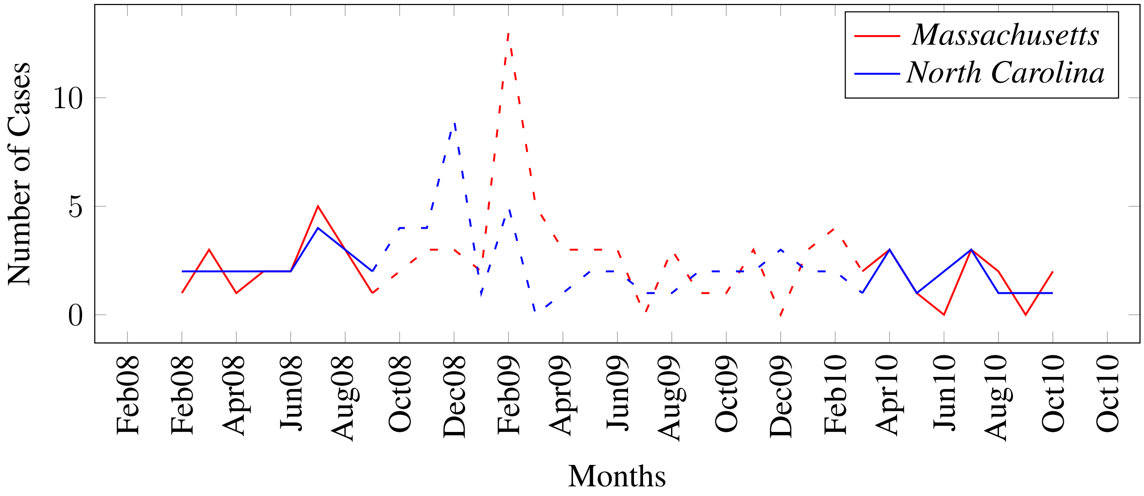


Figure 5.14: Co-occurrence of Large Breach Incidents (+1000 residents affected) in MA and NC

|  | Not in North Carolina | Yes, in North Carolina |
|---|---|---|
| Not in Massachusetts |  | 56 |
| Yes, in Massachusetts | 65 | 13 |

Similar to the prior case study, we can evaluate whether these major national incidents were concurrently reported in both states. Surprisingly, most larger incidents did not appear to have been reported in both states (at least not concurrently). Only 13 of the 134 incidents affecting 1000+ residents in one of the two states were reported to both.

Table 5.8: Comparative ITS Massachusetts Data Security Law vs. North Carolina for Large Incidents

| **Parameter** | **Interpretation** | **Estimate** | **Std Error** | **Probability** |
|---|---|---|---|---|
| $\alpha$ | Intercept | 1.79 | 0.88 | 0.5 . |
| $\beta_1$ | Control Pre-Trend | 0.13 | 0.17 | 0.46 |
| $\beta_2$ | Control Post-Level Change | –0.73 | 1.14 | 0.53 |
| $\beta_3$ | Control Post-Trend Change | –0.24 | 0.25 | 0.34 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | –0.18 | 1.24 | 0.89 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | 0.01 | 0.25 | 0.96 |
| $\beta_6$ | Treatment/Control Post-Level Difference | 0.08 | 1.61 | 0.96 |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | –0.01 | 0.35 | 0.97 |

The visual overlap between Massachusetts and North Carolina is confirmed with regression results that show no statistically significant change for any of the ITS variables except the intercept. The intercept is only weakly statistically significant at the 90% confidence level indicating an average of 1.79 incidents per month.

## 5.6 HITECH Act

Switching to the third case study, on the effect of the HITECH Act, all incidents from all collecting states through the year 2010 were coded for industry using Mergent Intellect by

FTSE Russell. This data source includes access to the D&B MDDI, which can be used to look up DUNS codes corresponding to each company manually. The DUNS codes allowed for the exporting of corporate metadata, which included NAICS codes. NAICS code 62 corresponds to the Health Care and Social Assistance sector, and NAICS code 52 corresponds to the Finance and Insurance sector. I consequently produce two quasi-experiments; the first is a comparative ITS where the health sector is contrasted with all other industries. In the second quasi-experiment, the health sector is compared to the finance industry.

### 5.6.1 Heath and Non-Health

The following chart shows the rate of change of incident frequency between the healthcare sector and all other sectors over the collection years that were coded for industry. Notably, the health sector appears to represent about 10% of breach incidents during the collection period. This ratio seems to correspond with what one would expect based on the size of that industry; US Bureau of Labor Statistics (2020) estimates suggest that 11.5 percent of jobs nationally were in the health care sector.

Subsetting this data to the relevant years, the collection period would start in the eight months before enacting the regulations (February 17, 2009) and the eight months following the enforcement of the regulations (May 27, 2009). This date range is simplified to August 2008 - November 2009.

Comparing health industry to non-health industry breaches, we see a statistically significant difference between treatment and control with $\beta_4$. This finding is intuitive, as the population of all non-health-related incidents would be significantly greater than health incidents, in this case approximately 26 fewer healthcare incidents than all other incidents. There is also a significant effect in $\beta_2$, which shows a drop of approximately 18 incidents for the control between the pre-period and the post-period. However, there was no relative change in slope or level between the treatment and control populations. In other words, there was not a statistically significant change in breach reporting in the health care indus-

Figure 5.15: HITECH Act: Health and Non-Health Coding

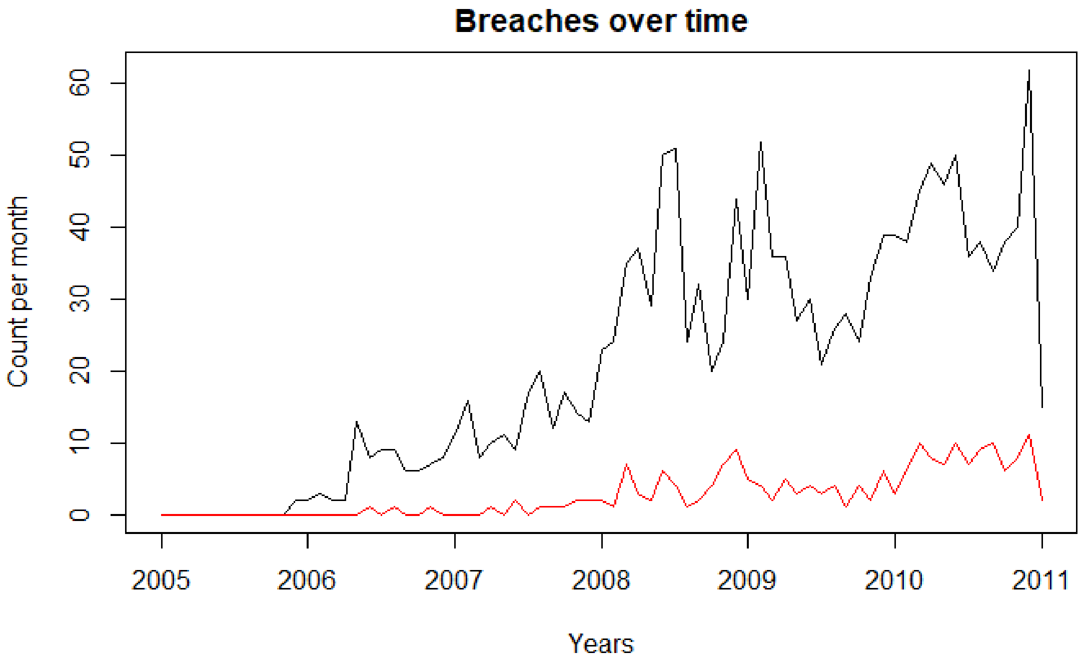
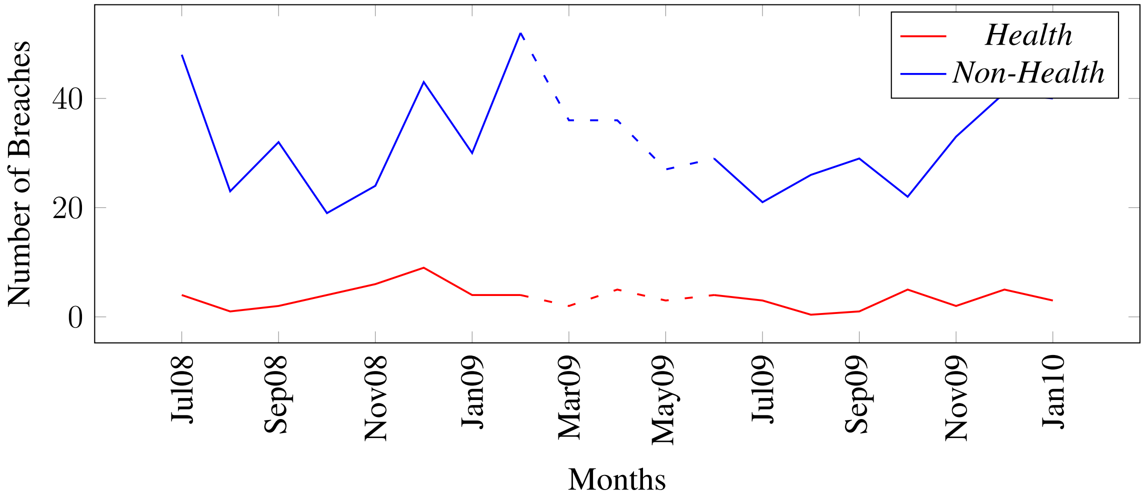


Figure 5.16: Comparing Health vs. Non-Health Breaches



try following the implementation of the HITECH Act.

### 5.6.2 Heath and Finance

The finance industry was identified as another potential control population as it represents a large portion of the overall number of incidents and has industry-specific regulations re-

Table 5.9: Comparative ITS HITECH Act (Health Sector Compared to All Other Sectors)

| **Parameter** | **Interpretation** | **Estimate** | **Std Error** | **Probability** |
|---|---|---|---|---|
| $\alpha$ | Intercept | 28.46 | 5.58 | 0.00 *** |
| $\beta_1$ | Control Pre-Trend | 1.20 | 1.10 | 0.29 |
| $\beta_2$ | Control Post-Level Change | −18.19 | 7.24 | 0.02 * |
| $\beta_3$ | Control Post-Trend Change | 1.07 | 1.56 | 0.50 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | −26.25 | 7.88 | 0.00 ** |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | −0.75 | 1.56 | 0.64 |
| $\beta_6$ | Treatment/Control Post-Level Difference | 15.68 | 10.24 | 0.14 |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | −1.51 | 2.21 | 0.50 |

quiring that it protect the privacy of its consumers (the Gramm-Leach-Bliley Act of 1999). Further, much of the more significant cybersecurity-relevant regulations affecting this industry would happen outside of the years captured (2008-2009). As mentioned above, the coding for these incidents reflects industry NAICS codes as an appropriate proxy. Where 62 covers the Health Care and Social Assistance sector and 52 covers Finance and Insurance.

The frequency of incidents for the corresponding eight months prior and post to the intervention shows that Finance has approximately three times as many reported data breaches as the health care sector during the covered dates. Visually, it is hard to tell whether the change is significant, but it does appear like there may be a modest increase in incident frequency in the finance sector and a decrease in the healthcare sector in the months following the regulatory intervention.

However, when evaluated with the regression in Table 5.10, there is no statistically significant change in the differences of level or slope for the treatment (Health) and control (Finance) sector. Instead, statistical significance is found with time $\beta_1$, with an increase of 7.7 incidents per month for pre-treatment control. Significance is also found in $\beta_2$, the difference between control pre-treatment and the control post-treatment, with a level drop of 12.7 incidents. With no statistical significance for $\beta_6$ and $\beta_7$, we can infer that the HITECH Act did not meaningfully change the level of incidents reported to the health care

Figure 5.17: Comparing Health vs. Finance Breaches

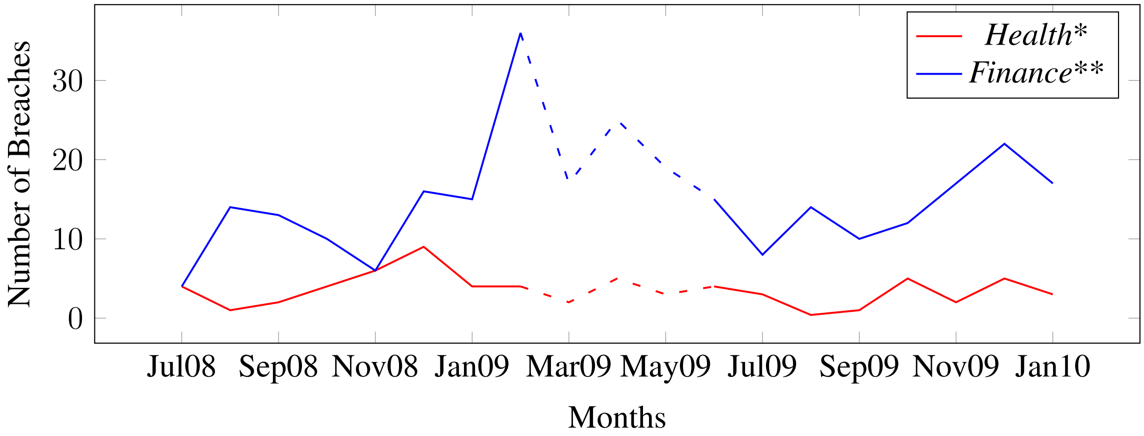


Table 5.10: Comparative ITS HITECH Act (Health Sector Compared to All Other Sectors)

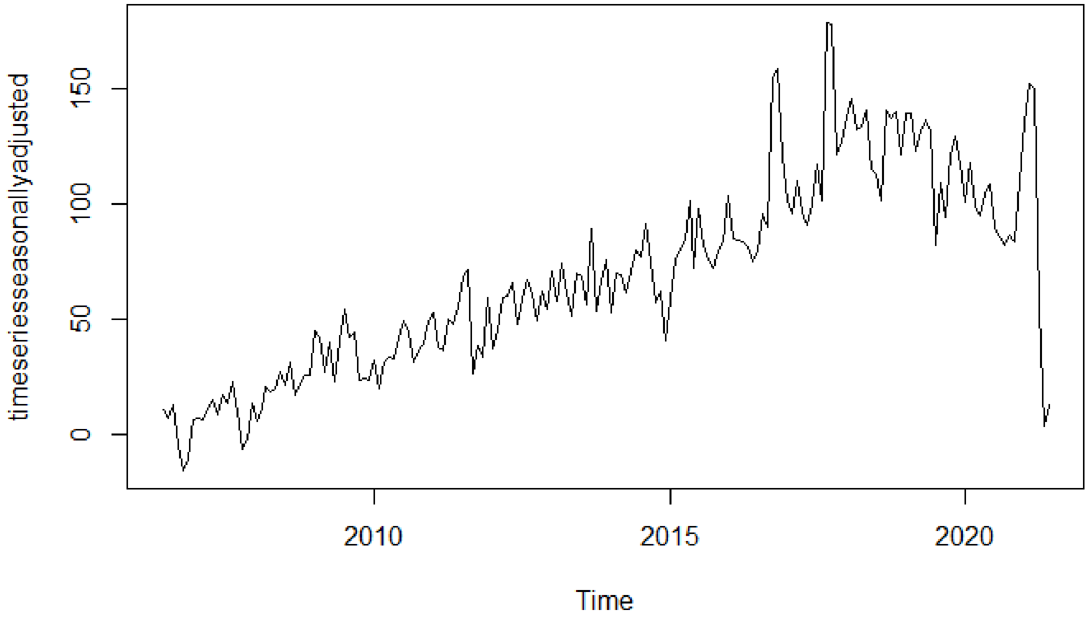
| Parameter | Interpretation | Estimate | Std Error | Probability |
|---|---|---|---|---|
| $\alpha$ | Intercept | 7.71 | 3.72 | 0.05 * |
| $\beta_1$ | Control Pre-Trend | 1.79 | 0.74 | 0.02 * |
| $\beta_2$ | Control Post-Level Change | −12.71 | 4.83 | 0.01 * |
| $\beta_3$ | Control Post-Trend Change | −0.65 | 1.04 | 0.54 |
| $\beta_4$ | Treatment/Control Pre-Level Difference | −5.50 | 5.26 | 0.31 |
| $\beta_5$ | Treatment/Control Pre-Trend Difference | −1.33 | 1.04 | 0.21 |
| $\beta_6$ | Treatment/Control Post-Level Difference | 10.20 | 6.83 | 0.15 |
| $\beta_7$ | Treatment/Control Change in Slope Difference Pre-to Post- | 0.21 | 1.47 | 0.89 |

sector.

## 5.7 Federal Trade Commission (FTC) Enforcement

In the case of the FTC's Wyndham Hotels lawsuit, a potentially fruitful quasi-experiment is produced across all states and sectors of the economy. The case was seen as an expansive extension of FTC oversight, particularly concerning PCI DSS compliance, the Payment Card Industry Data Security Standards, as well as extending potential cybersecurity liability from franchisors to their franchisees. This broad jurisdictional effect and the media attention of the incident could reasonably have shifted the curve. However, a look at the

Figure 5.18: Frequency of Breaches Across Seven States, Seasonally Adjusted

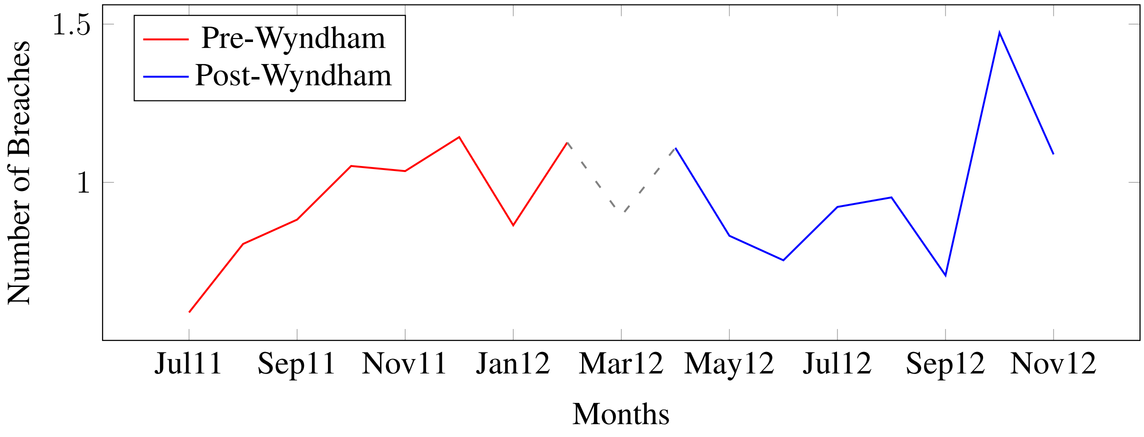


overall frequency of incidents from those states collecting during the period immediately preceding the Wyndham case (Massachusetts, New Hampshire, North Carolina, California, South Carolina, Hawaii, and Iowa) shows that the growth rate followed a macro level linear trend line from 2010 through 2017, with significant variance from month to month.

A subsample of aggregated incidents was complied based on whether they showed a dummy variable indicating a record in one of the aforementioned states. For purposes of regression analysis and to make this case consistent with the methods employed in other cases, the data was re-scaled to reflect incident frequency per population. However, as population rates of change were relatively low in the sample states and the pre and post-treatment aggregates were from the same sample population. This did not dramatically change the appearance of the above graph; however, it does change the scale such that incident frequency in 2012-2013 hovered around one breach per million residents per month in aggregate for the selected states.

### 5.7.1 FTC Complaint

Figure 5.19: The Wyndham FTC Complaint as Intervention



Subsetting the data to the relevant dates around the Wyndham case as shown in Figure 5.19, the eight months preceding the suit showed a generally positive trend line while the following eight months seem to drop before rising in the final two months. This trend demonstrates a particular challenge with this kind of stochastic data set, as drawing inference from the prior six months would fail to capture this October and November growth.

Running a regression on these results, as shown in Table 5.11, shows weak statistical significance in $\beta_1$ for a pre-intervention growth trend of .15 incidents per Million per month with a 90% confidence score. While the coefficient for $\beta_2$ is negative and would consequently suggest some evidence for a drop in incident reports in the months immediately following, this is not shown to be statistically significant. It may be possible with a smaller unit of analysis, such as with weeks and a different reporting window, to demonstrate statistical significance. However, as specified, there is no indication that the FTC's Wyndham suit reduced breach reports in the subsequent eight months.

### 5.7.2 Third Circuit Decision

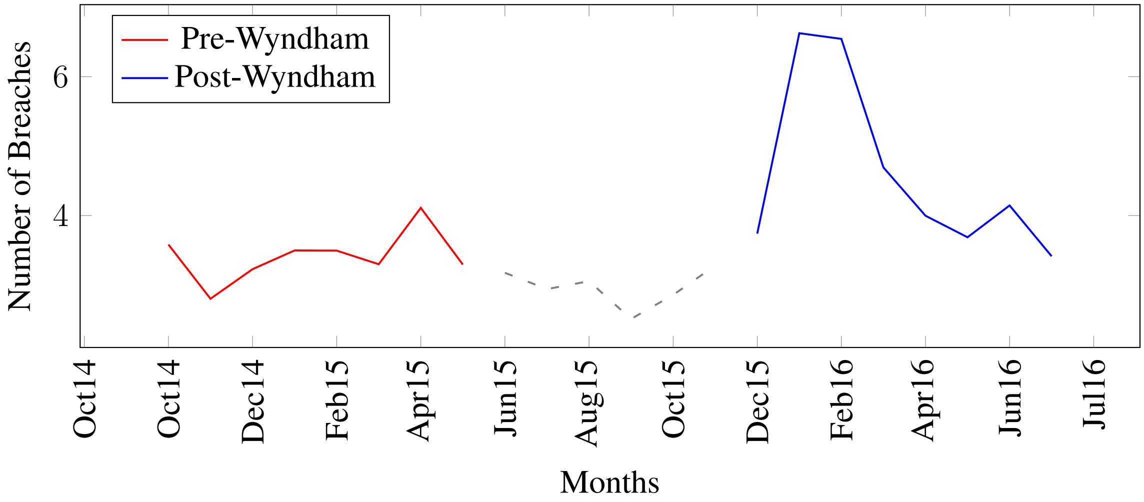
This first time-series and regression define the intervention as the initial complaint which the FTC directed against Wyndham Hotels. Alternatively, one could focus on the effect of

Table 5.11: Quasi-Experiment for Wyndham FTC Suit (FTC Complaint)

| Parameter | Interpretation | Estimate | Std Error | Probability |
|---|---|---|---|---|
| $\alpha$ | Intercept | 1.74 | 0.40 | 0.00 *** |
| $\beta_1$ | Pre-Trend | 0.15 | 0.08 | 0.09 . |
| $\beta_2$ | Post-Level Change | −0.82 | 0.52 | 0.14 |
| $\beta_3$ | Post-Trend Change | −0.06 | 0.11 | 0.62 |

Wyndham's motion to dismiss and subsequent appeal, which put the authority of the FTC in these matters into doubt. The subsequent decision by the Third Circuit Court (August 24, 2015), which confirmed the FTC's authority and the ultimate decision by Wyndham Hotel to settle (December 9, 2015), serves as dates that present another reasonable intervention. The difference-in-difference for this intervention shows a net increase rather than a decrease in the level of reported Breaches per capita from the targeted states.

Figure 5.20: The Wyndham FTC Third Circuit Decision as Intervention



This alternative time-series produces a distinct set of regression results, which show statistical significance in the increase of breaches post-treatment; however, it suggests a negative slope of the frequency of breaches in the eight months following the intervention.

Table 5.12: Quasi-Experiment for Wyndham FTC Suit (Third Circuit Decision)

| **Parameter** | **Interpretation** | **Estimate** | **Std Error** | **Probability** |
|---|---|---|---|---|
| $\alpha$ | Intercept | 3.16 | 0.67 | 0.00 *** |
| $\beta_1$ | Pre-Trend | 0.06 | 0.13 | 0.68 |
| $\beta_2$ | Post-Level Change | 2.28 | 0.87 | 0.02 * |
| $\beta_3$ | Post-Trend Change | –0.34 | 0.19 | 0.09 . |

## 5.8 Statistical Verification

Assumed in statistical time-series analysis is that monthly incident accounts are stationary time-series and that the incidents are sufficiently stochastic that one can ignore the possibility of auto-correlation or seasonal auto-regressive terms. The presence of autocorrelation can be demonstrated visually with a plot produced by Auto- and Cross- Covariance and -Correlation Functions as well as with Partial Autocorrelation Functions. The time-series data can also be tested as stationary (i.e., the statistical properties of a time-series such as mean, standard errors, and autocorrelation do not change over time). One such test is the Augmented Dickey-Fuller Test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests.

Time-series employed in the quasi-experiments were tested for both autocorrelation and partial autocorrelation. These charts are provided in the Appendix in Figure C.1 through Figure C.12. The ACF and PACF plots for incident frequency provided in the Appendix demonstrate that most of the spikes were not statistically significant. While occasionally trend spikes that dropped above or below a 95% confidence interval, consequently, the plots were not redesigned as autoregressive integrated moving average (ARIMA) models.

# CHAPTER 6
# IMPLICATIONS

This research is the first quasi-experimental policy evaluation of information security regulations to identify a cybersecurity reform that meaningfully reduces breaches. That said, in three of the four case studies developed for this research, statistically significant evidence of a reduction in breaches was not observed. Why was this case? What attributes about the NY DFS regulations made this activity seemingly more efficacious? This chapter will explore the potential impact of spatial and temporal factors discussed in section 5.3. The chapter's core will then address comparative questions between the four cases, developing hypotheses that might explain these observations. This chapter will also investigate the magnitude and duration of the NY DFS regulatory interventions' efficacy. Finally, it will provide some recommendations for future research and policy development that draws on these findings.

## 6.1 Impact of Spatial and Temporal Factors

Two findings from section 5.3 are particularly worthy of policy consideration. Preliminary evidence for localization and seasonality present promising topics for future research and valuable knowledge for businesses and policy entrepreneurs alike.

Given that breach reporting seems to spike around March and dips in November, how could businesses use this information to their advantage. If the cause of this seasonal increase (likely instigated by breaches that occurred in prior months) can be identified and isolated, then training materials could be aligned with their seasonal cause. Once linked to the holidays or the end of the fiscal year, employees could be warned of potentially timely fraud schemes, phishing attacks, or theft of devices. Even without identifying a cause, seasonal security hires might enable some firms to prepare for the worst.

The localization of breach incidents presents a separate set of policy expectations. If most breaches are not national in scope, perhaps the incidents are linked to smaller firms or localized branches with a local customer base. What resources do these firms need to adequately respond to more minor incidents? How can the costs of compliance be lowered? Perhaps states could employ online forms that export notification letter so that small businesses would not need to rely on lawyers to faciliate reporting. The Department of Homeland Security and the National Institute of Standards and Technology could develop more resources targeted at a lower level of cyber maturity. Localization also suggests that state-level interventions targeted at in-state firms (like the NY DFS regulations) may be more impactful on total breaches than one might anticipate.

## 6.2 Comparing Cyber Regulatory Efforts

There are four different aspects to the regulatory interventions that can be compared, their scope, substance, implementation period, and penalties. Across these four aspects, the NY DFS designed a unique structure for their cybersecurity regulations which is highly targeted, carries substantive measures, employs staged implementation phases, and uses a novel penalty structure. While I can not with the current results show a specific relationship between these measures and a change in efficacy, they present valuable guidelines for how future regulations might be structured and the dimensions regulators might experiment with to create new policy designs.

### 6.2.1 Policy Scope

While discussed in greater depth in chapter 3, I designed the structure for the case selection to include four quasi-experiments employing combinations of industry and state coverage. The degree of coverage for each case consequently varies. The Massachusetts Data Security Law affects any company with PII data covering residents of the state. The cybersecurity regulations in the HITECH Act apply to the handling of PHI. PHI is defined by 45 CFR § 160.103 as 'individually identifiable health information.' The FTC's Section 5 enforcement capability is nationally relevant across all industries and allows it to target any company that has failed to reasonably protect consumers' information. In contrast, the NY DFS targeted its cybersecurity provisions at regulated entities and licensed persons in the financial services and insurance sector for the state of New York.

This scoping issue becomes relevant if different affected populations vary in their compliance with the regulatory change. For example, the financial services sector presents an already heavily regulated industry that is likely to be more capable of effectively complying with new regulations than other sectors. Similarly, the health care sector is representative of a regulated sector capable of implementing new compliance measures. In contrast, the Massachusetts Law and FTC Section 5 enforcement cover all industries and consequently include lightly regulated business sectors, including companies less likely to track and comport with new regulations. Accordingly, the regulatory responsiveness of the covered population may serve as a moderating variable that could reduce the efficacy and potentially the statistical significance of analysis when viewed across all sectors.

### 6.2.2 Contents of Regulations

Perhaps most significantly, there are fundamental differences between these policies in what they require of organizations to comply. Of the four regulatory interventions, each employs a unique policy design, though often covering overlapping topics. This topical coverage can be compared across each intervention, including organizational requirements (Table 6.1) and computer security requirements (Table 6.2). The following tables identify whether the language in the four regulatory interventions addresses specific regulatory requirements, coded as either 'yes' or 'no.' However, as the FTC's complaint against Wyndham is a case rather than new regulation, the additional context of '(not new)' is appended to 'yes' to differentiate where previous FTC actions had covered the same ground. In the following paragraphs, I discuss each intervention's topical coverage in turn.

Table 6.1: Organizational Regulatory Requirements

| | MA Data Security Law | HITECH Act | FTC Section 5 (Wyndham Hotel) | NY Dept of Financial Services |
|---|---|---|---|---|
| Designation of specific personnel | Yes | Yes | No | Yes |
| Education and training of employees | Yes | Yes | No | Yes |
| Creation and maintenance of cyber policies | Yes | Yes | No | Yes |
| Notification of Breaches | Yes | Yes | No | Yes |
| Certification of compliance | No | No | No | Yes |

Table 6.2: Computer Security Requirements

| | MA Data Security Law | HITECH Act | FTC Section 5 (Wyndham Hotel) | NY Dept of Financial Services |
|---|---|---|---|---|
| Secure user authentication protocols | Yes | Yes | Yes (not new) | Yes |
| Secure access control measures | Yes | Yes | Yes (not new) | Yes |
| Encryption requirements | Yes | Yes | Yes | Yes |
| Reasonably up-to-date security software, patches, virus definitions | Yes | Yes | Yes (not new) | Yes |
| Control third-party access to network | Yes | Yes | Yes | Yes |

The substance of the Massachusetts Data Security Law includes both provisions written into the law and the regulations mandated by OCABR that were instigated by the law. Necessary for the quasi-experimental design, the law contained a requirement for state-level breach notification (M.G.L. § 93H-1 et seq.) with the earliest reports submitted in November and December of 2007. OCABR only passed the regulations instigated by the law (201 C.M.R. 17.00, et. seq) in September of 2008. Consequently, the quasi-experiment was made possible because a breach reporting requirement was already in effect in the months preceding OCABR's implementation of new regulations. The corresponding regulatory language that addresses each category is cited here:

- Designation of specific personnel (*201 CMR 17.00: Standards for the Protection of Personal Information of Residents of the Commonwealth*, 2020, 201 CMR 17.03.2a)
- Education and training of employees (201 CMR 17.03.2b1)
- Creation and maintenance of cyber policies (201 CMR 17.03.1)
- Notification of breaches (MGLA 93H § 3(a)) (*Regulations to Safeguard Personal*

*Information of Commonwealth Residents*, 2020)

- Secure user authentication protocols (*201 CMR 17.00: Standards for the Protection of Personal Information of Residents of the Commonwealth*, 2020, 201 CMR 17.04.1)
- Secure access control measures (201 CMR 17.04.2)
- Encryption Requirements (201 CMR 17.04.3 and 17.04.5)
- Reasonably up-to-date security software, patches, virus definitions (201 CMR 17.04.7)
- Control third-party access to network (201 CMR 17.03.2f)

HIPAA's Security Rule discussed in section 3.2 had since 2003 established administrative and technical regulations affecting cybersecurity. Consequently, while not directly addressed in the HITECH Act legislation or regulation, the expansion of HIPAA rules to covered entities addresses a wide range of organizational and computer security requirements. Like the Massachusetts Data Security Law, the HITECH Act's notification of breach requirements led to the earliest report being submitted in October of 2009, while the regulation itself was not implemented until March 1, 2010. The corresponding regulatory language for each category is cited here:

Organizational Regulatory Requirements

- Designation of specific personnel (*Federal Register* 2003, 45 CFR § 164.308(a)(2)))
- Education and training of employees (45 CFR § 164.308(a)(5)(i))
- Creation and maintenance of cyber policies (45 CFR § 164.308(a)(1)(i))
- Notification of breaches (*Federal Register* 2009, 45 CFR § 164.408)

Computer Security Requirements

- Secure user authentication protocols (*Federal Register* 2003, 45 CFR § 164.312(d))

- Secure access control measures (*Federal Register* 2003, 45 CFR § 164.312(a)(1))
- Encryption Requirements (45 CFR §164.312(e)(1))
- Reasonably up-to-date security software, patches, virus definitions (45 CFR § 164.308(a)(5)(ii)(A))
- Control third-party access to network (45 CFR § 164.314.2)

While the Wyndham Hotel case is remembered for its subsequent litigation, it expanded on prior FTC precedent. The FTC had previously employed its Section 5(a) enforcement powers before its complaint against Wyndham Hotel. For example, looking at an earlier August 2008 complaint against TJ Max shows that the complaint covered similar issues to those of the Wyndham case: including failing to authenticate users, adequately restrict network access, left PII in plain text, and did not update their systems. Given that the FTC complaint resembled prior cases, the intervention should be understood as limited to topics not previously addressed. Novel content issues covered by the complaint include the role of the Payment Card Industry (PCI) encryption standard and an expansion of third-party access requirements covering subsidiaries. As described above, the appending of (not new) was added to content categories addressed in prior complaints to clarify this distinction. As the complaint is not a regulation, the regulation categories are matched to sections in the complaint enumerated as "Defendants' Inadequate Data Security Practices." The assessment of inadequacy seems to imply the FTC's authority over these substantive domains.

- Secure user authentication protocols (*FTC vs Wyndham Worldwide Corporation First Amended Complaint for Injunctive and Other Equitable Relief* 2012, 24(h))
- Secure access control measures (24(a))
- Encryption Requirements (24(b))
- Reasonably up-to-date security software, patches, virus definitions (24(d))

- Control third-party access to network (*FTC vs Wyndham Worldwide Corporation First Amended Complaint for Injunctive and Other Equitable Relief* 2012, 24(j))

The NY DFS directly expanded new regulations under 23 NYCRR 500. This regulatory expansion included new notification of breach requirements to the Superintendent of Financial Services. However, this additional regulation need not be treated separately from the core intervention, as was the case in the Massachusetts Data Security Law or the HITECH Act, as alternative pre-existing notification requirements in other jurisdictions were already in place. References to the corresponding policy language are cited here:

Organizational Regulatory Requirements

- Designation of specific personnel (Vullo 2017, Section 500.10)
- Education and training of employees (Section 500.14)
- Creation and maintenance of cyber policies (Section 500.03)
- Notification of breaches (Section 500.17)

Computer Security Requirements

- Secure user authentication protocols (Section 500.14(a))
- Secure access control measures (Section 500.12(b))
- Encryption Requirements (Section 500.15)
- Reasonably up-to-date security software, patches,
  virus definitions (Section 500.05)
- Control third-party access to network (Section 500.11)

The content of these regulatory interventions demonstrates that all four regulations seek to establish standards on similar topics. However, simply demonstrating coverage does not suggest the degree of cyber maturity required. For example, while the FTC's complaint against Wyndham highlights simple passwords reused for both usernames and passwords, the NY DFS regulations require dual-authentication. The FTC's authority, as demonstrated by the Wyndham case, seems the most different from the others as its complaint was limited in scope to technical rather than organizational best practices. Perhaps uniquely, the NY DFS regulations require a certification of compliance with the Superintendent. This suggests that the NY DFS regulations are more extensive in their requirements than the other three interventions.

### 6.2.3 Implementation Period

Within the policy process literature, 'implementation' follows policy 'selection' (Brewer 1974). Consequently, in both the Massachusetts Data Security Law and the HITECH Act, the passage of the legislation was only one step. The regulatory process that followed involved propagating rules, collecting feedback, and setting an ultimate date for enforcement. Complex regulatory changes require time for the regulated business sector to adapt. Cybersecurity regulations that require changes to corporate policy, purchasing solutions, and employee training are particularly complex. The period of implementation modeled below in Figure 6.1 measures the number of days for compliance between the propagation of proposed rules and their enforcement dates.

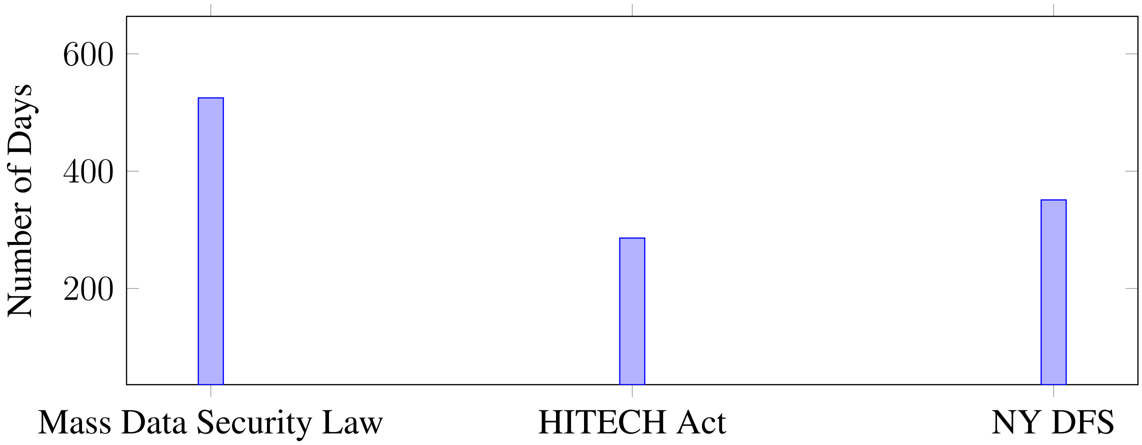
The Massachusetts Data Security Law had a single deadline for compliance. While the initial legislation passed on August 3, 2007, the final regulations were not filed by OCABR until September 22, 2008. These regulations were initially expected to require compliance on January 1, 2009, but were extended on three separate occasions, ultimately becoming effective on March 1, 2010. The first of these extensions was justified "in light of intervening economic circumstances" (Brenner 2008). An interim May 1 deadline was considered by OCABR to be in line with the FTC's Red Flag Rule. The regulations were also modified on August 17, 2009, to reflect a more risk-based and technology-neutral approach. Ultimately, businesses were left with 525 days to achieve compliance.

Congress signed the HITECH Act on February 17, 2009. Regulations covering the Section D cybersecurity components were first published on April 27, 2009. The interim final rule became effective on November 30, 2009. In total, this includes just 286 days for companies to prepare for compliance. There were separate deadlines for compliance with the breach reporting requirements to the HHS Secretary, which went into effect on September 23, 2009 (Civil Rights 2009).

NY DFS designed the regulation to be implemented in stages. The initial rules were drafted in September 2016 and open for comments through December 28, 2016. These initial rules assumed that companies would have a transitional period to prepare for compliance. Covered entities would be given 180 days from the effective date to prepare. A two-month delay pushed the initial January 1st effective date to March 1, 2017. After the rules became effective, certification with the state was not required until February 15, 2018. Combined, this extended out compliance for covered business by a year, ultimately leaving 351 days between formalizing the rules (their effective date) and their final enforcement through certification. This date range was selected for the quasi-experiments, as the earlier propagated rules were modified. Two additional deadlines would follow this initial certification. Compliance with monitoring and encryption requirements was initially estimated to take 18 months, and third-party service requirements were expected to take two years.

Identifying implementation dates to use as part of the quasi-experiments requires some degree of interpretive evaluation. Should one initiate the quasi-experiment upon the date legislation is passed, when regulations are drafted, or when regulations come into effect? Critically, this period presents an opportunity for businesses to come into compliance. If this period was too short, we might expect the effects of the regulation to not be observable until many months later. If the implementation phase is too long, some of these changes could become subject to broader trends affecting the relevant states or industries.

Figure 6.1: Implementation Days

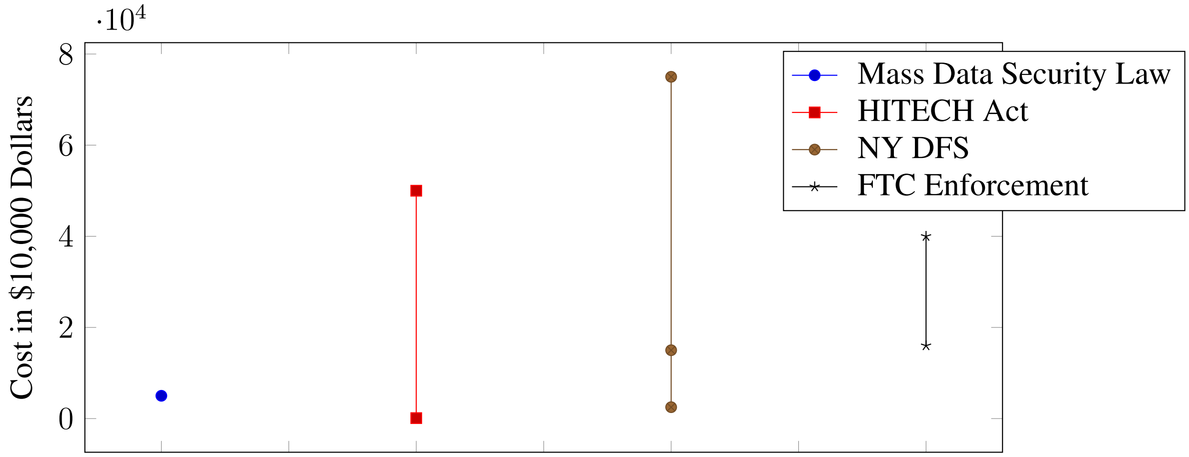


### 6.2.4 Regulatory Penalties

In the policy process literature, the implementation phase extends beyond the date of enforcement to enforcement actions themselves. These enforcement actions serve in some ways as additional quasi-experiments, moments where companies' sense of regulatory risk adjusts to new realities. However, the magnitude of penalties regulators can assign to would-be violators is frequently constrained under the law. These maximum limits on penalties provide the private sector with information that informs a regulatory risk assessment. When fines are constrained, the costs of non-compliance are more acceptable. The four cases identified in this research assign various maximum penalties described in the following list and graphed in Figure 6.2.

- MA Data Security Law Penalties have a maximum limit per violation of $5,000
- HITECH Act Penalties are limited per violation at $100 to $50,000.
- NY DFS regulations Penalties have a maximum limit per day of $2,500 (any-violation) $15,000 (negligence) $75,000 (knowing)
- FTC Enforcement Penalties have a maximum limit per violation of $16,000 (pre-2016) $40,000 (post-2016)

Figure 6.2: Maximum Penalty



A review of the regulatory costs shows that the NY DFS regulations have a higher penalty than either of the three other cases. However, this compares non-equivalent units, as the NY DFS regulations have penalties that are limited per day, where the other three have penalties that are limited per violation. Further, the NY DFS regulations are more tiered than those demonstrated in the other three laws. The distinction between any violation, negligence and a knowing violation pre-determines how the regulator will penalize would-be violators.

Comparing early regulatory enforcement (discussed in greater depth in chapter 3), the first enforcement action by Massachusetts was against the Briar Group, and the first settlement following the HITECH Act was with Blue Cross Blue Shield of Tennessee. While not the first penalty for violating the NY DFS cybersecurity regulations, the first payment was made as part of a settlement by Residential Mortgage Inc. Comparing these three penalties for non-compliance (see Table 6.3), it is notable that none occurred within the first year of enforcement. However, these incidents occurred several years before they were settled. The timing of these enforcement actions means that while compliance in the first year following implementation may matter, companies would not know how aggressive the regulatory agency would be until much later. As for the magnitude of the penalties, proportionate to the incident, it does appear that the Massachusetts Data Security Law is the least severe, and the NY DFS regulations are the most severe.

Table 6.3: First Regulatory Enforcement

| State Law | Massachusetts Data Security Law | HITECH Act | NY DFS cyber regulations |
|---|---|---|---|
| First Enforcement | Briar Group | Blue Cross Blue Shield of Tennessee | Residential Mortgage Inc. |
| Scope of Incident | Hack lasted 8 months, took credit card data. | Theft of hard drives over 1 million individuals affected | Phishing attack accessed mailing list (unreported) |
| Penalty | \$110,000 | \$1,500,000 | \$1,500,000 |
| Days Since Regulation Has Been Enforced | 372 days | 1,017 days | 1,112 days |

### 6.3 Estimating Savings from NY DFS Regulations

In the prior findings section, the data from Connecticut suggested that the NY DFS regulations produced a reduction in 2.24 observed breaches in the post-treatment period. If we extend this out across the 12 post-treatment months, this totals 26.88 incidents. Given this incident count, we can use industry cost estimates and apply simplified formulas to approximate the total regulatory savings. To this end, I leverage data from the 2020 Cyentia study and the 2017 Ponemon study. While the Cyentia study is better designed to assess cost estimates for a range of potential breach sizes, the Ponemon study leverages data closer to the appropriate year and distinguishes between direct and indirect costs. Applying different methodologies drawing on these data sources, we get a range of estimates of potential savings. These calculations start with a loss table copied from the Cyentia report shown in Table 6.4.

This loss table can be modified to estimate the percentage of incidents expected at each loss value by subtracting each column by the preceding one (Table 6.5). The first column can be subtracted from 100% to define some percentage of incidents as costing less than $10K, a continuation of the exponential sequence, would assign $1000 for these incidents. Given a distribution of incident magnitude and a total number of incidents, we could use

Table 6.4: Probable losses based on records affected in a breach (Cyentia Report)

| | | Probability of At Least This Much Loss | | | | | |
|---|---|---|---|---|---|---|---|
| | | $10K | $100K | $1M | $10M | $100M | $1B |
| No. of Records | 100 | 82.0% | 49.9% | 17.8% | 3.3% | 0.3% | 0.0% |
| | 1K | 88.4% | 60.9% | 26.0% | 5.9% | 0.7% | 0.0% |
| | 10K | 93.0% | 71.1% | 35.8% | 10.0% | 1.4% | 0.1% |
| | 100K | 96.0% | 79.8% | 46.7% | 15.8% | 2.7% | 0.2% |
| | 1M | 97.9% | 86.7% | 57.7% | 23.5% | 5.0% | 0.5% |
| | 10M | 99.0% | 91.8% | 68.2% | 32.8% | 8.6% | 1.1% |
| | 100M | 99.5% | 95.3% | 77.4% | 43.4% | 13.9% | 2.3% |
| | 1B | 99.8% | 97.4% | 84.9% | 54.5% | 21.0% | 4.2% |
| | 10B | 99.9% | 98.7% | 90.5% | 65.3% | 30.0% | 7.4% |

this loss table to approximate the total cost from a sample of incidents.

Table 6.5: Modified Loss Table Using Cyentia Data

| | | Probability of At Least This Much Loss | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $1K | $10K | $100K | $1M | $10M | $100M | $1B |
| No. of Records | 100 | 18.00 % | 32.10 % | 32.10 % | 14.50 % | 3.00 % | 0.30 % | 0.00 % |
| | 1K | 11.60 % | 27.50 % | 34.90 % | 20.10 % | 5.20 % | 0.70 % | 0.00 % |
| | 10K | 7.00 % | 21.90 % | 35.30 % | 25.80 % | 8.60 % | 1.30 % | 0.10 % |
| | 100K | 4.00 % | 16.20 % | 33.10 % | 30.90 % | 13.10 % | 2.50 % | 0.20 % |
| | 1M | 2.10 % | 11.20 % | 29.00 % | 34.20 % | 18.50 % | 4.50 % | 0.50 % |
| | 10M | 1.00 % | 7.20 % | 23.60 % | 35.40 % | 24.20 % | 7.50 % | 1.10 % |
| | 100M | 0.50 % | 4.20 % | 17.90 % | 34.00 % | 29.50 % | 11.60 % | 2.30 % |
| | 1B | 0.20 % | 2.40 % | 12.50 % | 30.40 % | 33.50 % | 16.80 % | 4.20 % |
| | 10B | 0.10 % | 1.20 % | 8.20 % | 25.20 % | 35.30 % | 22.60 % | 7.40 % |

The distribution of incident size can be approximated for the financial sector, using the 150 financial breach incidents reported to Maine in 2020.[1] The total number of incidents would, as discussed above, be 26.88 avoided incidents over a year. We can then multiply this total number (26.88) by the percentages assigned to each magnitude (22.0% between 100-1k, 38.6% between 1k-10k, 18.6% between 10k-100k, 4.0% between 100k-1M, 1.3% between 1M-10M, 0.6% between 10M-100M) to calculate the approximate number of incidents for each magnitude.[2] The loss levels can then be multiplied by the incident fre-

[1] Unlike Connecticut, the state of Maine reports total affected in addition to state residents affected

[2] Financial incidents larger than 100 million total affected (i.e., globally) were not reported to Maine in 2020.

quency for each breach size (sometimes a fraction of an incident) to estimate the number of incidents at each minimum loss level. These frequencies can then be multiplied by the corresponding minimum amount lost to create an expected cost for each pairing of breach size and loss level, as shown in Table 6.6.

Table 6.6: Estimated Cost for Different Categories

| | $1K | $10K | $100K | $1M | $10M | $100M | Total |
|---|---:|---:|---:|---:|---:|---:|---:|
| 100 | $686 | $16,262 | $206,385 | $1,188,634 | $3,075,072 | $4,139,520 | $8,626,559 |
| 1k | $726 | $22,723 | $366,262 | $2,676,925 | $8,923,085 | $13,488,384 | $25,478,105 |
| 10k | $200 | $8,099 | $165,489 | $1,544,901 | $6,549,581 | $12,499,200 | $20,767,471 |
| 100k | $23 | $1,204 | $31,181 | $367,718 | $1,989,120 | $4,838,400 | $7,227,646 |
| 1M | $3 | $252 | $8,247 | $123,702 | $845,645 | $2,620,800 | $3,598,648 |
| 10M | $3 | $288 | $12,256 | $232,805 | $2,019,924 | $7,942,752 | $2,404,415 |
| Total | $1,639 | $48,608 | $780,450 | $5,956,716 | $21,858,278 | $39,457,152 | $68,102,843 |

Taking the summation of this table provides an estimate for the saving from the NY DFS regulations in the year after the implementation period as 68.10 Million dollars. This estimate is a product of a summation of probabilities. Notably, more than half of this total estimate comes from breaches costing 100+ million. However, the model estimated less than one incident (0.41) of the expected 26.88 would have that loss amount. The magnitude of losses is consequently highly dependent on a single high-cost incident being avoided.

Nevertheless, the methodology of estimation proscribed above would be conservative in its estimation. For one, incident frequencies were multiplied by minimum losses. Breach trends are often non-linear; for example, it is likely that the average cost of breaches between one and ten million will be closer to one million, but we would still expect the average cost to be greater than this minimum. Secondly, incidents smaller than 100 were excluded from the sample because Cyentia's loss table lacked a corresponding distribution. However, based on the Maine distribution, smaller incidents were only expected to account for 17 of the 141 incidents. If we assumed the same distribution of costs for smaller incidents as with those affecting 100-1000, these incidents would add approximately 2.5 million to the estimated 68.1 million for a total of 70.6 million. Further, the observed breaches from Maine did not include larger incidents, which, as demonstrated in Cyentia's loss estimates, are much more likely to result in severe losses. It is possible that the size distribution of incidents that affected Maine are not generalizable; there are good reasons to believe they could either understate or overstate the average total breach size. Data from Connecticut could not be used as they only collect the number of state residents affected.

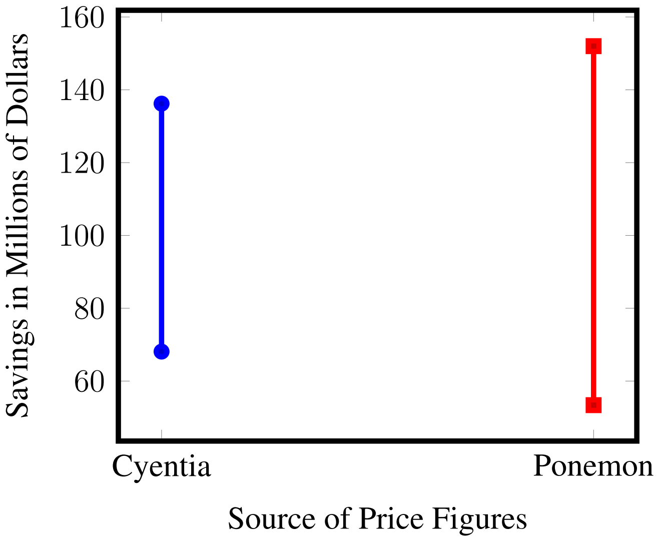
The work of Cyentia highlights how the data reported by the Ponemon studies are frequently misapplied to create inaccurate estimates of the overall costs of breaches. Consequently, a simple application of Ponemon's average cost record using the distribution of breach sizes from Maine would skew the average incident to unrealistic proportions (213k on average) and produce unrealistic estimates. However, as the Cyentia report notes, the Ponemon study is derived using data in the range of 10-100 thousand affected users. Limiting the Maine data to just this subset suggests an average incident size of 33,353 users per record. However, this incident type accounts for only 18.6% of the data.

Cyentia itself claims to include indirect costs such as competitive advantage and reputation, however, notes that these attributes might be less systematically accounted for in their estimate. In a discussion of their cost estimates, Cyentia described the role of Loss Magnitude in the Factor Analysis of Information Risk (FAIR) framework.

> "FAIR$^{TM}$ breaks that down further by bucketing losses into six forms: productivity, response, replacement, competitive advantage, fines and judgments, and reputation. Advisen tracks similar categories of losses, but we do not differentiate among them in this study. Some of these are probably better represented by our data (e.g., fines and judgments) than others (e.g., productivity and reputation)."

In contrast, the Ponemon study provides breakdowns of direct and indirect costs. In 2017, they estimated that 64% of total costs for US breaches were a result of indirect expenses (like customer loss and reputation). Ponemon clarifies that direct costs would include activities likes the hiring of forensic experts or a law firm for assistance and offering

Figure 6.3: Savings from Regulation over 1 Year



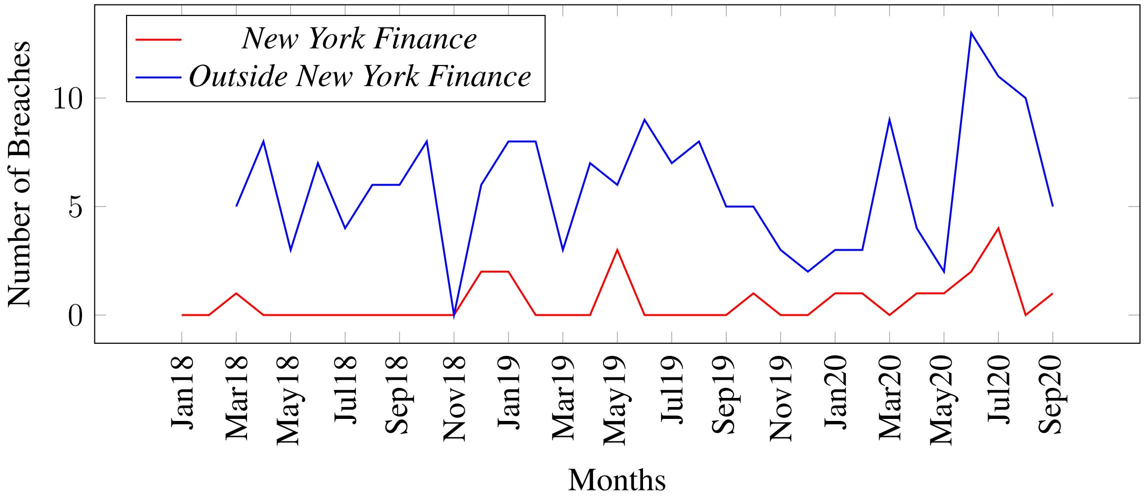
victims identity protection services. We can consequently identify how Cyentia's estimates fall between the direct costs and the indirect costs reported by Ponemon, and may use Ponemon's estimates to approximate a range of potential cost estimates. This approach assumes that the proportion of costs associated with incidents affecting 10-100k would continue across the size distribution. Based on this lower and upper bound, the Ponemon costs range from 78.4% to 223.2% of the Cyentia estimates. A more accurate estimate of the Cyentia figures might also employ a 200% multiplier that would approximate the average cost rather than a minimum (i.e., the average cost for incidents costing more than one million and less than ten million could easily average two million. These estimates are shown in Figure 6.3.

While many indirect costs are internalized by the business, for example, customer loss and brand damage, these would not be socially felt in a competitive marketplace. Consequently, the additional $149 of indirect expenses estimated by Ponemon may significantly exaggerate the social cost. However, there are certainly other social factors, relating to societal trust and consumer privacy, that would be additive to these direct costs. A cost-benefit analysis could attempt to identify the value of these societal costs. Overall, an estimate of approximately 100 million dollars in direct costs saved from the regulation can serve as a conservative estimation of the benefits of the intervention using only the change in incidents (where the submitter identified an industry) reported in the state of Connecticut. As for the costs, how compliance costs are passed onto consumers, and whether there are net social benefits from these changes is a question for future research.

## 6.4 Persistence of NY DFS Regulatory Effects

Given the savings observed from one year, we might ask for how long this trend of improved performance by the state of New York could continue. Temporarily, data challenges with New York and Connecticut records prevent us from drawing more direct conclusions. In the first case, the requisite information is not coded, in the second, the information from the open records request I filed ends in October 2019. However, data collected by Maine suggests that the observed gap between financial incidents nationally and those in New York would tighten in 2019 and 2020. Evidence for this breach growth is shown in Figure 6.4.

Figure 6.4: New York Financial Breach Growth in 2020



This increase is demonstrated by the .25 New York financial incidents observed per month in 2018, doubling to .5 in 2019, and based on the first nine months of 2020, more than doubling again to 1.22 incidents. In contrast, the growth rate was slower for other reported breaches from 5.25 per month in 2018, 5.91 in 2019, and 6.66 in 2020. This suggests a potential tightening in the level difference between New York and the rest of the country. Additional data from Connecticut, or New York may confirm this finding once processed and collected. Assuming the change in breach performance is not a statistical anomaly, this may imply the benefits from regulation are temporary. A temporary effect could be explained by the wider industry catching up to a new standard of best practices. Alternatively, the regulations might be effective by directing organizational attention to the problem that might wane over time. A third potential hypothesis might look for the delayed impact of additional regulations that the NY DFS implemented the following year.

## 6.5 Integrating Findings

This chapter has sought to explore the implications of the NY DFS regulatory reduction in breaches. Why was this case demonstratively effective while others were not? Is this effect significant? To address these questions, the four case studies have been comparatively evaluated, the degree of expected breach reduction from the NY DFS regulations has been assessed in dollars, and the persistence of this effect was explored. Compared to the other three regulatory interventions, the NY DFS regulations were more targeted in their scope (i.e., just the NY financial and insurance market), more extensive in the content of their regulatory requirements, provided an intermediate length of time for implementation, and employed a novel tiered maximum penalty formula. These actions all seem consistent with the capacity of the NY DFS agency to learn from prior regulatory actions, both internal and external.

The magnitude of the direct breach costs avoided by the NY DFS regulations is estimated in section 6.3 as approximately 100 million dollars. This number drew on the change in the pre to the post-intervention difference between treatment and control for the Connecticut quasi-experiments. This robustness check presented the strongest statistical evidence, but I reduced estimation in the size of the effect. Additional costs could easily accrue if additional New York breaches were not registered in Connecticut but still avoided or if we measured indirect societal benefits. Whether these savings are worth the costs for regulatory compliance is an important question. Regardless, conservative estimates for savings in the hundreds of millions of dollars are sufficiently large to invoke attention. Many of the compliance costs would likely be front-loaded, a cost-benefit assessment of the regulations will be highly dependent on the persistence of benefits. A preliminary analysis of more recent data from Maine, suggests that this effect may not be enduring.

## 6.6 Future Research

New cybersecurity, data security, and privacy legislation have started to proliferate, particularly in response to the implementation of the European Union's General Data Protection Regulation (GDPR). Yet, as new data becomes available and experiments present themselves, an empirical and policy-focused community that can evaluate these changes is still needed. Future comparative cybersecurity research leveraging this kind of analysis is likely to grow in its usefulness. Global trends in data collection and an increasingly large legal compliance regime means that even if US reporting requirements remain defederalized, a large and engaged non-academic audience (both legal and technical) exists who will hopefully be interested in this and future research findings. Yet, a window exists for experimentation before national norms are proposed, increasingly the business community is frustrated with the complexity of compliance requirements both in the United States and around the globe. If national model legislation is to be developed, rich empirical findings should supplement subjective expert opinion.

Consequently, this work has aimed for generativity, by creating a method for the restructuring of irregularly published state data. New research can build on this shared resource. While the Privacy Rights Clearinghouse remains a popular data source for analysis, it lacks important metadata and state-level dummy variables. Private sector firms like Advisen Ltd have built impressive data sources, however, access to these sources is understandably limited. Bringing some of this information into the public domain will hopefully assist the policy community.

So what kind of questions might be asked that could build on this research? Critically, this research only evaluated four cases. Future research will want to explore other regulatory interventions including privacy legislation, new data security legislation, changes to data breach reporting requirements, and other state-level laws. Further, this research should inform researchers globally to look at the effects of national and provincial legislation elsewhere. GDPR presents a particularly promising comparative data source. Further, this research may start to attempt to connect its findings with other data sources. Much cyber activity is not necessarily captured in these findings. For example, what is the role of vulnerabilities in tracking with increased data breach activity? Has the rise or fall of alternative hacking activities like ransomware or DDoS techniques reduced the likelihood of breaches? Perhaps these techniques which may not require breach reporting, provide an alternative source of revenue for hackers. I'm particularly interested in seeing if the 2016 plateau in reported breaches may be attributed to the rise of ransomware. Additional data is also present downstream of breaches, how can these findings be integrated with reports of cybercrime to the Federal Bureau of Investigation (FBI) at ic3.gov or reports of identity theft to the FTC.

Beyond macro empirical findings, there is a lot that can be done to start identifying important micro firm-level information. What are the costs for compliance with these regulations in an average firm? How sensitive are firms to these regulatory changes? How do CISO's leverage regulatory interventions to justify new security investments? This micro-level insight might be particularly useful at parsing whether new security investments decrease the frequency of incidents or do these new investments lead to the discovery of more incidents. If both, how can we better understand this moving variable?

The cybersecurity domain is constantly evolving as hackers and information security professionals adapt reflexively. This historical evolution presents ongoing challenges with claims of the generalizability of cybersecurity regulations the further we get from the initial observed effect. Quasi-experiments may help to identify those interventions with significant effect, but an ongoing need for future research will be to identify the mechanism involved. This is critically important to understanding how generalizable these findings are to other sectors, states, or nations. Cory Doctorow, quoting his friend Katherine Myronuk, says “All complex ecosystems have parasites” (Doctorow 2014). The cyber domain will continue to evolve and be subject to criminality, yet sensible policy might just create barriers to this criminality that can increase Internet user’s confidence and trust in the network.

Most importantly, there is immediate policy relevance to this research. This work speaks to the broader efficacy of regulatory incentives in the cybersecurity domain. Increasingly, as the technology sector has grown and become more concentrated, the public demands for regulatory oversight have grown as well. This research shows demonstrable evidence that some, but not all, information security mandates can decrease breach reporting. This presents preliminary evidence for further expansion of interventions like the NY DFS regulations. This has already begun to occur at the state-level in 2019, Connecticut passed cybersecurity regulations for their Insurance sector modeled on the NY DFS regulations scheduled to take effect on October 1, 2020. The mixed experience of regulatory efforts should suggest some caution before the United States moves towards a national standard, but should also increase state-level experimentation when it can be paired with quality data collection. To this end, the cybersecurity domain should normalize policy evaluation as part of a broader move towards sector maturity. As Charles Babbage said, “errors using inadequate data are much less than those using no data at all.”

# Appendices

# APPENDIX A
# INTRODUCTION AND BACKGROUND

## A.1 Relevant Data Security Legislation

Table A.1 was created using data from the National Conference of State Legislatures, which has monitored and tracked Security Breach Legislation since 2010.

Table A.1: Security Breach Notification Laws Passed in States by Year

| State | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| Alabama | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Arizona | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 |
| Arkansas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| California | 0 | 1 | 0 | 0 | 3 | 2 | 1 | 0 | 1 | 1 | 0 |
| Colorado | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Connecticut | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Delaware | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| District of Columbia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Florida | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 0 |
| Georgia | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hawaii | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Idaho | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Illinois | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 |
| Iowa | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Kansas | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Kentucky | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| Louisiana | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| Maine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Maryland | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 3 | 0 |
| Massachusetts | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Michigan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Minnesota | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mississippi | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Missouri | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Montana | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| Nebraska | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Nevada | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| New Hampshire | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 |
| New Jersey | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| New Mexico | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| New York | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 |
| North Carolina | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| North Dakota | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 |
| Ohio | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Oklahoma | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Oregon | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| Rhode Island | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| South Carolina | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| South Dakota | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Tennessee | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| Texas | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| Utah | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| Vermont | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Virginia | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 0 |
| Washington | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 2 |
| West Virginia | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Wyoming | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 0 |

# APPENDIX B
# METHODS

## B.1 Comparative Interrupted Time Series

To produce the comparative ITS analysis described in chapter 4, I developed a novel R code based on the work of Caswell (2018), who produced a SAS macro for the same function. This code and the output it produced are included here. Table B.1 shows a sample data frame used for the segmented OLS regression. The rate information is pulled from the Massachusetts Data Security case.

Listing B.1: Code Used to Produce Comparative ITS

```
quasiexp <- experiment[experiment$type != ”test”,]

# Added dummy variables for ITS
treatment <- as.data.frame(t(rbind(quasiexp$yearmonth,
                                    quasiexp$frequency.x)))

treatment <- treatment %>%
  mutate(Ti = as.vector(1:nrow(treatment))) %>%
  mutate(X = c(rep(0,month_n),rep(1,month_n))) %>%
  mutate(TiX = c(rep(0,month_n),1:(nrow(treatment)-month_n))) %>%
  mutate(Z = c(rep(1,month_n),rep(1,month_n))) %>%
  mutate(ZTi = Z * Ti) %>%
  mutate(ZX = Z * X) %>%
  mutate(ZTiX = c(rep(0,month_n),1:(nrow(treatment)-month_n)))

control <- as.data.frame(t(rbind(quasiexp$yearmonth,
                                  quasiexp$frequency.y)))
```

```
control <- control %>%
  mutate(Ti = as.vector(1:nrow(control))) %>%
  mutate(X = c(rep(0,month_n),rep(1,month_n))) %>%
  mutate(TiX = c(rep(0,month_n),1:(nrow(control)-month_n))) %>%
  mutate(Z = c(rep(0,month_n),rep(0,month_n))) %>%
  mutate(ZTi = Z * Ti ) %>%
  mutate(ZX = Z * X) %>%
  mutate(ZTiX = Z * Ti * X)

AppendITS <- rbind(treatment,control)

AppendITS[2:ncol(AppendITS)] <- lapply(AppendITS[2:ncol(AppendITS)],
                                       as.numeric)
names(AppendITS) <- c(”yearmonth”,”incident_permil”,”Ti”,”X”,
                      ”TiX”,”Z”,”ZTi”,”ZX”,”ZTiX”)

regTest <- lm(incident_permil ~ Ti + X + TiX + Z + ZTi + ZX +
               ZTiX, AppendITS)
```

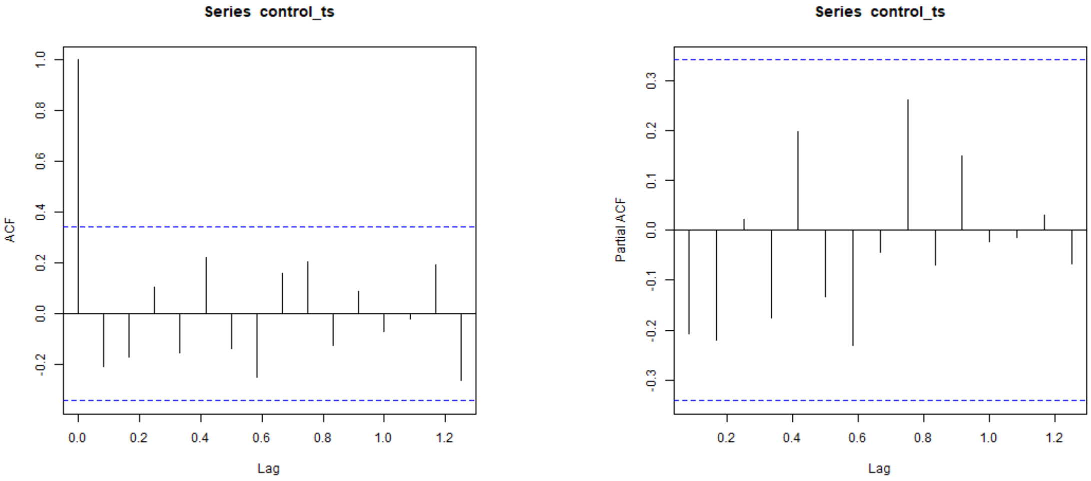
Table B.1: Sample Data Frame Used for Comparative ITS

| Description | Rate | T | X | TX | Z | ZT | ZX | ZTX |
|---|---|---|---|---|---|---|---|---|
| Treatment, Pre-Test Month 1 |  | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| Treatment, Pre-Test Month 2 |  | 2 | 0 | 0 | 1 | 2 | 0 | 0 |
| Treatment, Pre-Test Month 3 |  | 3 | 0 | 0 | 1 | 3 | 0 | 0 |
| Treatment, Pre-Test Month 4 |  | 4 | 0 | 0 | 1 | 4 | 0 | 0 |
| Treatment, Pre-Test Month 5 |  | 5 | 0 | 0 | 1 | 5 | 0 | 0 |
| Treatment, Pre-Test Month 6 |  | 6 | 0 | 0 | 1 | 6 | 0 | 0 |
| Treatment, Pre-Test Month 7 |  | 7 | 0 | 0 | 1 | 7 | 0 | 0 |
| Treatment, Pre-Test Month 8 |  | 8 | 0 | 0 | 1 | 8 | 0 | 0 |
| Treatment, Post-Test Month 1 |  | 9 | 1 | 1 | 1 | 9 | 1 | 1 |
| Treatment, Post-Test Month 2 |  | 10 | 1 | 2 | 1 | 10 | 1 | 2 |
| Treatment, Post-Test Month 3 |  | 11 | 1 | 3 | 1 | 11 | 1 | 3 |
| Treatment, Post-Test Month 4 |  | 12 | 1 | 4 | 1 | 12 | 1 | 4 |
| Treatment, Post-Test Month 5 |  | 13 | 1 | 5 | 1 | 13 | 1 | 5 |
| Treatment, Post-Test Month 6 |  | 14 | 1 | 6 | 1 | 14 | 1 | 6 |
| Treatment, Post-Test Month 7 |  | 15 | 1 | 7 | 1 | 15 | 1 | 7 |
| Treatment, Post-Test Month 8 |  | 16 | 1 | 8 | 1 | 16 | 1 | 8 |
| Control, Pre-Test Month 1 |  | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 2 |  | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 3 |  | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 4 |  | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 5 |  | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 6 |  | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 7 |  | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Pre-Test Month 8 |  | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 1 |  | 9 | 1 | 1 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 2 |  | 10 | 1 | 2 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 3 |  | 11 | 1 | 3 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 4 |  | 12 | 1 | 4 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 5 |  | 13 | 1 | 5 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 6 |  | 14 | 1 | 6 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 7 |  | 15 | 1 | 7 | 0 | 0 | 0 | 0 |
| Control, Post-Test Month 8 |  | 16 | 1 | 8 | 0 | 0 | 0 | 0 |

# APPENDIX C
# FINDINGS

## C.1 Autocorrelation and Partial Autocorrelation

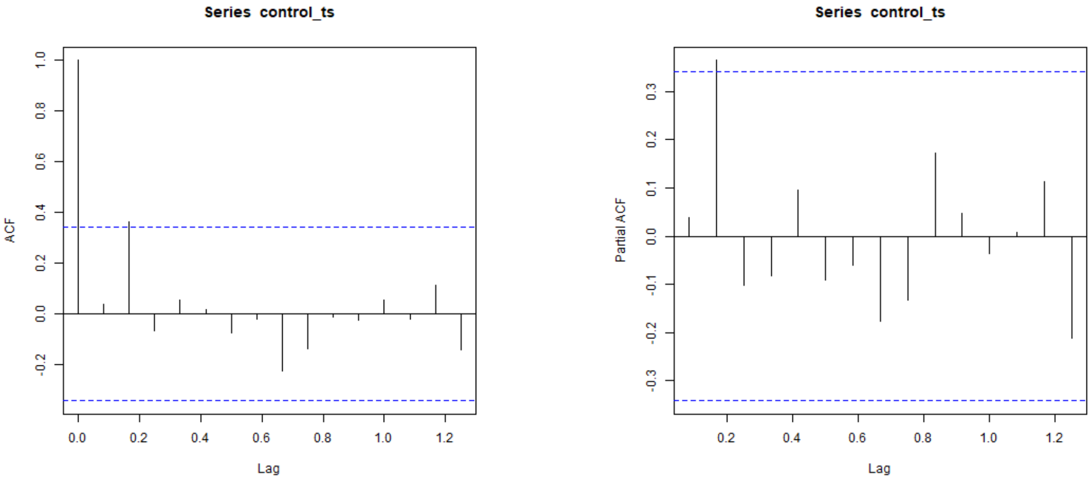
One serious issue with time series analysis is the presence of autocorrelation and partial autocorrelation. Autocorrelation indicates when the level of one observation is dependent on the observation of another point in time. The following plots labeled as (a) were created with the R function `ACF` to show the level of autocorrelation at different lags for the specified time series. Those plots labeled (b) show the partial autocorrelation, created with the R function `PACF`, which shows the relationship of an observation to a prior term in the time series after removing the relationships of prior intervening observations.



(a)

(b)

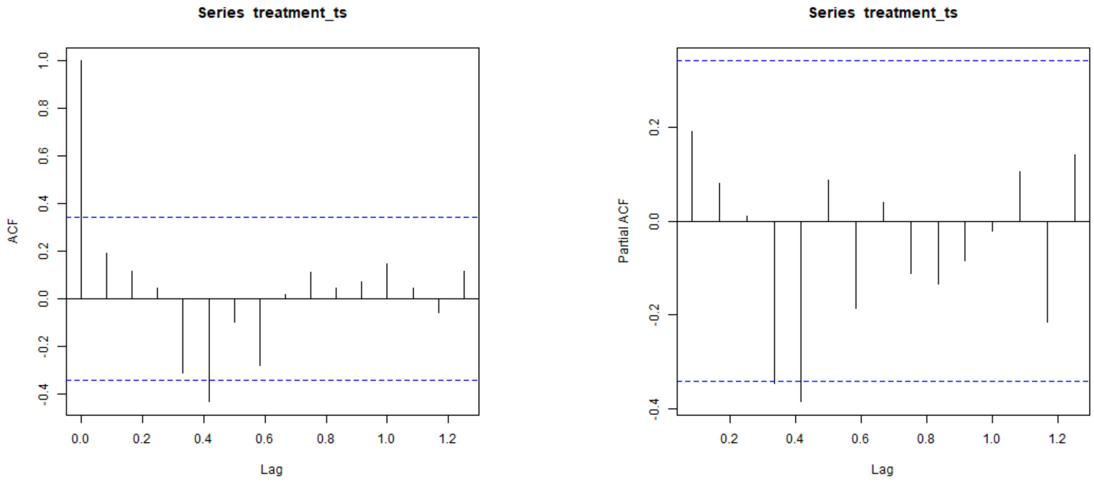
Figure C.1: Tests on Massachusetts (No Resident Limit) for (a) ACF (b) PACF



(a)

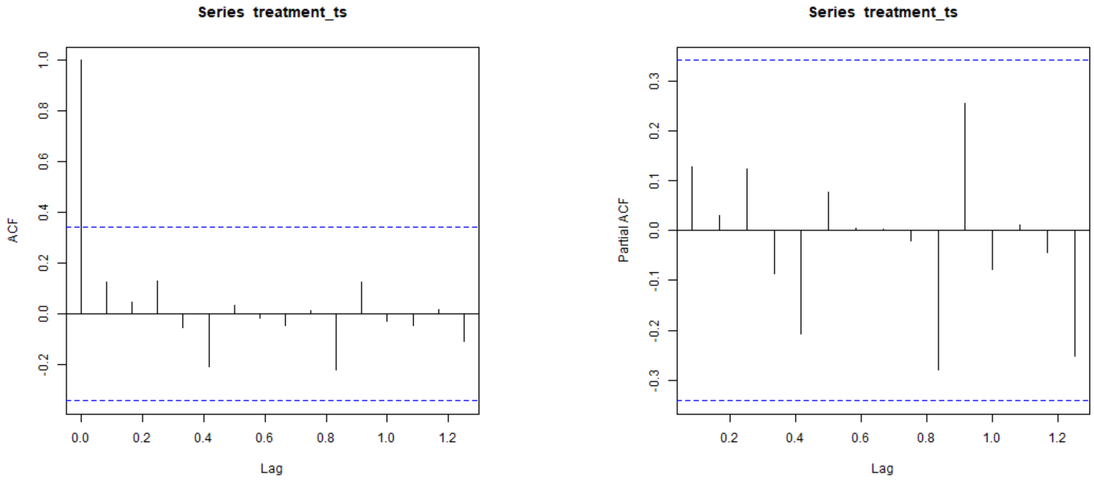
(b)

Figure C.2: Tests on Massachusetts (1000+ affected) for (a) ACF (b) PACF



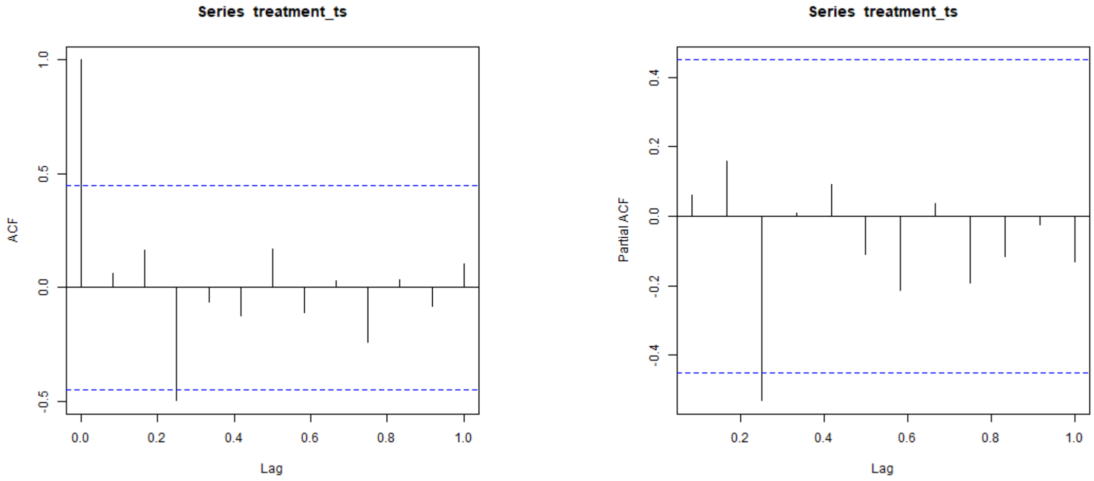
(a) (b)

Figure C.3: Tests on New Hampshire Incidents for (a) ACF (b) PACF



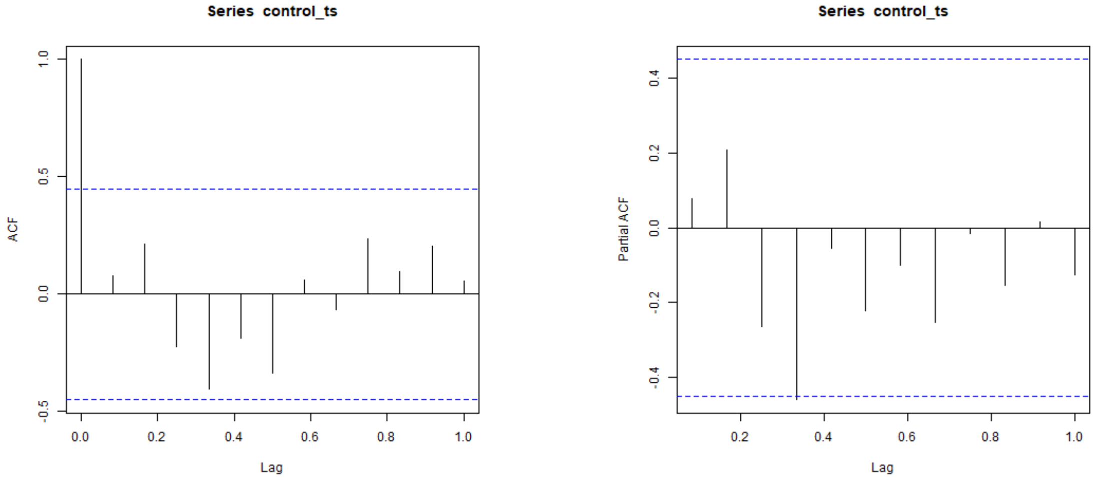
(a) (b)

Figure C.4: Tests on North Carolina (1000+ affected) for (a) ACF (b) PACF



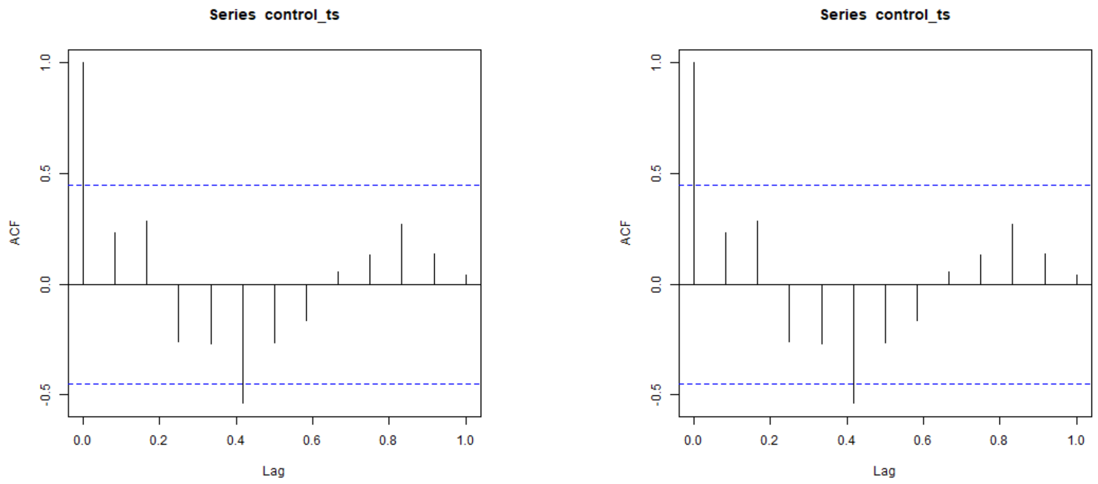
(a) (b)

Figure C.5: Tests on Health Related Incidents for (a) ACF (b) PACF



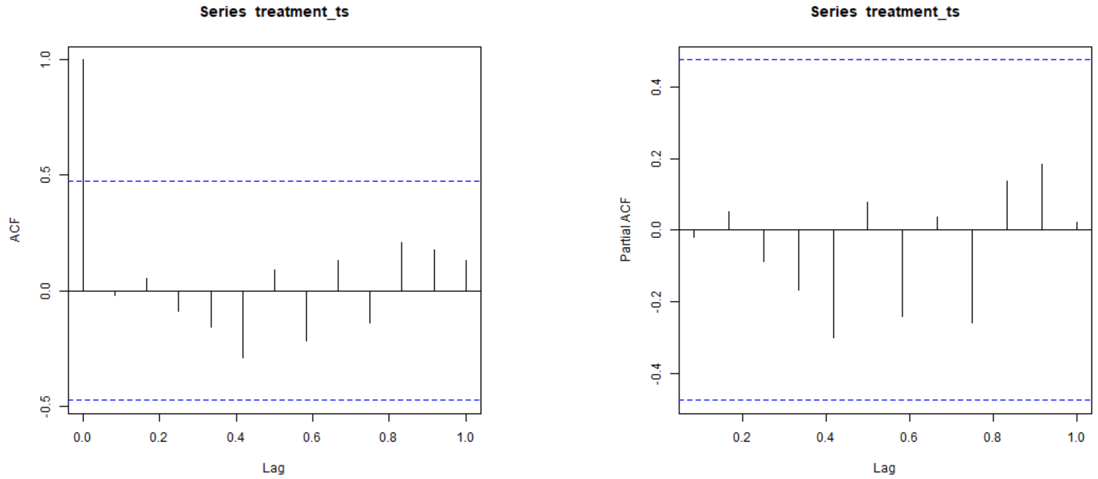
(a) (b)

Figure C.6: Tests on Non-Health Related Incidents for (a) ACF (b) PACF



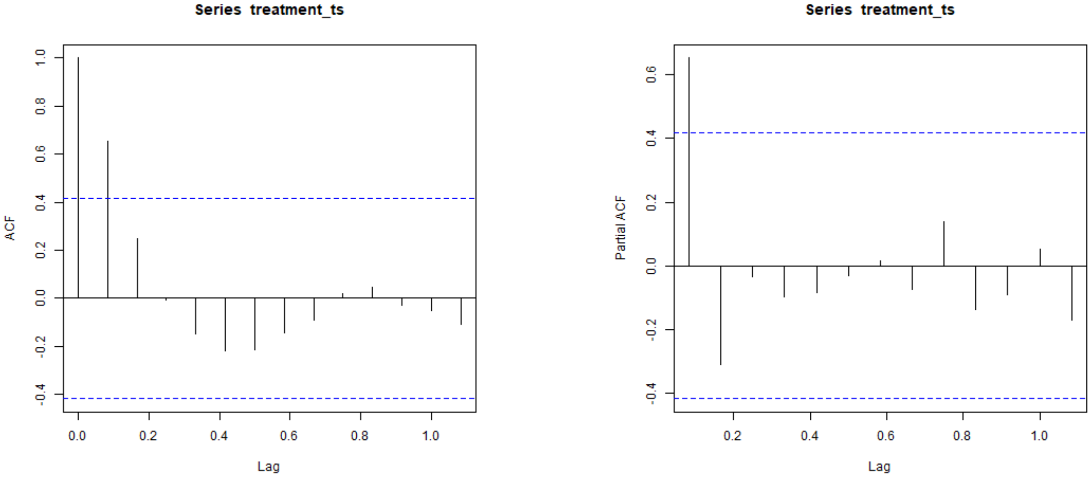
(a) (b)

Figure C.7: Tests on Finance Related Incidents as part of HITECH Act for (a) ACF (b) PACF



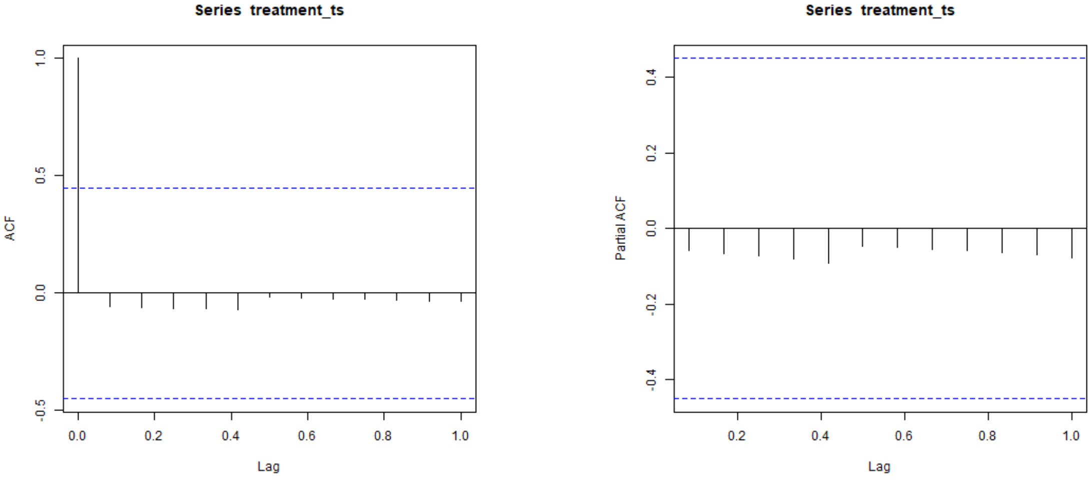
(a) (b)

Figure C.8: Tests on All Incidents Across Six States (Used in Wyndham Case - July 2011 to November 2012) for (a) ACF (b) PACF



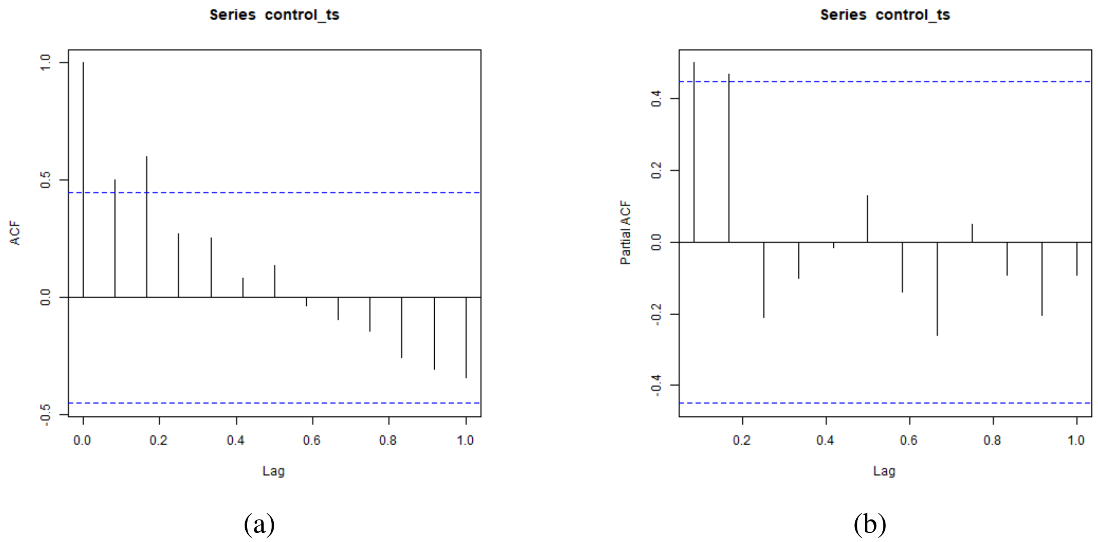
(a) (b)

Figure C.9: Tests on All Incidents Across Six States (Used in Wyndham Case - Oct 2014 to June 2016) for (a) ACF (b) PACF



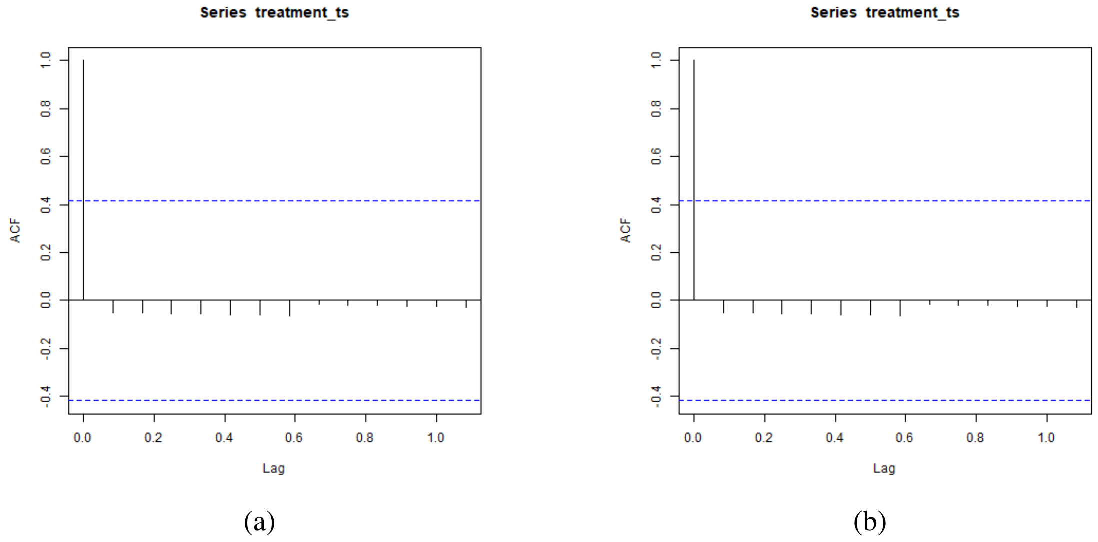
(a) (b)

Figure C.10: Tests on Finance Incidents in New York for (a) ACF (b) PACF



(a) (b)

Figure C.11: Tests on Finance Incidents Outside of New York for (a) ACF (b) PACF



(a) (b)

Figure C.12: Tests on All Non-Finance Incidents in New York for (a) ACF (b) PACF

# APPENDIX D
# AVAILABILITY OF CODE AND DATA

All relevant project code and data files have been uploaded to GitHub were they are accessible to researchers at: https://github.com/kgrindal/Dissertation_Code. The files uploaded to the GitHub repository are structured as an R Studio project, which can be downloaded in its entirety (over 200 MB). Where possible, file referencing was made easier by employing the `here` package which allows for easy referencing within the existing file structure (Müller and Bryan 2020). The project is structured around three folders: Data, Output, and Scripts. The data folder contains relevant tabular data stored as .csv, .txt, .xlsx files. The output folder is used as a destination folder to export output files. The Scripts folder contains R and R Markdown (.Rmd) files that contain the bulk of the analytical work. To give context to the function of this coding work, Table D.1 references a subset of available files and their intended function.

Table D.1: Function of code used in dissertation research

| File Names | Function |
|---|---|
| AllSource_BreachData.Rmd | Merges state data with breach metadata associated with HHS and Clearinghouse |
| Breach_Laws.Rmd | Analysis of data breach laws and bills |
| Cleaning_RawBreachData.Rmd | Primary document for cleaning and integrating state data breach reports |
| Descriptive_State_Breaches.Rmd | Identified the frequency of data breaches for each state and decomposes breach data to find seasonal patterns |
| Descriptive_Stats.Rmd | Script for identifying the overall distribution of data breach total size (i.e., number affected) |
| FTC_CaseStudy.Rmd | Identify FTC data breach cases within state breach reports |
| FTC_Wyndham.Rmd | Analysis of FTC Wyndham case using segmented regression of a comparative interrupted time series |
| HITECH_Act.Rmd | Analysis of HITECH Act case using segmented regression of a comparative interrupted time series |
| Mass_Law.Rmd | Analysis of the Massachusetts Data Security Law case using segmented regression of a comparative interrupted time series |
| NewYorkFS.Rmd | Analysis of the NY DFS cybersecurity regulation case using segmented regression of a comparative interrupted time series |
| Ny_Descriptive.Rmd | Assess distribution of breach incidents reported in Maine |
| State_Pop.Rmd | Imputing monthly population data for US states using US Census records |

# BIBLIOGRAPHY

*Federal Register*. 2002. "16 CFR Part 314 Standards for Safeguarding Customer Information; Final Rule "67, no. 100 (May 23, 2002): 36484–36494. Accessed February 10, 2020. https://www.ftc.gov/sites/default/files/documents/federal_register_notices/standards-safeguarding-customer-information-16-cfr-part-314/020523standardsforsafeguardingcustor pdf.

201 CMR 17.00: Standards for the Protection of Personal Information of Residents of the Commonwealth. 2020. Accessed February 10, 2020. https://www.mass.gov/doc/201-cmr-17-standards-for-the-protection-of-personal-information-of-residents-of-the/download.

2019. "2018 Security Breach Legislation." National Conference of State Legislatures, February 8, 2019. Accessed April 27, 2021. https://www.ncsl.org/research/telecommunications-and-information-technology/2018-security-breach-legislation.aspx.

*Federal Register*. 2009. "45 CFR Parts 160 and 164 Breach Notification for Unsecured Protected Health Informati "74, no. 162 (August 24, 2009): 42740–42770. Accessed July 14, 2021. https://www.govinfo.gov/content/pkg/FR-2009-08-24/pdf/E9-20169.pdf.

———. 2003. "45 CFR Parts 160, 162, and 164 Health Insurance Reform: Security Standards; Final Rule "68, no. 34 (February 20, 2003): 8334–8381. Accessed April 27, 2021. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/securityrule/securityrulepdf.pdf?language=es.

*A Budget for A Better America, Fiscal Year 2020*. 2020, 305–310. Analytical Perspectives. Washington D.C. Accessed February 10, 2020. https://www.whitehouse.gov/wp-content/uploads/2019/03/spec-fy2020.pdf.

*Actions by Attorney General; Notice; Venue; Injunctions*. 2020. Accessed February 10, 2020. https://malegislature.gov/Laws/GeneralLaws/PartI/TitleXV/Chapter93A/Section4.

Agranoff, Robert, and Beryl A. Radin. 1991. "The Comparative Case Study Approach in Public Administration." Edited by James L Perry. In *Research in Public Administration: A Research Annual.* 203–231. Vol. 1. N.p.: JAI Press, Inc.

*An Act Relative to Consumer Protection from Security Breaches*. 2018 H.4806. July 24, 2018. Accessed February 10, 2020. https://malegislature.gov/Bills/190/H4806.

*An Act Relative to Security Freezes and Notification of Data Breaches*. 2007. August 2, 2007. https://malegislature.gov/Laws/SessionLaws/Acts/2007/Chapter82.

*An Act to Amend, Renumber, and Add Section 1798.82 of, and to Add Section 1798.29 to, the Civil Code, Relating to Personal Information.* 2002. September 25, 2002.

*An Act to Enhance Protections Against Identity Theft and to Protect the Credit of Crime Victims Compensation Fund Applications and Appeals*. 2021 Senate Bill 1017. Accessed March 3, 2021. https://www.ncleg.net/Sessions/2009/Bills/Senate/PDF/S1017v7.pdf.

Anderson, R., and T. Moore. 2006. "The Economics of Information Security." *Science* 314, no. 5799 (October 27, 2006): 610–613. Accessed April 12, 2020. https://www.sciencemag.org/lookup/doi/10.1126/science.1130992. 10.1126/science.1130992.

Anderson, Ross, Chris Barton, Rainer Böhme, Richard Clayton, Michel J. G. van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. 2013. "Measuring the Cost of Cybercrime." Edited by Rainer Böhme. In *The Economics of Information Security and Privacy*, 265–300. Berlin, Heidelberg: Springer Berlin Heidelberg. Accessed April 24, 2021. http://link.springer.com/10.1007/978-3-642-39498-0_12. 10.1007/978-3-642-39498-0_12.

Andrew Fielder, Emmanouil Panaousis, Pasquale Malacaria, Chris Hankin, and Fabrizio Smeraldi. 2014. "Game Theory Meets Information Security Management." In *29th IFIP TC 11 International Conference, SEC 2014*, 15–29. Marrakech, Morocco.

Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019 (NST-EST2019-01). 2019. Accessed May 1, 2021. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

*Annual Report to Congress on Breaches of Unsecured Protected Health Information: For Calendar Years 2011 and 2012*. 2015, 1–29 HHS-0945–F-9452. Washington, D.C. Accessed April 27, 2021. https://web.archive.org/web/20150326145755/https://www.hhs.gov/ocr/privacy/hipaa/administrative/breachnotificationrule/breachreport2011-2012.pdf.

*Annual Report to Congress on HIPAA Privacy, Security, and Breach Notification Rule Compliance*. 2021, 24 HHS-0945-1905-F-1431. Washington, D.C. Accessed April 27, 2021. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/enforcement/compliancereport2011-2012.pdf.

Arnold, Chris, and Michael Costello, eds. 2007. *Bill Would Tie Retailers to Costs of ID Theft*. Boston, MA: NPR News. Accessed July 16, 2021. https://www.npr.org/templates/story/story.php?storyId=7599116.

Asghari, Hadi, Michel van Eeten, and Johannes M. Bauer. 2016. "Chapter 13: Economics of Cybersecurity." Edited by Johannes Bauer and Michael Latzer. In *Handbook on the Economics of the Internet*: Edward Elgar Publishing. Accessed April 12, 2020. http://www.elgaronline.com/view/9780857939845.xml. 10.4337/9780857939852.

Bai, Ge, John (Xuefeng) Jiang, and Renee Flasher. 2017. "Hospital Risk of Data Breaches." *JAMA Internal Medicine* 177, no. 6 (June 1, 2017): 878. Accessed February 10, 2020. http://archinte.jamanetwork.com/article.aspx?doi=10.1001/jamainternmed.2017.0336. 10.1001/jamainternmed.2017.0336.

Baldwin, Robert, Martin Cave, and Martin Lodge. 2012. *Understanding Regulation: Theory, Strategy, and Practice*. N.p.: OUP Oxford.

Bauer, Johannes M., and Michel J.G. van Eeten. 2009. “Cybersecurity: Stakeholder Incentives, Externalities, and Policy Options.” *Telecommunications Policy* 33, nos. 10-11 (November): 706–719. Accessed April 12, 2020. https://linkinghub.elsevier.com/retrieve/pii/S0308596109000986. 10.1016/j.telpol.2009.09.001.

Benjamin Dean. 2016. “Natural and Quasi-Natural Experiments to Evaluate Cybersecurity Policies.” *Journal of International Affairs* Vol. 70, No. 1 (The Cyber Issue (Winter 2016)): 129–160.

Benkler, Yochai. 2002. “Coase’s Penguin, or, Linux and ”The Nature of the Firm”.” *The Yale Law Journal* 112, no. 3 (December): 369. 10.2307/1562247.

Blumenthal, David, and Georgina Verdugo. 2021. Building Trust in Health Information Exchange: Statement on Privacy and Security. Accessed June 25, 2021. https://www.hhs.gov/sites/default/files/statement-privacy-security.pdf.

Brenner, Bill. 2008. “Why Mass. 201 CMR 17 Deadline Was Extended.” CSO Online, November 24, 2008. Accessed April 2, 2021. https://www.csoonline.com/article/2123439/why-mass--201-cmr-17-deadline-was-extended.html.

Brewer, Garry D. 1974. “The Policy Sciences Emerge: To Nurture and Structure a Discipline,” January 1, 1974: 3. Accessed April 24, 2021. https://www.rand.org/pubs/papers/P5206.html.

Campbell, Donald T., and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* 1st edition. Boston: Cengage Learning.

2020. “Cases Tagged with Data Security.” Federal Trade Commission. Accessed February 1, 2020. https://www.ftc.gov/enforcement/cases-proceedings/terms/249.

Caswell, Joseph M. 2018. “Interrupted Time Series Analysis for Single Series and Comparative Designs:Using Administrative Data for Healthcare Impact Assessment,” Toronto, ON, April. Accessed May 1, 2021. https://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Health-User-Groups/ITS_SAS.pdf.

———. 2019. “Interrupted Time Series Analysis for Single Series and Comparative Designs: A Guide for Beginners with SAS Macro.” *Academia.edu*, September. Accessed May 2, 2021. https://www.academia.edu/35275583/Interrupted_Time_Series_Analysis_for_Single_Series_and_Comparative_Designs_A_Guide_for_Beginners_with_SAS_Macro.

Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. 2004. “The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers.” *International Journal of Electronic Commerce* 9, no. 1 (October): 70–104. Accessed February 9, 2020. https://www.tandfonline.com/doi/full/10.1080/10864415.2004.11044320. 10.1080/10864415.2004.11044320.

Civil Rights, Office for. 2009. “HITECH Breach Notification Interim Final Rule.” HHS.gov, August 18, 2009. Accessed April 2, 2021. https://www.hhs.gov/hipaa/for-professionals/breach-notification/laws-regulations/final-rule-update/HITECH/index.html.

Coglianese, Cary, and Evan Mendelson. 2012. "Meta-Regulation and Self-Regulation." *The Oxford Handbook of Regulation*, January 1, 2012. 10.1093/oxfordhb/9780199560219.003.0008.

2021. "Current Employment Statistics Survey Data for New York State." New York Department of Labor. Accessed April 24, 2021. https://statistics.labor.ny.gov/cesemp.asp.

*Cybersecurity Requirements for Financial Services Companies*. 2020. Accessed November 10, 2020. https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsrf500txt.pdf.

Data Breach Notification Report, 2007. 2019. Accessed February 10, 2020. https://www.mass.gov/doc/data-breach-report-2007-0/download.

2020. "Database of Unclassified Cyber Spending." Taxpayers for Common Sense. Accessed February 10, 2020. http://cyberspending.taxpayer.net/.

Dembosky, Luke, Jeremy Feigelson, Avi Gesser, Jim Pastore, Lisa Zornberg, Zila Reyes Acosta-Grimes, Michael Bloom, Christopher S. Ford, and Mengyi Xu. 2020. "First Enforcement Action by New York DFS Under Its Cyber Rules Shows Where Companies Face Regulatory Risk – Six Quick Takeaways — Compliance and Enforcement." Program on Corporate Compliance and Enforcement at New York University School of Law, July 23, 2020. Accessed April 28, 2021. https://wp.nyu.edu/compliance_enforcement/2020/07/23/first-enforcement-action-by-new-york-dfs-under-its-cyber-rules-shows-where-companies-face-regulatory-risk-six-quick-takeaways/.

2020. "Department of Financial Services Announces Cybersecurity Charges Against A Leading Title Insurance Provider For Exposing Millions of Documents with Consumers' Personal Information." Department of Financial Services, July 22, 2020. Accessed April 28, 2021. https://www.dfs.ny.gov/reports_and_publications/press_releases/pr202007221.

2017. "DFS Continues Innovative Regulatory Initiatives with the Launch of New Online Cybersecurity Portal for Businesses Seeking to Report Cybersecurity Events in New York." Department of Financial Services, July 31, 2017. Accessed June 25, 2021. https://www.dfs.ny.gov/reports_and_publications/press_releases/pr1707311.

2021. "DFS Superintendent Lacewell Announces Cybersecurity Settlement With Licensed Insurance Company." Department of Financial Services, April 14, 2021. Accessed April 28, 2021. https://www.dfs.ny.gov/reports_and_publications/press_releases/pr202104141.

Dingell, John D. 2008. "Text - H.R.6357 - 110th Congress (2007-2008): PRO(TECH)T Act of 2008." October 3, 2008. Accessed June 29, 2021. https://www.congress.gov/bill/110th-congress/house-bill/6357/text.

Doctorow, Cory. 2014. *All Complex Ecosystems Have Parasites*. Paris, France: Feedbooks.

2020. "Download & API." CyberGreen, February 10, 2020. https://stats.cybergreen.net/download/.

Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. 2016. "Hype and Heavy Tails: A Closer Look at Data Breaches." *Journal of Cybersecurity* 2, no. 1 (December 1, 2016): 3–14. Accessed February 9, 2020. https://academic.oup.com/cybersecurity/article/2/1/3/2736315. 10.1093/cybsec/tyw003.

*Efficient, Effective, Accountable An American Budget, Fiscal Year 2019*. 2020, 273–288. Analytical Perspectives. Washington D.C. Accessed February 10, 2020. https://www.whitehouse.gov/wp-content/uploads/2018/02/spec-fy2019.pdf.

Eling, Martin, and Kwangmin Jung. 2018. "Copula Approaches for Modeling Cross-Sectional Dependence of Data Breach Losses." *Insurance: Mathematics and Economics* 82 (September): 167–180. Accessed February 9, 2020. https://linkinghub.elsevier.com/retrieve/pii/S0167668717306200. 10.1016/j.insmatheco.2018.07.003.

Eling, Martin, and Jan Wirfs. 2019. "What Are the Actual Costs of Cyber Risk Events?" *European Journal of Operational Research* 272, no. 3 (February): 1109–1119. Accessed February 9, 2020. https://linkinghub.elsevier.com/retrieve/pii/S037722171830626X. 10.1016/j.ejor.2018.07.021.

2020. "Employment by Major Industry Sector." US Bureau of Labor Statistics, September 1, 2020. Accessed March 2, 2021. https://www.bls.gov/emp/tables/employment-by-major-industry-sector.htm.

*Federal Trade Commission v. Wyndham Worldwide Corporation, et Al.* 2015 Civil Action No. 2:13-CV-01887-ES-JAD. August 24, 2015. Accessed April 30, 2021. https://www.ftc.gov/system/files/documents/cases/150824wyndhamopinion.pdf.

*Federal Trade Commission v. Wyndham Worldwide Corporation, et Al.* 2015. December 8, 2015. Accessed April 30, 2021. https://www.ftc.gov/system/files/documents/cases/151209wyndhamstipulated.pdf.

*Federal Trade Commission v. Wyndham Worldwide Corporation, et al., Stipulated Order for Injunction*. 2014 Civil Action No. 13-1887 (ES). April 7, 2014. Accessed April 30, 2021. https://www.ftc.gov/system/files/documents/cases/140407wyndhamopinion.pdf.

*Federal Trade Commission v. Wyndham Worldwide Corporation, et al., Stipulated Order for Injunction*. 2020 2:13-CV-01887-ES-JAD. Accessed February 10, 2020. https://www.ftc.gov/system/files/documents/cases/151209wyndhamstipulated.pdf.

Franke, Ulrik. 2017. "The Cyber Insurance Market in Sweden." *Computers & Security* 68 (July): 130–144. Accessed February 9, 2020. https://linkinghub.elsevier.com/retrieve/pii/S0167404817300883. 10.1016/j.cose.2017.04.010.

Friedman, Allan. 2012. "The Economics of Cybersecurity," January 31, 2012. Accessed April 12, 2020. https://digitalscholarship.unlv.edu/brookings_lectures_events/34.

*FTC vs Wyndham Worldwide Corporation First Amended Complaint for Injunctive and Other Equitable Relief*. 2012 2:12-cv-01365-PGR. August 9, 2012. Accessed July 14, 2021. https://www.ftc.gov/sites/default/files/documents/cases/2012/08/120809wyndhamcmpt.pdf.

FTC.gov. 2013. "Consumer Sentinel Network." Federal Trade Commission, July 16, 2013, 1:22 p.m. -04:00. Accessed April 21, 2021. https://www.ftc.gov/enforcement/consumer-sentinel-network.

———. 2021. "Commissioners." Federal Trade Commission, 2021. Accessed April 27, 2021. https://www.ftc.gov/about-ftc/commissioners.

Gao, Xing, Weijun Zhong, and Shue Mei. 2015. "Security Investment and Information Sharing under an Alternative Security Breach Probability Function." *Information Systems Frontiers* 17, no. 2 (April): 423–438. Accessed February 9, 2020. http://link.springer.com/10.1007/s10796-013-9411-3. 10.1007/s10796-013-9411-3.

2017. "Governor Cuomo Announces New Actions to Protect New Yorkers' Personal Information in Wake of Equifax Security Breach." Press Release, September 18, 2017. Accessed February 10, 2020. https://www.governor.ny.gov/news/governor-cuomo-announces-new-actions-protect-new-yorkers-personal-information-wake-equifax.

2016. "Governor Cuomo Announces Proposal of First-In-The-Nation Cybersecurity Regulation to Protect Consumers and Financial Institutions." Department of Financial Services, September 13, 2016. Accessed June 25, 2021. https://www.dfs.ny.gov/reports_and_publications/press_releases/pr1609131.

Governor Deval Patrick. 2021. *Identity Theft Legislation*. Governor Deval Patrick's Podcast 1. Accessed June 25, 2021. https://web.archive.org/web/20080802183429/http://www.mass.gov/Agov3/podcasts/2007-08-01_id_theft.rtf.

*Gramm-Leach-Biley Act*. 1999 106–102. November 12, 1999. Accessed November 10, 2020. https://www.congress.gov/106/plaws/publ102/PLAW-106publ102.pdf.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with Lubridate." *Journal of Statistical Software* 40, no. 1 (April 7, 2011): 1–25. Accessed April 28, 2021. https://www.jstatsoft.org/index.php/jss/article/view/v040i03. 10.18637/jss.v040.i03.

Hazlett, Thomas W. 1990. "The Rationality of U. S. Regulation of the Broadcast Spectrum." *The Journal of Law and Economics* 33, no. 1 (April): 133–175. Accessed April 22, 2020. https://www.journals.uchicago.edu/doi/10.1086/467202. 10.1086/467202.

2009. "HITECH Act Breach Notification Guidance and Request for Public Comment." hhs.gov, April 27, 2009. Accessed February 10, 2020. https://www.hhs.gov/hipaa/for-professionals/security/guidance/hitech-act-breach-notification-guidance/index.html.

Homan, Timothy R. 2021. "Biden Takes Quick Action on Cyber in First 100 Days." TheHill, April 30, 2021, 6:00 a.m. -04:00. Accessed May 2, 2021. https://thehill.com/policy/cybersecurity/551092-biden-takes-quick-action-on-cyber-in-first-100-days.

Hosp, Edward A., and Starr T. Drum. 2018. "The Alabama Data Breach Notification Act of 2018." *The Alabama Lawyer* 79, no. 5 (September): 333–340. Accessed July 9, 2021. https://www.maynardcooper.com/wp-content/uploads/2018/09/The-Alabama-Data-Breach-Notification-Act-of-2018.pdf.

*In the Matter of Tuition Options LLC and Edvantage LLC Consent Order*. 2020. Accessed February 10, 2020. https://www.dfs.ny.gov/system/files/documents/2019/08/ea190815_tuition_options.pdf.

Insurance Data Security Model Law. 2017. Accessed February 10, 2020. https://www.naic.org/store/free/MDL-668.pdf.

Judy, Henry L., Holly K. Towle, and Sean P. Mahoney. 2007. Financial Services Alert: New Massachusetts Identity Theft Prevention Law Breaks New Ground. Accessed April 27, 2021. https://studylib.net/doc/13818030/financial-services-alert-new-massachusetts-identity-theft....

Kennedy, Edward M, Christopher J Dodd, Tom Harkin, Barbara A Mikulski, Jeff Bingaman, Patty Murray, Jack Reed, et al., eds. 2009. *Investing in Health IT: A Stimulus for a Healthier America*, 60 S Hrg. 111-157. Washington, D.C. Accessed June 29, 2021. https://www.govinfo.gov/content/pkg/CHRG-111shrg46710/pdf/CHRG-111shrg46710.pdf.

Kennedy, Edward M. 2007. "S.1693 - 110th Congress (2007-2008): Wired for Health Care Quality Act." October 1, 2007. Accessed June 29, 2021. https://www.congress.gov/bill/110th-congress/senate-bill/1693.

Kesari, Aniket. 2020. *The Effect of State Data Breach Notification Laws on Medical Identity Theft* ID 3700248. Rochester, NY. Accessed April 29, 2021. https://papers.ssrn.com/abstract=3700248. 10.2139/ssrn.3700248.

Kunreuther, Howard, and Geoffrey Heal. 2003. "Interdependent Security." *Journal of Risk and Uncertainty* 26 (2/3): 231–249. Accessed April 13, 2020. http://link.springer.com/10.1023/A:1024119208153. 10.1023/A:1024119208153.

Lai, Willy. 2016. Fitting Power Law Distributions to Data. Accessed March 2, 2021. https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Willy_Lai.pdf.

Laube, Stefan, and Rainer Böhme. 2016. "The Economics of Mandatory Security Breach Reporting to Authorities." *Journal of Cybersecurity* 2, no. 1 (December): 29–41. Accessed February 9, 2020. https://academic.oup.com/cybersecurity/article-lookup/doi/10.1093/cybsec/tyw002. 10.1093/cybsec/tyw002.

Lenchik, Kostiantyn. 2016. "The Economics of Cybersecurity: A Practical Framework for Cybersecurity Investment," Norwegian University of Science and Technology, May 31, 2016. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2403055/KLenchik_2016.pdf?sequence=1&isAllowed=y.

Liu, Emily Yunfeng. 2020. "The Effect of State Characteristics and Cybercrime Legislation on Internet Crime." Honors Project, Smith College, Northampton, MA. Accessed May 2, 2021. https://scholarworks.smith.edu/theses/2242?utm_source=scholarworks.smith.edu%2Ftheses%2F2242&utm_medium=PDF&utm_campaign=PDFCoverPages.

2011. "Major Boston Restaurant Group That Failed to Secure Personal Data to Pay $110,000 Under Settlement with AG Coakley." Attorney General of Massachusetts, March 28, 2011. Accessed April 27, 2021. https://web.archive.org/web/20110402043100/https://www.mass.gov/?pageID=cagopressrelease&L=1&L0=Home&sid=Cago&b=pressrelease&f=2011_03_28_briar_group_settlement&csid=Cago.

2020. "Massachusetts Data Breach Notification Report." Mass.gov, Office of Consumer Affairs and Business Regulation. Accessed February 10, 2020. https://www.mass.gov/lists/data-breach-notification-reports.

Massachusetts Office of the Governor. 2007. 2007 Press Releases (Revised). Accessed June 25, 2021. http://digitalarchives.sec.state.ma.us/uncategorised/digitalFile_9a354973-cdd6-4a82-8bde-226617660fbc/.

Matland, Richard E. 1995. “Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation.” *Journal of Public Administration Research and Theory*, April. Accessed April 22, 2020. https://academic.oup.com/jpart/article/5/2/145/880350/Synthesizing-the-Implementation-Literature-The. 10.1093/oxfordjournals.jpart.a037242.

Mueller, Milton. 1999. “Digital Convergence and Its Consequences.” *Javnost - The Public* 6, no. 3 (January): 11–27. Accessed April 22, 2020. http://www.tandfonline.com/doi/abs/10.1080/13183222.1999.11008716. 10.1080/13183222.1999.11008716.

Müller, Kirill, and Jennifer Bryan. 2020. *Here: A Simpler Way to Find Your Files*. December 13, 2020. Accessed April 30, 2021. 1.0.1. https://CRAN.R-project.org/package=here.

n.d. “National Vulnerabilities Database.” National Institute of Standards and Technology, Dept. of Commerce. https://nvd.nist.gov/vuln-metrics/cvss.

*New York Banking Law, Sec. 44 Violations; Penalties*. 2020. Accessed February 10, 2020. https://newyork.public.law/laws/n.y._banking_law_section_44.

New York State Department of Financial Services. 2018. “DFS Launches Online Registration Form for Credit Reporting Agencies to Protect the Private Information of New Yorkers.” Press Release, August 22, 2018. Accessed February 10, 2020. https://www.dfs.ny.gov/about/press/pr1808221.htm.

*New York State Department of Financial Services, In the Matter of Residential Mortgage Inc.* 2021. March 3, 2021. Accessed April 28, 2021. https://www.dfs.ny.gov/system/files/documents/2021/03/ea20210303_residential_mortgage_0.pdf.

2019. “NY State Senate Bill S5575B.” NY State Senate, June 14, 2019, 12:55 p.m. -04:00. Accessed April 28, 2021. https://www.nysenate.gov/legislation/bills/2019/s5575/amendment/b.

2017. “NYC NYDFS 23 NYCRR 500 Cybersecurity Event A Big Success.” Fasoo, May 22, 2017. Accessed April 29, 2021. https://en.fasoo.com/nyc-nydfs-23-nycrr-500-cybersecurity-event-a-big-success/.

OCABR. 2009. “Fiscal Effect and Small Business Impact Statement.” Mass.gov, Office of Consumer Affairs and Business Regulation, April 1, 2009. Accessed April 21, 2021. https://web.archive.org/web/20090401001442/http://www.mass.gov/?pageID=ocaterminal&L=3&L0=Home&L1=Business&L2=Identity+Theft&sid=Eoca&b=terminalcontent&f=idtheft_sbimpact&csid=Eoca.

Office of Consumer Affairs and Business Regulation. 2017. 201 CMR 17.00: Standards for the Protection of Personal Information of MA Residents. Accessed June 25, 2021. https://www.mass.gov/regulations/201-CMR-1700-standards-for-the-protection-of-personal-information-of-ma-residents.

Parker, Geoffrey G., and Marshall W. Van Alstyne. 2005. "Two-Sided Network Effects: A Theory of Information Product Design." *Management Science* 51, no. 10 (October): 1494–1504. Accessed April 22, 2020. http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1050.0400. 10.1287/mnsc.1050.0400.

*Part 312—Children's Online Privacy Protection Rule*. 2013. January 17, 2013. Accessed February 10, 2020. https://www.ecfr.gov/cgi-bin/text-idx?SID=4939e77c77a1a1a08c1cbf905fc4b409& node=16%3A1.0.1.3.36&rgn=div5.

*Part 681—Identity Theft Rules*. 2007. November 9, 2007. Accessed February 10, 2020. https://www.ecfr.gov/cgi-bin/text-idx?SID=fddfe88d36b1e7881a1b76f4e8437d65&mc=true&node=pt16.1.681&rgn=div5#se16.1.681_11.

Penfold, Robert B., and Fang Zhang. 2013. "Use of Interrupted Time Series Analysis in Evaluating Health Care Quality Improvements." *Academic Pediatrics* 13, no. 6 (November): S38–S44. Accessed May 1, 2021. https://linkinghub.elsevier.com/retrieve/pii/S1876285913002106. 10.1016/j.acap.2013.08.002.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Accessed May 1, 2021. https://www.R-project.org/.

2020. "Red Flags Rule." Federal Trade Commission. Accessed February 10, 2020. https://www.ftc.gov/tips-advice/business-center/privacy-and-security/red-flags-rule.

Redhead, C Stephen. 2009. *The Health Information Technology for Economic and Clinical Health (HITECH) Act*, 32. CRS Report R40161. April 27, 2009. Accessed June 29, 2021. https://crsreports.congress.gov/product/pdf/R/R40161/9.

*Regulations to Safeguard Personal Information of Commonwealth Residents*. 2020. Accessed February 10, 2020. https://malegislature.gov/Laws/GeneralLaws/PartI/TitleXV/Chapter93H/Section2.

Reinke, Jennifer L. 2006. "New York's Information Security Breach and Notification Act." *NYSBA/MLRC Municipal Lawyer* 20 (1): 19–21. Accessed April 28, 2021. https://nysba.org/NYSBA/Publications/Section%20Publications/Local%20and%20State%20Government%20Law/PastIssues/Winter2006/Winter2006Assets/MuniNewsWin06.pdf.

*Report on Cyber Security in the Banking Sector*. 2014, 1–12. New York, NY. Accessed June 25, 2021. https://www.dfs.ny.gov/system/files/documents/2020/03/dfs_cyber_banking_rpt_052014.pdf.

*Report on Cyber Security in the Insurance Sector*. 2015, 1–14. New York, NY. Accessed July 9, 2021. https://www.dfs.ny.gov/system/files/documents/2020/03/dfs_cyber_insurance_rpt_022015.pdf.

Romanosky, Sasha. 2016. "Examining the Costs and Causes of Cyber Incidents." *Journal of Cybersecurity*, August 25, 2016: tyw001. Accessed February 9, 2020. https://academic.oup.com/cybersecurity/article-lookup/doi/10.1093/cybsec/tyw001. 10.1093/cybsec/tyw001.

Romanosky, Sasha, Lilian Ablon, Andreas Kuehn, and Therese Jones. 2017. "Content Analysis of Cyber Insurance Policies: How Do Carriers Write Policies and Price Cyber Risk?" *SSRN Electronic Journal*, 2017. Accessed February 9, 2020. https://www.ssrn.com/abstract=2929137. 10.2139/ssrn.2929137.

Romanosky, Sasha, David Hoffman, and Alessandro Acquisti. 2014. "Empirical Analysis of Data Breach Litigation: Empirical Analysis of Data Breach Litigation." *Journal of Empirical Legal Studies* 11, no. 1 (March): 74–104. Accessed February 9, 2020. http://doi.wiley.com/10.1111/jels.12035. 10.1111/jels.12035.

Romanosky, Sasha, Rahul Telang, and Alessandro Acquisti. 2011. "Do Data Breach Disclosure Laws Reduce Identity Theft?: Do Data Breach Disclosure Laws Reduce Identity Theft?" *Journal of Policy Analysis and Management* 30, no. 2 (March): 256–286. Accessed February 9, 2020. http://doi.wiley.com/10.1002/pam.20567. 10.1002/pam.20567.

2021. "S05827 Summary:" New York State Assembly. Accessed May 2, 2021. https://www.nyassembly.gov/leg/?bn=S05827&term=2005.

Sanches, Linda. 2021. "2012 HIPAA Privacy and Security Audits." *HHS Office for Civil Rights*:1–35. Accessed April 27, 2021. https://csrc.nist.gov/CSRC/media/Presentations/OCR-Audit-Program-2012-HIPAA-Privacy-and-Security/images-media/day2-2_lsanches_ocr-audit.pdf.

Sanches, Linda, and Verne Rinker. 2013. "Lessons Learned from OCR Privacy and Security Audits." *HHS Office for Civil Rights*, 2013:55. Accessed April 27, 2021. https://clearwatercompliance.com/wp-content/uploads/2014/06/4-1.-Lessons-Learned-from-OCR-Privacy-and-Security-Audits-Sanches_Rinker_03-07-2013.pdf.

Sarabi, Armin, Parinaz Naghizadeh, Yang Liu, and Mingyan Liu. 2016. "Risky Business: Fine-Grained Data Breach Prediction Using Business Profiles." *Journal of Cybersecurity* 2, no. 1 (December): 15–28. Accessed February 9, 2020. https://academic.oup.com/cybersecurity/article-lookup/doi/10.1093/cybsec/tyw004. 10.1093/cybsec/tyw004.

2020. "Search Results for: Data Breach." General Court of the Commonwealth of Massachusetts. Accessed February 10, 2020. https://malegislature.gov/.

Simitian, Joseph. 2009. "UCB Security Breach Notification Symposium: March 6, 2009 How a Bill Becomes a Law, Really." *Berkeley Technology Law Journal* 24 (3): 1009–1017.

Simon, F. C. 2016. *Meta-Regulation in Practice: Beyond Normative Views of Morality and Rationality*. Routledge Advances in Sociology. London ; New York, NY: Routledge, Taylor & Francis Group.

Sperry, Todd. 2012. "Feds Sue Wyndham Hotels over Repeated Computer Hacks." CNN, June 26, 2012. Accessed April 30, 2021. https://www.cnn.com/2012/06/26/travel/wyndham-hacking/index.html.

Stark, P. 2011. "Congressional Intent for the HITECH Act." *American Journal of Managed Care* 16:SP24–SP28. Accessed June 29, 2021. http://ajmc.s3.amazonaws.com/_media/_pdf/AJMC_10decHIT_Stark_SP24tp28.pdf.

The New Massachusetts Mandatory Security Regulations and Guidelines. 2008. Accessed February 10, 2020. https://bostonbar.org/pub/bw/0809/111708/sescurity_regs.pdf.

Therier, Adam. 2018. "The Pacing Problem and the Future of Technology Regulation." Mercatus Center, August 8, 2018, 1:02 p.m. -04:00. Accessed July 9, 2021. https://www.mercatus.org/bridge/commentary/pacing-problem-and-future-technology-regulation.

Tom, Willard K., Lisa Schifferle, Kristin Cohen, Kevin Moriarty, Katherine McCarron, and John Krebs. 2012. *Complaint for Injunctive and Other Equitable Relief*, 1–26. June 26, 2012. Accessed April 29, 2021. https://www.ftc.gov/sites/default/files/documents/cases/2012/06/120626wyndamhotelscmpt.pdf.

2020. "U.S. State Data Breach Lists." International Association of Privacy Professionals. Accessed February 9, 2020. https://iapp.org/resources/article/u-s-state-data-breach-lists/.

*United States of America Federal Trade Commission In the Matter of ELI LILLY and COMPANY, a Corporation. COMPLAINT*. 2020 C-4047. Accessed February 10, 2020. https://www.ftc.gov/sites/default/files/documents/cases/2002/05/elilillycmp.htm.

*United States of America v Choicepoint Inc.* 2010 1 :06-cv-0198-JTC. September 2, 2010. Accessed February 2, 2019. https://www.ftc.gov/sites/default/files/documents/cases/2006/01/0523069complaint.pdf.

Vullo, Maria T. 2017. "23 NYCRR 500 Cybersecurity Requirements for Financial Service Companies." *Official Compilation of Codes, Rules and Regulations of the State of New York* 23, no. 500 (March 1, 2017): 14.

Weatherford, Holly, Jennifer McAdam, and Chara Bradstreet. 2020. *State Legislative Brief: The NAIC Insurance Data Security Model Law*. Kansas City, MO. Accessed April 28, 2021. https://www.naic.org/documents/cmte_legislative_liaison_brief_data_security_model_law.pdf.

Wheatley, Spencer, Thomas Maillart, and Didier Sornette. 2016. "The Extreme Risk of Personal Data Breaches and the Erosion of Privacy." *The European Physical Journal B* 89, no. 1 (January). Accessed February 9, 2020. http://link.springer.com/10.1140/epjb/e2015-60754-4. 10.1140/epjb/e2015-60754-4.

Wickham, Hadley. 2007. "Reshaping Data with the Reshape Package." *Journal of Statistical Software* 21, no. 1 (November 13, 2007): 1–20. Accessed May 2, 2021. https://www.jstatsoft.org/index.php/jss/article/view/v021i12. 10.18637/jss.v021.i12.

Wyatt, Edward. 2012. "F.T.C. Charges Hotel Group Over Data Breaches." *The New York Times: Business*, June 26, 2012. Accessed May 2, 2021. https://www.nytimes.com/2012/06/27/business/ftc-charges-wyndham-worldwide-over-data-breaches.html.

2015. "Wyndham Settles FTC Charges It Unfairly Placed Consumers' Payment Card Information At Risk." Press Releases, December 9, 2015. Accessed November 10, 2020. https://www.ftc.gov/news-events/press-releases/2015/12/wyndham-settles-ftc-charges-it-unfairly-placed-consumers-payment.

Xu, Maochao, Kristin M. Schweitzer, Raymond M. Bateman, and Shouhuai Xu. 2018. "Modeling and Predicting Cyber Hacking Breaches." *IEEE Transactions on Information Forensics and Security* 13, no. 11 (November): 2856–2871. Accessed February 9, 2020. https://ieeexplore.ieee.org/document/8360172/. 10.1109/TIFS.2018.2834227.

Yang Liu, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, and Mingyan Liu. 2015. "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents." In *Proceedings of the 24th USENIX Security Symposium*, 1009–1024. Washington, D.C. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/liu.

Yin, Robert K. 2017. *Case Study Research and Applications: Design and Methods*. 6th edition. Los Angeles: SAGE Publications, Inc.

Yixin Zou, Abraham H. Mhaidli, Austin McCall, and Florian Schaub. 2018. ""I've Got Nothing to Lose": Consumers' Risk Perceptionsand Protective Actions after the Equifax Data Breach." In *Proceedings of the Fourteenth Symposium on Usable Privacy and Security*, 197–216. Baltimore, MD.

# VITA

Karl Grindal is a policy analyst and information security researcher based in Atlanta, GA. He is a collaborator with the Internet Governance Project and will join the new School of Cybersecurity and Privacy at Georgia Institute of Technology as a Postdoctoral Fellow in the Fall of 2021. He completed his Ph.D. studies at Georgia Tech's School of Public Policy.

Karl previously served as the Director of Research for Intelligent Cyber Research (ICR), where he developed the Geocyber Risk Index (GCRI), a comparative assessment of the cyber threats of operating a network in different countries in collaboration with the Eurasia Group. Before joining ICR, he provided strategic, policy, and research services as a Senior Analyst at Delta Risk LLC.

From 2014-2017, Karl was the Executive Director of the Cyber Conflict Studies Association (CCSA), a non-profit dedicated to advancing a research agenda on cyber conflict. Earlier with CCSA, Karl had collaborated with Jason Healey as the Associate Editor to the book A Fierce Domain: Conflict in Cyberspace 1986 to 2012.

Karl completed a Master's of Public Policy with a concentration in Technology Policy from Georgetown University. He completed his undergraduate studies with a Bachelor's of Arts in Government and Certificate in International Relations from Wesleyan University.