Essays on U.S. Data Protection Law & Policy

by

Aniket Kesari

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Jurisprudence & Social Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Daniel L. Rubinfeld, Chair
Adjunct Full Professor Chris Jay Hoofnagle
Professor Sonia Katyal
Assistant Professor Avi Feller
Assistant Professor Peng Ding

Spring 2020

Essays on U.S. Data Protection Law & Policy

Abstract

Essays on U.S. Data Protection Law & Policy

by

Aniket Kesari

Doctor of Philosophy in Jurisprudence & Social Policy

University of California, Berkeley

Professor Daniel L. Rubinfeld, Chair

Privacy and cybercrime law in the United States typically focuses on disclosure and deterrence by denial, but obtaining evidence about this regime's efficacy has eluded policymakers and researchers. This dissertation evaluates various pieces of U.S. data protection law, and offers data-driven approaches to longstanding questions in the literature. Chapter 2 reframes cybercrime from a causal question to a predictive one, and presents a machine learning model that predicts which publicly traded companies are likely to suffer data breaches. Chapter 3 examines state data breach notification laws, the primary mechanism for responding to data breaches in the U.S., and offers evidence about their effect on medical identity theft rates. Chapter 4 looks at how governments, intellectual property owners, and technology companies police cybercrime by disrupting cybercriminals' access to intermediaries. Taken together, the three chapters suggest a path forward for researching and evaluating cybercrime policy in a data-driven manner.

To the memory of my father, Krishna Kesari, 1955-2012.

Contents

List of Figures

List of Tables

Acknowledgments

I owe a huge debt to my committee members: Dan Rubinfeld, Chris Hoofnagle, Sonia Katyal, Avi Feller, and Peng Ding. Dan, Avi, and Peng provided incredible advice that helped me write at the intersection of law, policy, and statistics. Chris and Sonia taught me most of what I know about privacy and cybercrime law, and actively trained me as a scholar by providing me with ample opportunities to research and co-author with them over the years.

Cathryn Carson, Bob Cooter, Justin McCrary, Kevin Quinn, and Bin Yu were also incredibly supportive throughout my graduate career. In various ways, they facilitated my intellectual shift toward law & economics and data science. I am grateful that I had the opportunity to learn from such incredible faculty.

Several organizations also provided financial and intellectual support for my endeavors over the last few years. The Google Public Policy Fellowship that funded my time at Engine, the Data Science for Social Good Fellowship at the University of Chicago, and GitHub's internship program provided rewarding summer experiences. I also gratefully acknowledge research funding from various centers at UC Berkeley including the D-Lab, the Data Science Education Program, the Law, Economics, and Politics Center, the Center for Long-Term Cybersecurity, and the Berkeley Center for Law & Technology.

Several people inspired my journey to graduate school and deserve thanks. Professors Milton Heumann, Dennis Bathory, Daniel Kelemen, and Lisa Miller at Rutgers University cultivated my interest in social science research during my undergraduate years, and they have continued to be a source of support ever since. Their warmth, wisdom, and generosity will forever be the standard I strive for with my own students. I am also thankful for my many lifelong friends from Rutgers and East Brunswick who have continued to be my confidants and supporters after so many years.

My friends and colleagues at Berkeley Law deserve special mention. In particular, I want to thank Gabe Beringer, Griffin Brunk, Ben Chen, Ryan Copus, James Dillon, Kyle DeLand, James Hicks, Melissa McCall, Julian Nyarko, Lawrence Liu, Joel Sati, and Reid Whitaker. Their friendship and insights over the years were the best parts of graduate school.

I also thank my friends who I've made over the last few years: Renata Barreto-Montenegro, Harrison Chan, Mark Patanta, Alex Settle, Carly Strauss, Thomas Ryland Rembert, Vetri Velan, Stephanie Wang, Tyler Westenbroek, and Justin Zik. Our jaunts to Cornerstone and game nights were always the highlight of my week. My puppy, Silla, has been my best friend and a constant source of pure joy throughout the dissertation writing process. My mom and brother, Anupam, always gave me a place to come home and endless support.

Chapter 1

Introduction

Data protection law is rapidly evolving at all levels of government. In recent years, New York City launched an Automated Decision Systems Task Force, California enacted the California Consumer Privacy Act (CCPA), the U.S. Congress debated multiple privacy bills, and the European Union established the General Data Protection Regulation. As governments continue to legislate in this area, an examination of the theories underlying these policies is both prudent and timely. This dissertation focuses on the regulation of private sector responses to cybercrime. A major lesson from this endeavor is that designing effective data protection policy is not an intractable problem. Although the economics of cybercrime differ from the economics of other types of crime, there are still several features of cybercrime that policymakers can exploit to effectively deter it. By orienting regulatory focus from cybercriminals to the organizations that collect and maintain consumer data, policymakers can create an effective regime for protecting privacy and minimizing harms.

1.1 Data-Driven Privacy Law Research

Empirical researchers who study privacy and cybercrime law confront the perennial lack of high-quality data. Identity theft victimization rates are inconsistently collected and reported. Companies' adherence to their own privacy policies is largely unknown. Consumer valuations of their own privacy is difficult to measure because of bounded rationality. Cybercriminals themselves are difficult to detect and identify. Although cybercrime can be costly, and clearly a priority for lawmakers, this lack of data poses challenges for conducting credible research.

One objective in these chapters is to highlight the potential for researchers to be creative in their approach to collecting and analyzing data in the privacy context, as well as legal scholarship more broadly. Empirical legal studies traditionally borrows data collection and analysis techniques from econometrics – in large part because of the movement's close alignment with Law & Economics. This tradition forms the foundation for this dissertation's theoretical and empirical approach. However, there is substantial value in borrowing frameworks from other quantitative disciplines as well. In particular, I make a case for infusing legal research with computational techniques because code enables approaches that would otherwise be difficult to manually replicate.

One example of this approach is the use of text data in legal applications. Chapter 2 draws on a combination of annual financial disclosures and firm-level financial data to generate predictions for companies that are likely to suffer data breaches or cybersecurity incidents. The premise of this effort is that publicly traded companies are required to make disclosures to their investors about their potential cybersecurity risks. The text of these disclosures is potentially ripe with information, and techniques from natural language processing (NLP) can be used to exploit that information to inform policy decisions. The use of NLP in law is a nascent research area that seeks to leverage the vast amounts of text in the legal reasoning underlying judicial decisions, legislation, regulations, and disclosures to derive new insights. By using mandatory disclosures to enhance a regulatory decisionmaking process, I illustrate one possible application of NLP that is relevant to both privacy law scholars and empirical legal researchers.

Chapter 3 further highlights the importance of expanding legal research to incorporate interdisciplinary approaches. In this chapter, I examine the effect of state data breach notification laws on medical identity theft. I collect identity theft data from the Consumer Financial Protection Bureau's Consumer Complaint Database, and analyze the effect of state-level regulations with a synthetic control. While common in economics, statistics, comparative politics, and public policy, the synthetic control has yet to be widely adapted in legal settings. Despite weaknesses relative to methods that rely on clean identification strategies, it offers advantages over purely descriptive methods or regression models in legal contexts. In particular, the ability to more transparently evaluate the effect of laws that can rarely be described as exogenous shocks gives researchers and policymakers a baseline to inform future policymaking and evaluation.

In both privacy and other legal contexts, the lower barriers to entry to utilize data science presents an exciting opportunity. Legal scholarship has long looked to cognate disciplines in the social sciences and humanities to drive new insights. As computational and statistical thinking start to transform research agendas in those fields, legal scholarship is also poised to explore new avenues. By highlighting various methods of data collection, illustrating studies in both prediction and causal inference, and connecting data science to legal questions, I provide a blueprint for other researchers interested in developing similar agendas across legal contexts.

1.2 Deterrence by Denial

Moving more specifically to lessons for U.S. data protection law, each chapter points to the conclusion that the law should focus on a deterrence by denial strategy. Deterrence theory is a common term across several disciplines and sub-disciplines. In criminology, deterrence generally refers to ”deterrence by punishment,” or the idea that crime can be deterred with sufficiently costly punishments. In international relations, deterrence theory refers to the idea that states can deter other states from engaging in certain behaviors because of disproportionately destructive threats (i.e. mutually assured destruction theory). The thread that unifies these ideas is that deterrence is achieved by making an action too costly to be worthwhile.

As detailed in each chapter, for various reasons, cybercrime is not easily deterred simply by increasing the cost associated with punishment. Cybercriminals rarely suffer punishments for perpretrating cybercrimes. Because they are difficult to find, much less prosecute, cybercriminals face low probabilities of punishment. Thus even large punishments become heavily discounted to become nearly meaningless, especially compared to the potential financial rewards that cybercriminals may reap from their activity. This low-cost/high-benefit feature of cybercrime can lead to the conclusion that cybercrime is impossible to solve with the legal tools that are conventionally used in other criminal justice applications.

However, these pieces eschew the economics of crime that typically focuses on the costs and benefits borne by the individual, and instead adopt a deterrence by denial approach that focuses on licit organizations. Rather than deter cybercrime by making the punishment for cybercriminal activity large, each chapter suggests an approach for making cybercrime impractical or financially unattractive. Each study presupposes that cybercrime is sophisticated, targeted, and financially motivated. These cybercrime features make defensive cybersecurity posturing more realistic because regulatory activity can be channeled through a handful of legitimate businesses in the U.S.

Each chapter focuses on a different type of organization to focus deterrence by denial strategies. Chapter 2 looks at publicly traded companies and conceptualizes the policy problem as a predictive one where the SEC ensures that vulnerable companies are adequately investing in cybersecurity. Chapter 3 focuses on HIPAA-covered organizations that collect and maintain health data. Chapter 4 examines the role intermediaries such as payment processors and online marketplaces play in choking off cybercriminal activity. The key takeaway here is that policymakers can effectively deter cybercrime by making it more difficult to breach and steal from legal businesses.

1.3 Overview

This dissertation proceeds as follows. Chapter 2, ”Predicting Cybersecurity Incidents Through Mandatory Disclosure Regulation,” develops a predictive model of publicly traded companies that are likely to suffer data breaches or other cybersecurity incidents. Chapter 3, ”The Effect of Data Breach Notification Laws on Medical Identity Theft,” uses a synthetic control to estimate the effect of California’s data breach notification law on medical identity theft. Chapter 4, ”Deterring Cybercrime: Focus on Intermediaries” explores how cybercrime is policed through licit intermediaries. Chapter 5 concludes by tying together each piece and summarizing key themes.

Chapter 2

Predicting Cybersecurity Incidents Through Mandatory Disclosure Regulation

2.1 Introduction

"Sunlight is said to be the best of disinfectants, electric light the most efficient policeman.",
Louis Brandeis in *Other People's Money and How the Bankers Use It* (1914)

In 1914, Louis Brandeis wrote this powerful statement in response to the emergence of consolidated banks and trusts (Brandeis, 1914). He was concerned about the power these institutions would have in American democracy, and prescribed several solutions. Among these was the notion of "sunlight as disinfectant" - that transparency and openness were effective means to regulate these large enterprises that could perpetuate a range of social ills. At the time, Brandeis called for the creation of a government agency that could force transparency and investigate wrongdoings. These ideas were the foundation that formed what became the Federal Trade Commission.

Today, large corporations deal not only with other people's money, but also their data. Over the last several decades, the U.S. adopted several data protection laws that regulate particular economic sectors that deal with especially sensitive data. Mandatory disclosures are a popular tool for encouraging good corporate behavior. "Sunlight as disinfectant" is the main theory underlying mandatory disclosure laws, and the notion is that consumers and regulators can punish corporations that engage in bad behavior. So long as there is adequate information, the public and government agencies are well positioned to prevent the types of social ills that stem from consolidated corporate power.

Cybersecurity incidents that result in the loss of consumer data, especially losses attributable to external breaches, pose a serious threat to consumer privacy. Such incidents are becoming more severe, as evidenced by recent news headlines surrounding Cambridge Analytica (Lapowsky, 2018) and the Equifax breach (Cowley, 2017). These most recent events each implicated at least 80 million records, and their unprecedented scale has prompted policymakers at both the federal and state levels to consider and pass legislation to prevent future events. Academics have seriously engaged in the theoretical, technical, economic, and policy dimensions underlying privacy and cybercrime for decades, but there remain a number of open empirical questions.

The cybercrime literature is largely concerned with detecting and deterring cybercrime. However, relatively little attention has been paid to predicting incidents of cybercrime. Compared to traditional crime, cybercrime's spatial dimensions are difficult to conceptualize, and cybercrimes are somewhat rare events. Despite this difficulty, successfully predicting cybercrime could potentially yield enormous benefits. Compensating victims of cybercrime for their losses after a breach is difficult because it is hard to measure the damage (Mayer, 2016). Finding and punishing the perpetrators of cybercrime may be nearly impossible, especially if they live in a non-U.S. jurisdiction. However, deterring cybercrime by making it economically impractical may be more effective. Much of cybercrime is financially motivated, and choking off financial incentives is a powerful way to deter it.

In this piece, I propose predicting incidents of cybercrime primarily by looking at the potential risk factors a company may exhibit. Cybercriminals may exploit vulnerabilities that can lead to massive data losses, or other catastrophic consequences. From policymakers' and law enforcement's perspective, it is difficult to identify risk factors without firms' close cooperation, which may be impractical. Developing a tool that uses publicly available information to develop cyberrisk profiles can help policymakers and auditors prioritize their regulatory activities.

I utilize the fact that the Securities and Exchange Commission (SEC) requires numerous disclosures from publicly traded companies. In particular, every publicly traded company must provide a statement of its risk factors, and financial statements detailing the overall health of the company. The goal of these disclosures is to signal potentially relevant information to investors. In addition, companies must disclose financial information about their stock performance, tax liabilities, and assets to investors and regulators. The SEC is generally interested in harnessing its massive troves of data for Artificial Intelligence applications. I propose using machine learning and Natural Language Processing (NLP) techniques to train an algorithm that predicts future cybersecurity incidents based on firm-level the text of a company's filings. If successful, this could prove to be a valuable tool for regulators as they attempt to identify risky companies, and develop interventions to prevent cybercriminals from exploiting those risks. Just as the FTC emerged in response to the growing problem of trusts, this study highlights the potential for the SEC to take on a similar role with regards to cybersecurity.

2.2 Law & Economics of Cyber Risk Disclosure

Cybersecurity regulation and auditing is an asymmetric information problem. Firms have private information about their cybersecurity posture, and this information is not available to regulators without intervention. Absent incentives to publicize this information, firms will prefer to keep this information private. In the cybersecurity context, the law has thus far addressed this problem by mandating disclosure of relevant cybersecurity risks alongside with other mandatory financial disclosures. Failure to adequately disclose this information can result in an investor lawsuit, thus imposing a cost on firms that fail to disclose risks, suffer an adverse cybersecurity event, and are subsequently sued.

From a regulator's perspective, obtaining information through these disclosures raises additional questions. Assuming the regulator is interested in obtaining the maximum amount of information about a firm's cyberrisk, it will craft disclosure requirements with an eye toward optimizing this quantity. These requirements' design is critical because firms are not likely to disclose relevant information unless there are other incentives to do so (such as signaling preparedness to regulators and investors). This simple model is the basis for the relationship between regulators and firms in a wide variety of contexts that involve audits, such as food safety inspections.

This simple model can be expanded by considering firms' own abilities to understand their cybersecurity postures. Although firms have private information about their policies, estimating cyberrisk requires domain expertise and involves uncertainty with regards to relative risk compared to similar firms. Again, because each firm's cybersecurity posture is private information, firms are unlikely to know what similarly situated firms are doing, and therefore cannot assess their own risk relative to their competitors.

There are several vendors who develop risk assessment tools that provide companies with cyberrrisk scores. For example, Security Scorecard uses a combination of information volunteered by a firm along with information scraped from a variety of security risk databases. It scores companies on ten different categories, and returns an A-F letter grade, along with access to a dashboard that helps companies pinpoint areas for improvement. Similarly, FICO offers a Cyber Risk Score service. Like its consumer credit scores, the scores seem to range between 300 and 850. Both of these services sell enterprise editions to companies and provide them with an comprehensible metric. These services therefore ameliorate the costs firms face with regards to processing their own information about their cyberrisks, and understanding their position relative to similar firms. These scores are also sold to insurers who underwrite cybersecurity incident policies, thus potentially solving the problem of distributing risk across similarly situated firms for rare events.

However, these scores remain private information for the firms in question, and therefore do not solve the problem of regulators having less information about firms' relative cyberriskiness. A tool that provides information about firms' security postures to regulators would therefore help bridge this information gap. Moreover, while these services advertise that they use machine learning tools to generate their scores, and that the scores are a direct measure of riskiness with respect to suffering a breach, the validation strategies are unclear and not publicly available. Benchmarking firms' private assessments of their riskiness against real-world outcomes and making that information available to regulators would be helpful for refining and targeting regulatory efforts such as audits. Critically, creating a risk assessment based on public data gives regulators the ability to prioritize their decisionmaking even if firms do not volunteer to assess their own cybersecurity postures.

2.3 Literature Review

There are few studies that directly predict future cybersecurity incidents. In large part, the cybercrime literature is more concerned with causation rather than prediction. This is not unique to cybercrime however, as the social sciences traditionally focus on causation. However, techniques originating from data science open up opportunities to engage in useful prediction exercises as well.

Kleinberg et. al. argue this point in ”Prediction Policy Problems.” In this paper, the authors argue that machine learning techniques do not get adequate attention in the social sciences, and in economics in particular. They make the case that social scientists frequently miss interesting prediction questions because of the traditional focus on causal inference techniques. To illustrate, they use the toy example of a doctor deciding whether to perform a hip replacement on a patient. The catch is that the hip replacement is very painful in the short term, and would only improve the patient’s qualify of life after about six months or so. Thus, it is only worthwhile to provide this treatment if the doctor can be reasonably sure that the patient will live at least that long. The prediction problem is trying to accurately predict whether a patient will live for six more months. If so, the hip replacement would be worthwhile. If not, there would be no need to put the patient through unnecessary pain (and expend the time and money on the needless procedure). To do this exercise, the decision maker does not need to know *why* the patient will live or not in the next six months, but rather simply whether or not they will. This insight is the key to understanding the motivation underlying this project (Kleinberg et al., 2015).

Susan Athey extends this discussion by discussing the intersection of machine learning, causal inference, and policy evaluation. In particular, she highlights the importance of rigorously mapping an algorithmic decision to an actual policy decision. While Kleinberg highlights a useful example of applying simple off-the-shelf methods to a problem, Athey argues that some understanding of the domain problem and causal mechanisms is still necessary for successfully implementing machine learning in policy. Pairing predictive decisions with techniques drawn from causal inference will help guide optimal policy decisionmaking (Athey, 2017)

Within the legal literature, Joshua Mitts makes a similar argument in ”Predictive Regulation.” He notes that regulatory agencies frequently design rules and interventions that respond to ”crises” or other events, but oftentimes, it would be better if these rules could be designed in a way that anticipated crises rather than correct them after the fact. He motivates this line of reasoning by pointing out that it is unlikely that the next financial crisis will be caused by subprime mortgage lending the way the 2008 crisis was. Regulators could avert the consequences of future crises by anticipating them and enacting relevant rules beforehand.

He argues that statistics provides many of the essential tools that can predict adverse events, and therefore enable policymakers to pro-actively intervene. Specifically, he demonstrates that natural language processing techniques could have flagged speculative language in the housing market before the 2008 financial crisis. He points to the potential for using these techniques across domain contexts could dramatically improve regulatory efforts (Mitts, 2014).

These papers provide a basis for exploring predictive cybersecurity policy. One need not understand the exact mechanics of the motivations of cybercrime to predict which companies are most at risk of suffering an attack. Instead, one must simply do a reasonably good job of predicting accurately, and therefore better informing decisions about where to target interventions.

The most direct study of predicting cybersecurity incidents is a paper from the University of Michigan entitled, ”Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents.” The authors in this study created an incidents database from a combination of the VERIS, Hackmageddon, and Web Hacking Incidents Database. These datasets constituted the outcome data, and they were joined with features drawn from each organization's cyber practices. These included features such as DNS misconfiguration, spam/phishing activity, etc. Overall, using a random forest algorithm, the authors report a high accuracy rate (90% True Positive, 10% False Positive) (Liu et al., 2015).

Aside from the cybercrime literature, there is a rich literature surrounding SEC disclosures. The theoretical foundations of corporate disclosure as a regulatory tool have been explored at length in the economics, business, and law literatures. These literatures ask questions about the optimal amount of disclosure to require, the incentives underlying honest and dishonest signaling in disclosure statements, and whether insiders use information to their advantage prior to a disclosure. These are all key questions that motivate the use of disclosure as a tool, and are particularly attuned to the SEC's disclosure requirements because of the high stakes involved with publicly traded firms, and the relatively consistent and stringent regulations placed on them.

Christian Leuz and Peter Wysocki provide a general overview of the various literatures in ”Economic Consequences of Financial Reporting and Disclosure Regulation: A Review and Suggestions for Future Research.” They identify a gap across the board, namely that the study of disclosure has largely focused on voluntary disclosures made by individual firms. In comparison, there is relatively little work done that studies the effects of mandatory disclosures, and how well those regulations achieve certain policy outcomes (Leuz and Wysocki, 2016). Various other studies that look at the effects of mandatory disclosure regulations generally focus on the effect of disclosure regulation and legislation on capital markets. The authors cite a number of studies that look at SEC regulations dating back to the 1930's that mainly look at how firms adjust behavior when a regulatory regime is imposed, and how capital markets react when disclosures are made. However, these studies by and large do not make extensive use of natural language processing or qualitative analysis that examines the content of the disclosures themselves, and therefore miss key questions about the relationship between the regulations, disclosure content, and outcomes.

Kogan et. al. use the text of 10-K disclosures to predict stock price volatility. In particular, they used a tf-idf featurization and support vector regression technique to predict price volatility. They find that the text of 10-K disclosures provides substantial information to make predictions about historic price volatility (Kogan et al., 2009).

Otherwise, there is also a growing interest in the use of data science, machine learning, artificial intelligence, and other quantitative methods in the SEC. In a recent statement, Scott W. Bauguess, Acting Director and Acting Chief Economist of the SEC, articulated the SEC's goals in thinking about the rise of these methods. He emphasized that SEC regulators would benefit from being able to predict likely outcomes in a range of domains, and these tools provide unprecedented potential to do so. As part of its commitment to developing such technologies, the SEC makes troves of its own data and the raw text of disclosures available on its EDGAR interface (Bauguess, 2017).

The availability of these data has encouraged some preliminary work in implementing data science approaches to regulation. Joshua Mitts and colleagues wrote a piece entitled, ”The 8-K Trading Gap” that looked at whether there was evidence of insider trading in the days preceding a damaging disclosure statement. Similarly in the cybersecurity context, Mitts and Eric Talley conducted a study that found evidence of insider trading prior to a cybersecurity breach disclosure (Mitts and Talley, 2018).

2.4 Data

Outcome Data

For outcome data, meaning reported data breaches and cybersecurity incidents, I combine several data sources that independently collect information about these events. In particular I use the Veris Community Database (VCDB) (which feeds into the Verizon Data Breach Investigations Report), the Privacy Rights Clearinghouse Chronology of Data Breaches Database, and the Identity Theft Resource Center. Each of these database maintainers collects different information and the nature of the incident. The most important distinction between these databases is the definition of breaches and incidents. Simply put, an incident can encompass a variety of events including loss of equipment, mismanagement of cybersecurity training, etc. Data breaches are one example of a cybersecurity incident. In general, companies do not always need to report incidents because they are not always material (in terms of securities regulation), but breaches are almost certainly material. Thus, the outcome data includes both breaches and material incidents, but it is important to note that these account for *reported* breaches and incidents. Because certain events, even material ones, may be unreported (and even undetected), focusing on reported breaches necessarily undercovers the universe of actual breaches and material incidents.

I use the Privacy Rights Clearinghouse (PRC) database as the baseline for data breaches and incidents, and augment the outcome data with breaches that are missing from PRC. I do this primarily because PRC collects information that is useful for sketching out the policy problem that may be missing from other databases, namely the type of incident, its location, and a description of the incidents. Figure 2.1 shows the number of breaches and incidents in the PRC dataset from 2010 onward. Note that these are breaches among publicly traded companies successfully matched in the dataset, as there are far more in the PRC database as a whole.

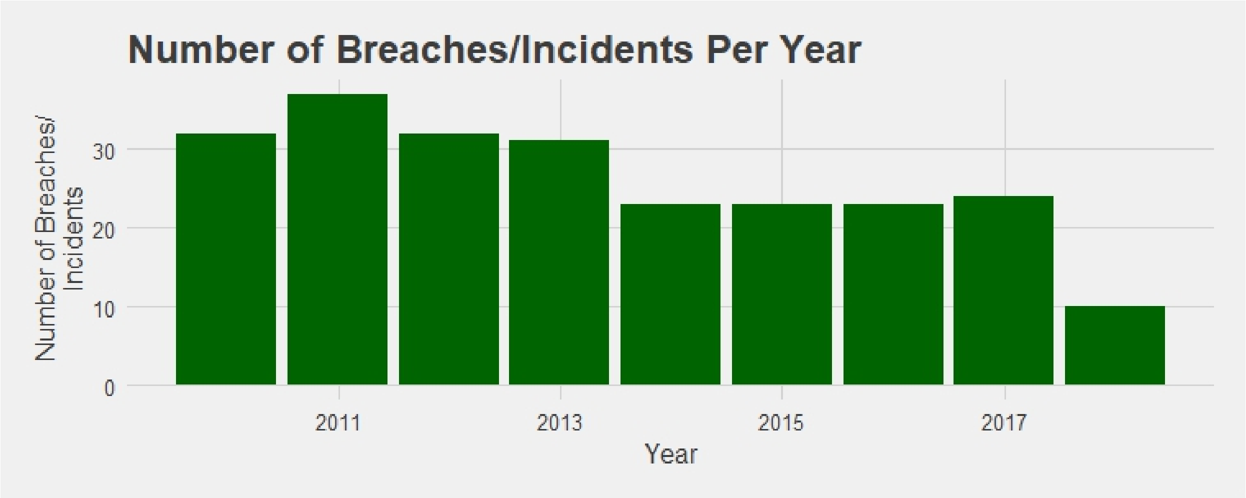


Figure 2.1: PRC Breaches Per Year

Notably, there are relatively few breaches among publicly traded companies in any given year. Some years have slightly over 30 breaches in this data, and closer to 25 in others. Compared to a universe of approximately 2000 companies [1], this makes breaches quite rare. In computer science terms, this is referred to an ”imbalanced learning” problem because one class (”no breach”) dominates in numbers over the other class (”breach”).

Broken down by incident type, it is clear that the bulk of incidents is quite serious. In Figure 2.2, STAT refers to stationary computer loss, DISC to unintended disclosures, and PORT refers to portable device loss. Meanwhile HACK refers to outside hacking or malware infections, and INSD refers to a company insider intentionally breaching information. The number of incidents in the HACK category grows over time, while unintentional data losses become less frequent over time. This trend may suggest that companies are becoming more careful and better at preventing data losses that result from carelessness. On the other hand, outside attacks have grown over time, which can point to increased cybercriminal activity,

[1] According to the Wall St. Journal, there are approximately 3500 publicly traded companies in the U.S. However because of inconsistencies in how companies report their disclosures under different central index key numbers (ciks), matching disclosure text, financial information, and incident information is difficult. Future iterations of this work will work to complete the dataset used in this paper to include all companies across all U.S. stock exchanges. That being said, aside from a handful of notable exceptions (e.g. McDonald’s), there are few breached firms that did not make it into the dataset.

or a substitution away from techniques like phishing toward more sophisticated techniques like malware.

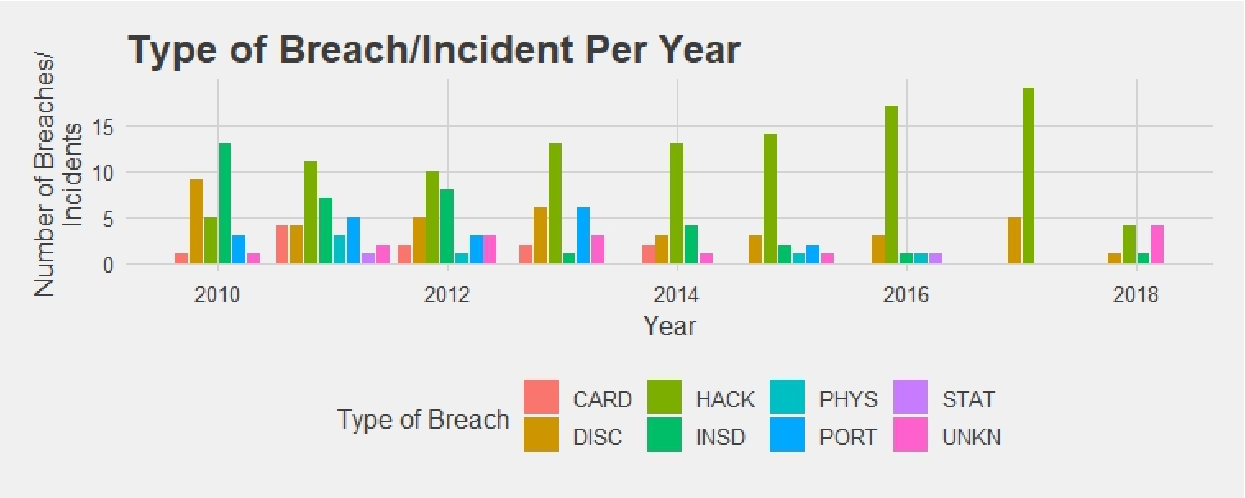


Figure 2.2: PRC Breaches Per Year and Type

Looking at the descriptions of the events paints a similar picture. Figure 2.3 shows a word cloud visualizing the most common words in the descriptions of the incidents. PRC writes these descriptions summarizing the description of the events from the source of the information about the breach (newspaper article, mandatory disclosure, etc.). Social security, credit card, bank, and email information are among the things talked about in these descriptions. These words give some idea of the sort of information that is most frequently compromised in these sorts of incidents among publicly traded companies. Geographically, incidents are concentrated in a handful of places. Firms in New York, New Jersey, and California make up the bulk of the outcome data. Given the prevalence of publicly traded companies in industries like finance and technology, this is unsurprising. Figure 2.4 shows the geographic spread.

Firm-Level Data

In machine learning applications, text features tend to perform best when combined with non-text features. In this case, I collect firm-level data on each publicly traded company in my dataset. These features are helpful primarily because they are already publicly available and easy to use, and therefore can provide a reasonable baseline for how regulators may try to predict data breaches without leveraging text information.

First, I extract industry codes and addresses. This information is helpful primarily because firms belonging to different industries will likely prepare for and respond to cybersecurity incidents in different ways. For example firms that handle health information are

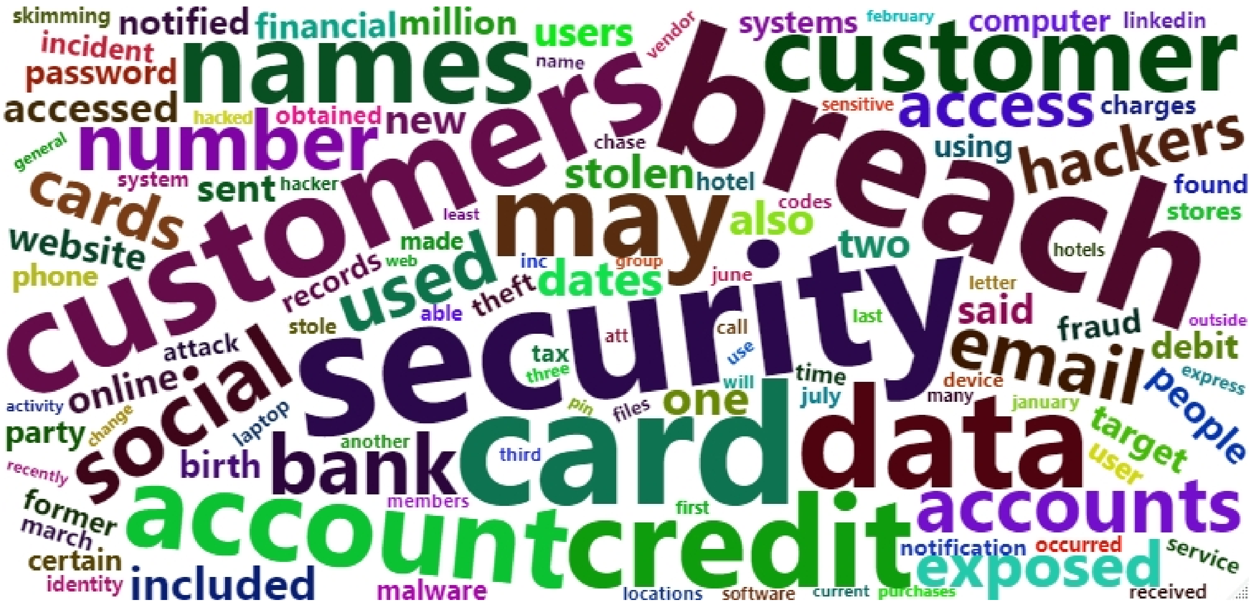


Figure 2.3: Word Cloud of Description of Incidents

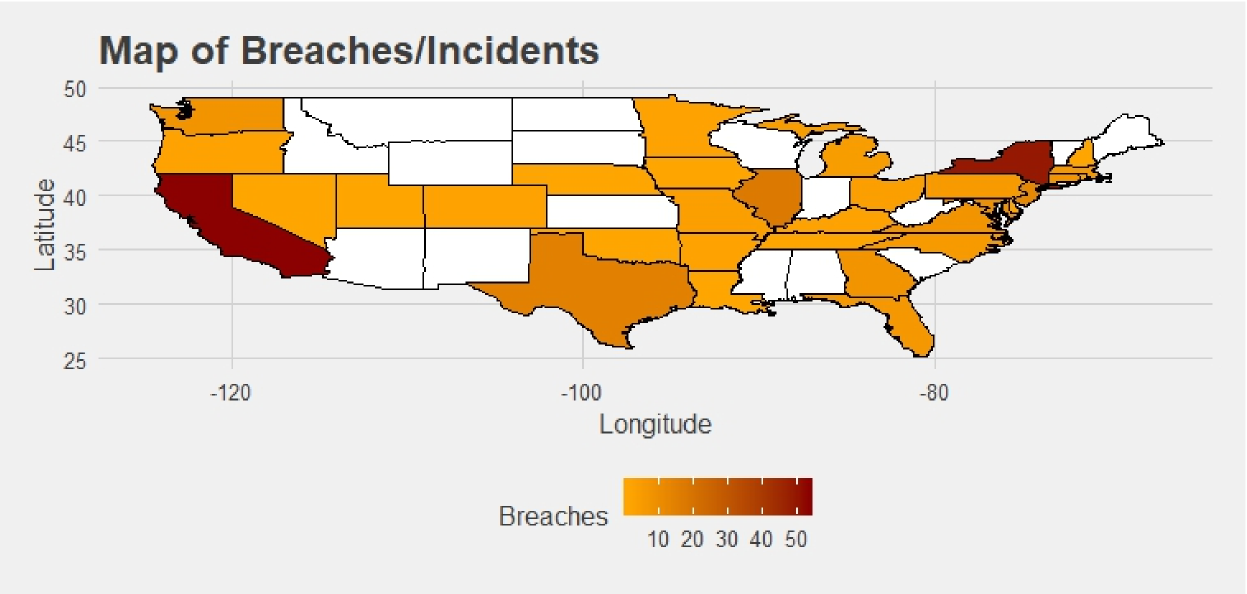


Figure 2.4: Map of Breaches and Incidents

susceptible to stronger negative consequences stemming from breaches, and may be more likely to invest more in precaution as a result. One example of different incentives is that generally consumers do not have individual causes of action after the announcement of a breach, but generally do enjoy causes of action when protected health information is compromised. Industry codes are therefore potentially valuable information, and geographic information may also be relevant insofar as it may serve as a proxy for things like firm size, products, etc.

Industry codes are also interesting because different industries have varying cyberrisk profiles. Figure 2.5 shows the number of breached and non-breached observations among a subset of the most represented industries in the dataset. Some industries, such as real estate, are well-represented in the dataset, but suffer relatively few breaches or incidents. Figure 2.6 shows the ratio of breached observations relative to non-breached observations per industry. Although there are over 200 industries represented in the dataset, only approximately 40 suffer cybersecurity incidents at all. Notably, 50% of observations associated with the financial services industry also correspond to breaches. Telecommunications, software, and retail also have fairly high risk profiles.

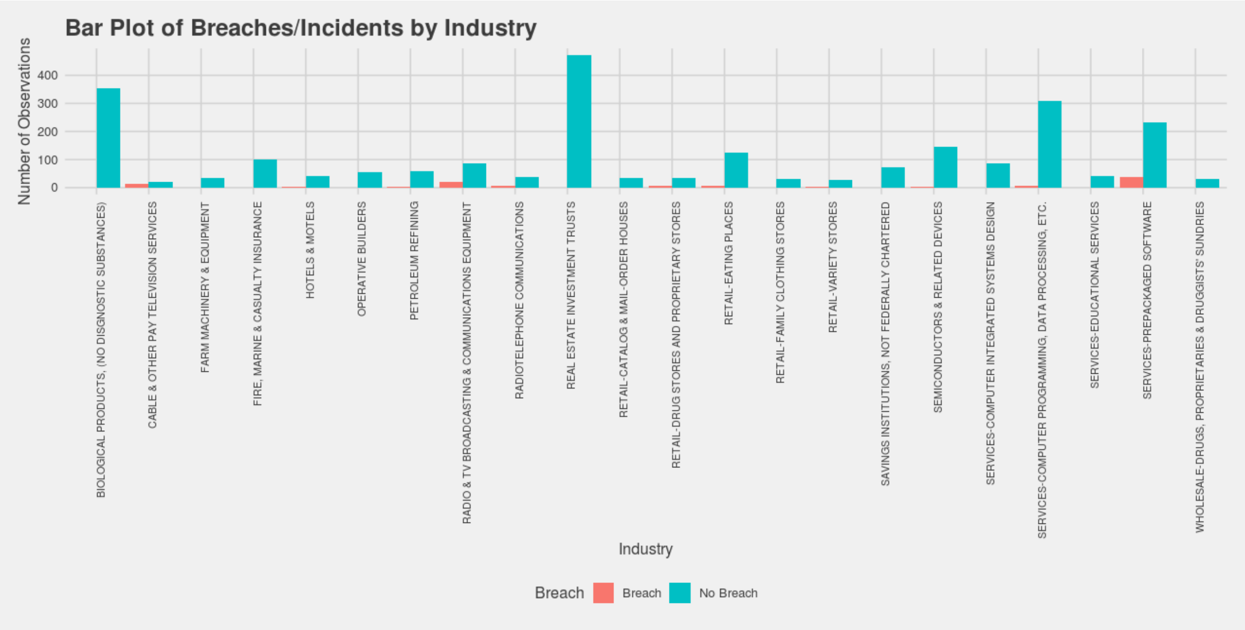


Figure 2.5: Industries with Breaches/Incidents

I also incorporate firm level data from US Stocks Database maintained by the Center for Research in Security Prices. Table 2.1 summarizes the features drawn from the US Stocks Database. Critically, I avoid trying to predict how stocks may respond to cybersecurity incidents. Rather, I use stock volatility as a proxy for a firm's general riskiness, as measured by how investors respond in capital markets. (Kogan et al., 2009) already demonstrated that text analysis successfully predicts an asset's stability fairly well. The basic idea here is to take that measure of riskiness, and use it as a feature to predict cybersecurity riskiness.

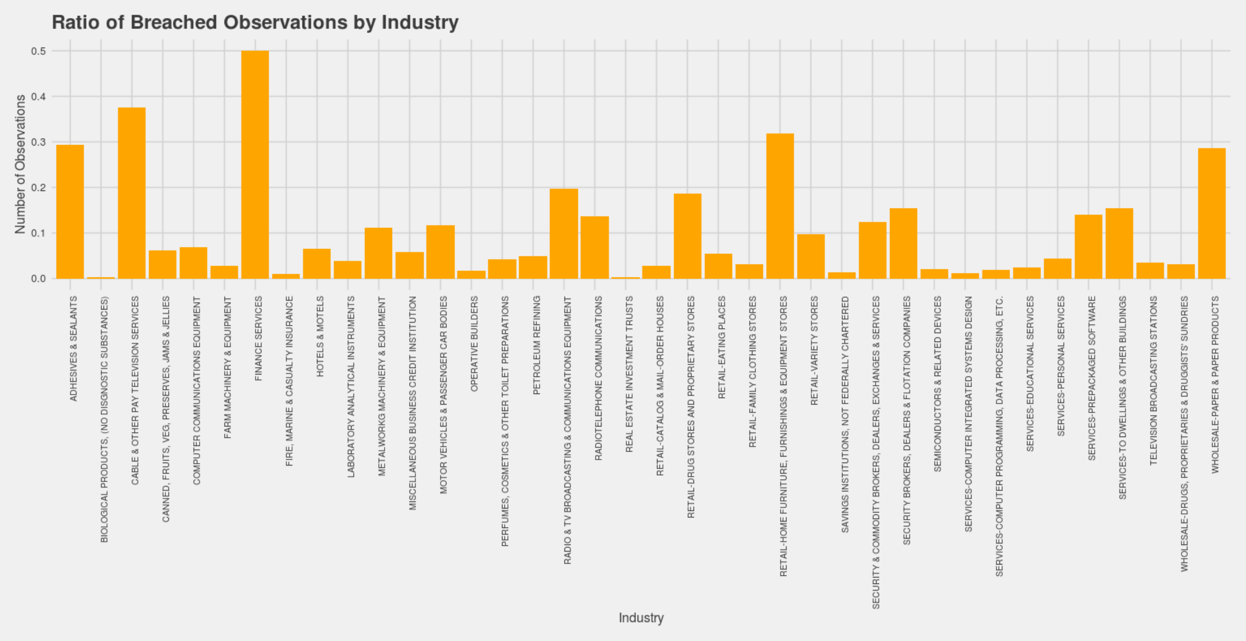


Figure 2.6: Ratio of Breaches/Incidents

Text Data

The text data source is the Securities and Exchange Commission's (SEC) datasets that collect companies' annual 10-K disclosures. In these 10-Ks, firms are required to disclose potential risk factors, including cybersecurity risks, to their investors. However, the SEC recognizes that companies need to manage the language in these disclosures so as to not create a roadmap for potential cybercriminals to exploit vulnerabilities.

Extracting Risk Disclosure Text

The most difficult data collection task is collecting all of the relevant SEC filings so that they can be matched to the outcome data. The SEC provides an online search tool (EDGAR) for looking up individual firms and their corresponding documents, but this does not lend itself to dataset construction.

Luckily, a number of open-source packages are available that aid with this task. In particular, I use the "edgar" and "edgarWebR" packages in the R computing environment. The edgar package provides a list of "Central Index Key (cik)" numbers that uniquely identify each publicly traded company. The edgarWebR package includes functions for looking up companies by their cik numbers, and extracting the raw text of their disclosures. A key feature here is that the package also includes a method for tagging parts of a disclosure, such that a user may tag all text that falls under the "Risk Disclosure" heading, which is always "Item 1A" on a 10-K disclosure form. Because some forms may be ill-formed, doing

Feature Name	**Explanation**
CURCD	Native Currency Code
TXDB	Deferred Taxes
TXDBCA	Deferred Tax Asset
TXDBCL	Deferred Tax Liability
TXDITC	Deferred Taxes and Investment Tax Credit
TXNDB	Net Deferred Tax Asset (Liability) - Total
TXNDBA	Net Deferred Tax Asset
TXNDBL	Net Deferred Tax Liability
TXNDBR	Deferred Tax Residual
TXP	Income Taxes Payable
CSHTR_C	Common Shares Traded - Annual - Calendar
DVPSP_C	Dividends Per Share
PRCC_C	Price Close - Annual - Calendar
PRCH_C	Price High - Annual - Calendar
PRCL_C	Price Low - Annual - Calendar
CSHTR_F	Common Shares Traded - Annual - Fiscal
MKVALT	Market Value - Total - Fiscal
ADDZIP	Zip Code
CITY	Headquarters City
State	Headquarters State
Industry Title	Standard Industry Code Industry

Table 2.1: Features Drawn from U.S. Stocks Database

this computationally may not capture every relevant aspect of every disclosure. However, it should be sufficient for most purposes.

After extracting the risk disclosure text, the next task is combining it with the outcome data and other features. Ultimately, the resulting dataset contains information about a firm's name, cik number, filing date, risk disclosure text, firm-level features drawn from the U.S. Stocks Database, and a logical indicator suggesting whether it was breached in the year following the publication of its risk disclosure. An example dataframe can be viewed here.

Exploratory Analysis

An example of the raw text of a disclosure can be seen here. This filing is from Apple's 10-K filing in 2011. Under its risk disclosure, it says the following about cybersecurity risks:

> The Company may be subject to breaches of its information technology systems, which could damage the Company's reputation, business partner and customer

> relationships, and access to online stores and services. Such breaches could subject the Company to significant reputational, financial, legal, and operational consequences.
>
> The Company's business requires it to use and store customer, employee, and business partner personally identifiable information ("PII"). This may include names, addresses, phone numbers, email addresses, contact preferences, tax identification numbers, and payment account information. Although malicious attacks to gain access to PII affect many companies across various industries, the Company may be at a relatively greater risk of being targeted because of its high profile and the amount of PII managed.
>
> The Company requires user names and passwords in order to access its information technology systems. The Company also uses encryption and authentication technologies to secure the transmission and storage of data. These security measures may be compromised as a result of third-party security breaches, employee error, malfeasance, faulty password management, or other irregularity, and result in persons obtaining unauthorized access to Company data or accounts. Third parties may attempt to fraudulently induce employees or customers into disclosing user names, passwords or other sensitive information, which may in turn be used to access the Company's information technology systems. To help protect customers and the Company, the Company monitors accounts and systems for unusual activity and may freeze accounts under suspicious circumstances, which may result in the delay or loss of customer orders.
>
> The Company devotes significant resources to network security, data encryption, and other security measures to protect its systems and data, but these security measures cannot provide absolute security. The Company may experience a breach of its systems and may be unable to protect sensitive data. Moreover, if a computer security breach affects the Company's systems or results in the unauthorized release of PII, the Company's reputation and brand could be materially damaged and use of the Company's products and services could decrease. The Company would also be exposed to a risk of loss or litigation and possible liability, which could result in a material adverse effect on the Company's business, results of operations and financial condition."

This disclosure represents just one case, and Apple may be more conscientious than most companies. That being said, this type of language reflects the sort of text that might distinguish various cybersecurity practices. If there are patterns in the language, details, and other information presented in cybersecurity risk disclosures, this may emerge through natural language processing.

More generally, we can see general patterns in the way that companies talk about their risks. Figure 2.7 shows a topic model for two topics, trained on the text of the risk disclosures. These topics give a sense of the sorts of terms that are likely to appear together in a disclosure. Specifically, the concepts of ”risk,” ”price,” and ”adverse” seem to come up, which should not be surprising given the nature of section 1A.

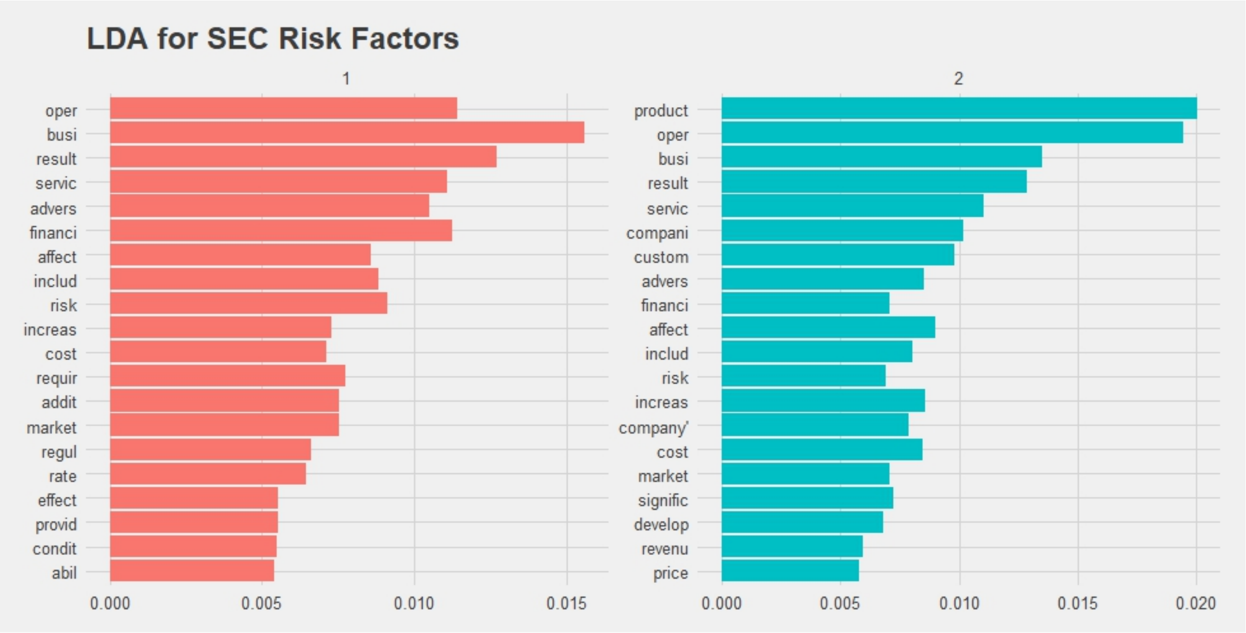


Figure 2.7: Latent Dirichlet Allocation for 2 topics

Feature Engineering

In addition to firm-level and textual data, I also conduct feature engineering to manually create some features that may be helpful for prediction purposes. From the firm-level data, I calculate the difference between high and low stock prices for the year to reflect stock volatility. I also make a logical indicator for companies that experiences breaches or incidents in previous years. Finally, I calculate the ratio of breached observations to unbreached observations within an industry.

I also used keyword searches of the disclosure text to create features that mapped to the SEC's interpretative guidance. Some examples of manually created features and the associated keywords can be seen in Table 2.2. Further feature engineering would use more sophisticated methods to pick up on the concepts underlying the SEC guidance, but keywords are a first attempt to see how basic models would do. Concretely, the SEC interpretative guidelines look at the following elements:

- Occurrence of prior cybersecurity incidents
- Probability of the occurrence and potential magnitude of cybersecurity incidents

- Preventative actions taken to reduce cybersecurity risks and associated costs
- Aspects of business that give rise to material cybersecurity risks
- Costs associated with maintaining cybersecurity protections
- Potential for reputational harm
- Existing or pending laws that might affect cybersecurity risk
- Litigation, regulatory investigation, and remeditation costs associated with cybersecurity incidents

Feature Name	Key Words
Probability of Occurrence	Cyberattack, Previous Incident
Preventative Actions	IT Security, Encryption, Cybersecurity Awareness Training
Aspects of Business	Personal Data, PII, PHI, Password
Reputational Harm	Harm to Our Reputation, Reputational Harm
Existing Laws and Regulation	Produce User Data, User Data Requests, Government Requests for User

Table 2.2: Features Engineered from SEC Interpretative Guidance

2.5 Policy Setup & Exploratory Analysis

In this section, I sketch out the decisionmaking process for SEC cybersecurity audits. I describe the substance of cybersecurity audits, as well as trends in how many have been conducted over the last few years. I then provide a simulation of how well randomly choosing firms to audit does at predicting future breaches. I then provide a simple model that uses a logistic regression to estimate the likelihood of a breach.

Background

The SEC is increasingly paying attention to cybersecurity risks and is taking active steps to safeguard investors. In 2017, the SEC established a Cyber Unit in its Division of Enforcement. According to the SEC's website, the "Cyber Unit focuses on violations involving digital assets, initial coin offerings and cryptocurrencies; cybersecurity controls at regulated entities; issuer disclosures of cybersecurity incidents and risks; trading on the basis of hacked nonpublic information; and cyber-related manipulations, such as brokerage account takeovers and market manipulations using electronic and social media platforms."

Most of the enforcement actions brought so far deal with initial coin offerings, but the SEC also pursues actions related to failure to adequately disclose material events and cyberrisks.

In 2017, the Office of Compliance Inspections and Examinations (OCIE) conducted a pilot program where it audited the cybersecurity policies and practices of 75 publicly traded firms. It found that while firms generally had written policies in place about how they should deal with cyberrisk and adverse events, oftentimes these written explanations were too vague to provide helpful guidance to employees. Moreover, it was not always clear that firms actually implemented some of their written policies, such as requiring and monitoring cybersecurity training. In general, the SEC is expanding its auditing and enforcement efforts, as the number of firms subject to some kind of audit (not just cybersecurity) increased from 8% to 13% from 2013 to 2018. As part of this general expansion, the SEC is paying particular attention to cybersecurity concerns.

SEC Cybersecurity Audits

In 2015, the SEC launched its Cybersecuriry Examination Initiative. With this notice, the SEC outlined the general procedure for its cybersecurity audits, and what minimum standard firms are expected to uphold. The specific areas that SEC examiners focus on are:

- Governance and Risk Assessment
- Access Rights and Controls
- Data Loss Prevention
- Vendor Management
- Training
- Incident Response

In these audits, the examiners look at both a company's written policies, as well as their actual practices. There is now a cottage industry surrounding preparedness for these cybersecurity audits. One source suggests that an audit may take about six days, and requires three SEC auditors (one of whom specializes in cybersecurity audits).

Metrics

Before providing baseline simulations to motivate the core policy problem, I define basic metrics for evaluating the efficacy of cybersecurity audits. Simply put, the prediction task here is predicting whether a firm will suffer a cybersecurity breach or incident. There are various ways to define whether a prediction task is working well. In this case, the task is predicting "breach" or "no breach," with "breach" being the "positive" class. Some foundational building blocks to think about predictions in this case include:

- **True Positives (TP)**: Predictions where the model accurately predicts the positive class. In this case, these are instances when a model predicts a "breach" and there was indeed a breach.
- **False Positives (FP)**: Predictions where the model erroneously predicts the positive class. In this case, these are instances when a model predicts "breach" when there was no breach.
- **True Negatives (TN)**: Predictions where the model accurately predicts the negative class. In this case, these are instances when a model predicts "no breach" and there was no breach.
- **False Negatives (FN)**: Predictions where the model erroneously predicts the negative class. In this case, these are instances when a model predicts "no breach" when there was actually a breach.

In this context, true positives and false negatives are the most consequential metrics. Successfully predicting a true positive indicates that the model found an ideal candidate for an audit, while predicting a false negative (failing to detect a breach) implies a situation where an audit may have helped but the model failed to direct the intervention toward that firm. False positives imply that the model would have a firm audited that may not have needed it, and while this imposes costs on the agency, is not as consequential as a false negative. Meanwhile, true negatives are trivial to predict in this context because relatively few firms are breached in any given year.

Delving deeper, these metrics can be combined in useful ways.

$$\textbf{Accuracy}: \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Accuracy is essentially a measure of how many times the model was correct in its predictions in either direction, divided by the total number of predictions it made.

$$\textbf{Recall}: \frac{TP}{TP + FN}$$

Recall is a measure of successful the algorithm was at detecting instances of the positive class. In this case, the ratio is an expression of what fraction of the actual breaches the algorithm successfully predicts.

$$\textbf{Precision}: \frac{TP}{TP + FP}$$

Precision is a measure of how successful a model at filtering out noisy predictions. Put differently, it is a statement of what fraction of all the predictions of the positive class were actually in the positive class. In this case, it is saying of all the firms that the model predicted would be breached, how many were actually breached.

Random Audits

In 2015, the SEC began its cybersecurity auditing program. That year, the SEC conducted 75 audits. I simulate these audits to provide a baseline for an algorithm to improve upon. To do this, I simulate randomly choosing 75 firms to audit, and plot the distributions of how well these audits predict eventual breaches. I ran 100,000 simulations of picking 75 firms to audit at random, and then plotted distributions for true positives, recall, and precision. To accomplish this, I looked at breaches for the 2015-2016 fiscal year, where there were 10 breaches.

Figure 2.8 shows the distribution of true positives across these simulations. Across 100,000 simulations, the modal outcome is to successfully detect 0 breaches in advance. In the tail of the distribution, randomly auditing may pick up on one or two eventual breaches, but hardly ever exceeds these figures. Similarly, recall follows the same pattern, as seen in Figure 2.9.

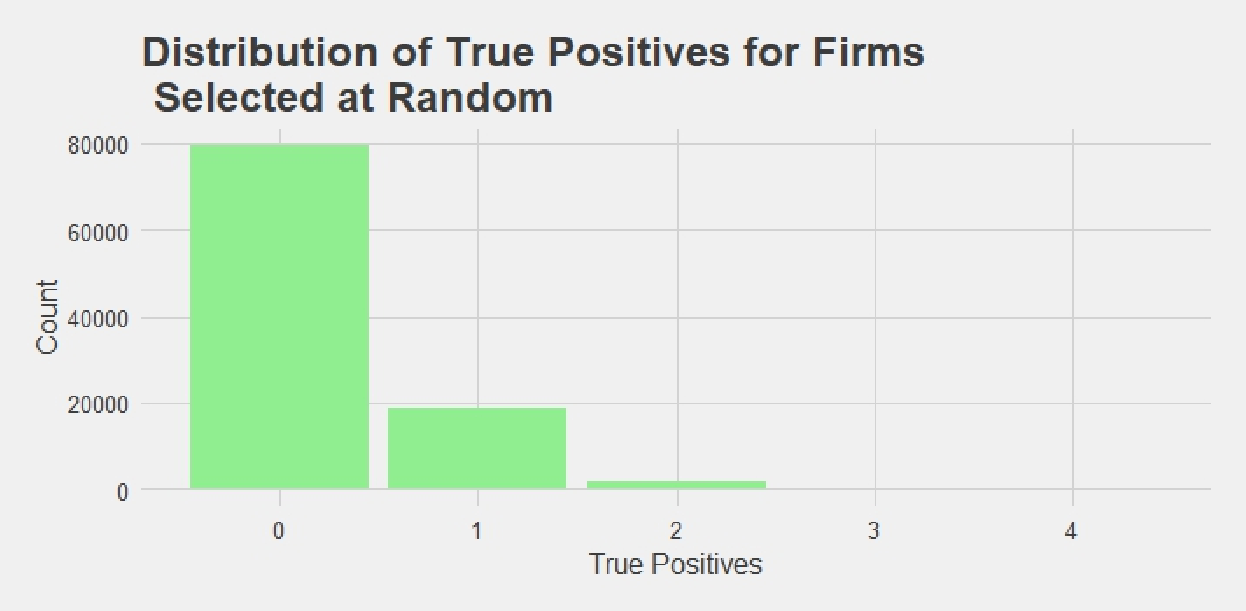


Figure 2.8: Distribution of True Positives in Random Audits

Even if an algorithm could not improve on true positive and recall measures, there is substantial room to improve on precision. Figure 2.10 shows the distribution of precision across simulations of random audits. Even in cases where a random audit successfully predicts a breach, the precision score lies somewhere between .015 and .025. The takeaway here is that of the 75 audits conducted, about 73-74 are potentially wasted. Any improvement over this precision score would make these audits more efficient.

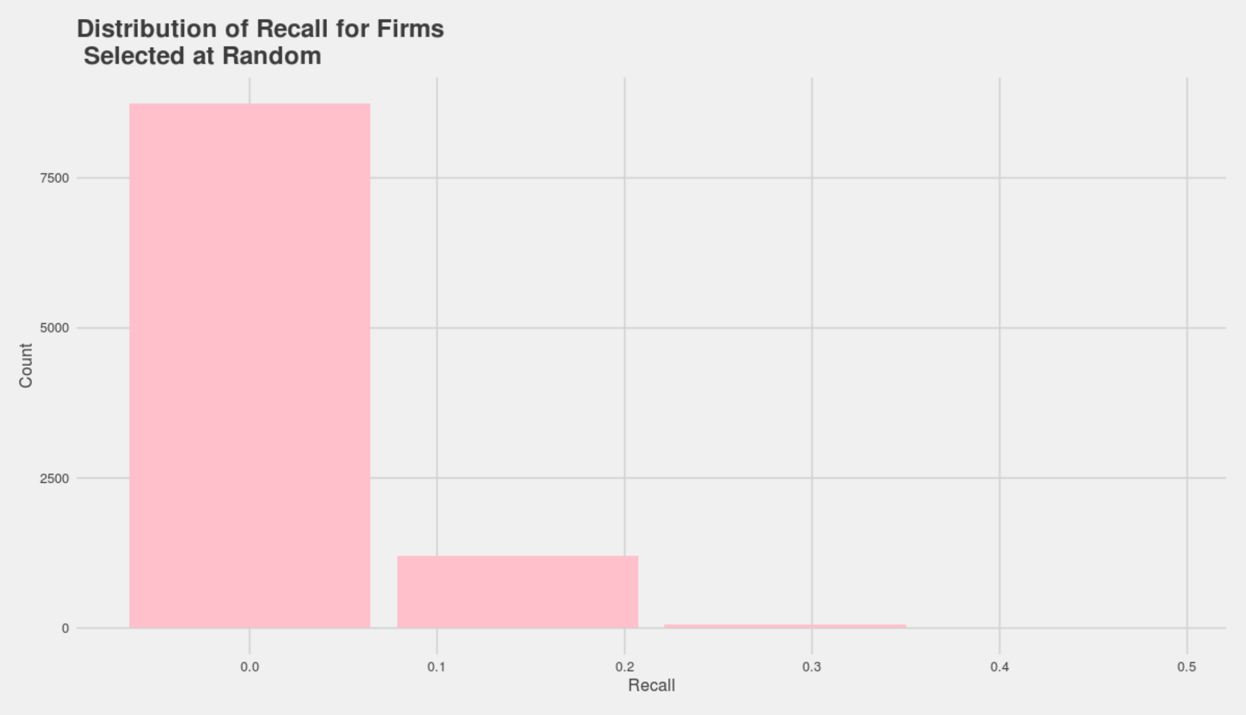


Figure 2.9: Distribution of Recalls in Random Audits

Logistic Regression

Next, I train a logistic regression to simulate how well a simple algorithm performs on this prediction task. Using the same subset as I did with the random audits, I train the logistic regression on the disclosures. In this case, I featurize the text of the disclosures using the term frequency-inverse document frequency (tf-idf) technique. The simplest natural language processing (NLP) model that could be used is the "bag of words" model where the columns in the dataset correspond to counts of how many times a given word appears in a document. In this context, a document is a 10-K disclosure for a particular company and year. Instead of using a bag of words, tf-idf takes the number of times a term appears in a given document (term frequency), and then multiplies that by the inverse of the number of documents that the term appears in (idf). The basic intuition here is that more weight is given the more times a term appears within a document, but then weight is decreased the more common a term is across documents. Thus, tf-idf does well classifying documents where individual documents have key terms that do not appear elsewhere in the corpus.

Table 2.3 shows a confusion matrix that shows how well a logistic regression does with tf-idf weighting at predicting outcomes for the 2015-2016 fiscal year. A confusion matrix is a useful tool for visualizing how well an algorithm did at a classification task, and common metrics like accuracy, recall, and precision are easily derived from it. Of the 4209 companies

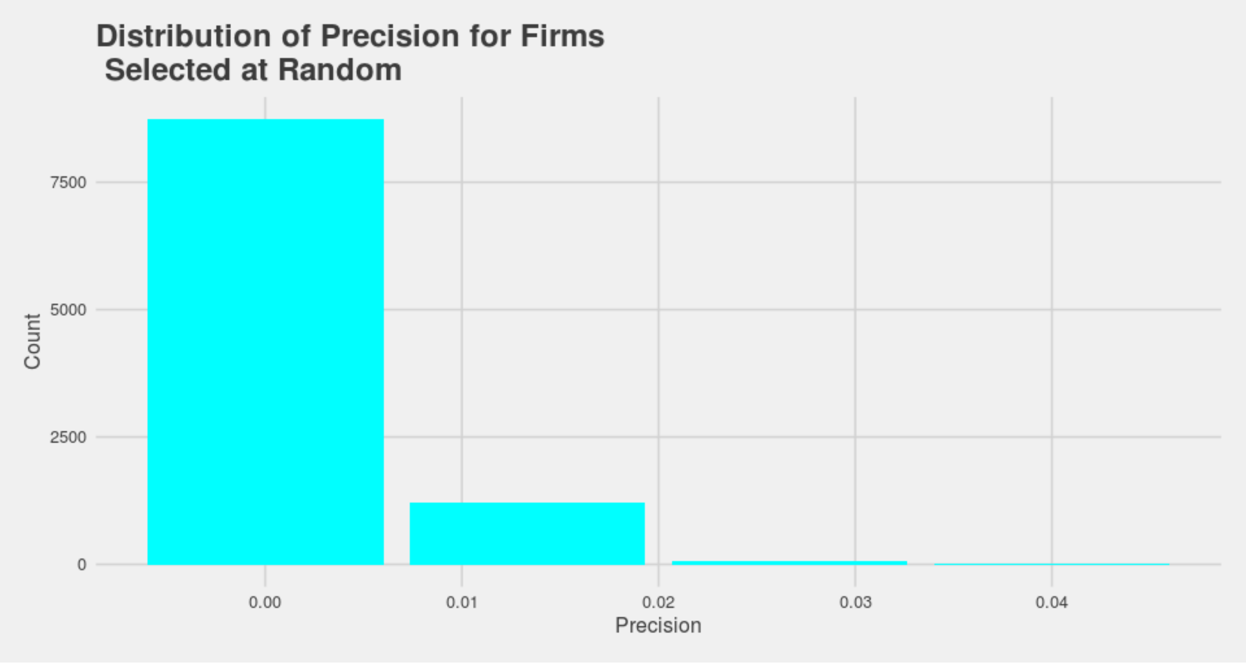


Figure 2.10: Distribution of Precisions in Random Audits

in this dataset, 5 suffered breaches. The model predicted 4 companies would suffer breaches, but none of these predictions overlapped with actual breaches. The model did successfully predict almost every case of ”no breach,” but this is of little value given the severe imbalance in the dataset.

	observed no breach	observed breach
predict no breach	4200	5
predict breach	4	0

Table 2.3: Logistic Regression Confusion Matrix

In this case, the simple model does even worse than random auditing. While a combination of logistic regression and tf-idf has advantages with regard to transparency and interpretability, more complex models are likely to do better in this case. Given that this logistic regression had 0 true positives, it similarly had a recall and precision of 0.

The low baselines implied by both random auditing and logistic regression motivate the possibility for exploring other methods that can enhance successful prediction of cybersecurity incidents. The poor performance of a simple logistic regression may also point to why the SEC and other regulatory agencies have thus far been slow to adopt algorithmic approaches to prediction policy problems.

2.6 Methodology

Featurizing Text

To featurize the text (turn text into quantitative information), I use word2vec. Word2vec is a set of popular word embedding models first introduced by Mikolov et. al. (Mikolov et al., 2013). Word2vec is a ”word embedding” technique, meaning it converts words into numerical vectors, and puts substantively similar words into vectors that are close together. Specifically, I use document-averaged word embeddings from word2vec to transform the raw text of annual disclosures into quantitative features.

Frequency-Based Featurization

The simplest model for featurizing text would be the ”bag-of-words” approach. A bag-of-words is a frequency-based scheme that essentially counts how many times a word appears in a document and creates a feature for that count. One popular extension of the bag-of-words technique is the ”term frequency-inverse document frequency” approach which counts the number of times a word appears in a document, but divides that figure by the number of times that word appears across a corpus. Thus, words that are unique to a document will get higher weights than words that appear frequently across documents.

Frequency-based approaches are useful because they are easy to implement and interpretable. Tf-idf in particular can be quite powerful when dealing with a classification task where there are words that are good as discriminating between class labels. For instance, in e-mail spam detection, certain key words show up in spam e-mails that rarely show up in legitimate ones. A model trained on tf-idf or bag-of-words features would be able flag spam simply by looking at whether words associated with spam class labels appear in a document. The disadvantage of these approaches is that in more complicated classification tasks, frequency based embeddings may sacrifice too much of the substantive meaning of words to be useful predictors. Whereas in the spam example there is a clear link between the presence of some words and the outcome label, this relationship is not always so strong. In the context of this study, word frequencies likely do not so neatly map into the outcome of a firm being breached, thus warranting considering more information.

One Hot Encoding

One way to capture more of a word’s meaning in context is the one-hot encoding approach. A one-hot encoder essentially takes a collection of words (sentence, paragraph, or document), and creates logical indicators for whether words in the corpus appear in that collection. For example, if we had five words in a feature space, ”I,” ”love,” ”data,” ”is,” and ”cool” then the following sentences would be encoded as follows:[2]

[2] This example is borrowed from: https://towardsdatascience.com/an-intuitive-explanation-of-word2vec-208bed0a0599

	I	love	data	is	cool
I love data	1	1	1	0	0
love is cool	0	1	0	1	1
data is cool	0	0	1	1	1

Table 2.4: One-Hot Encoder

This featurization approach is useful primarily because it encodes sentence-level (or paragraph/document) information in a numerical vector. [3] By representing sentences as vectors, more information about the distance between sentences is available to the analyst. However, the one-hot encoding still does not understand the meaning of the sentences. Although ”love is cool” is close to ”data is cool” in vector space, these vector representations still depend on the appearance of certain words, rather than their actual substantive meaning. Moreover, in this application I am looking at entire documents. The feature space quickly becomes high-dimensional when one-hot encoding thousands of long documents, which increases computational complexity.

Word2Vec

Word2Vec is a word embedding technique that uses a prediction-based approach to creating word vectors. The training process for word2vec involves predicting words based on their surrounding context. [4] Using the context surrounding a word as input, this context is passed through a neural network to produce vectors with probabilities to predict target words. The algorithm then uses a technique called backpropogation to adjust the weights it assigns to these vectors until it minimizes the loss function (or more simply, until it does as well as it can at predicting words). This process then outputs vector representations for each word, and words with similar contexts will have vectors that are closer together in vector space. The canonical example of the advantage of this approach is encapsulated in this relationship:

$$\vec{king} - \vec{man} + \vec{woman} = \vec{queen}$$

Taking the vector for king, subtracting the vector for man, and adding the vector for woman yields a vector that is very close to the vector for queen. Thus, the word2vec

[3] A ”vector” in this case should be understand in its linear algebra context. A vector represents an object with a magnitude and direction (for instance, the acceleration of an object), and vectors can be operated on in a vector space, which is a collection of vectors. In this case, the vectors encode information about a sentence, and situates each sentence in a vector space relative to other sentences.

[4] This prediction can either use Continuous Bag-of-Words (predicting a target word from surrounding words) or skipgram (predicting surrounding words from a target word). CBOW does better with larger datasets and common words, whereas skipgram is better for smaller datasets and rare words. See Figure 2.11 for an illustration.

representations are able to capture the idea that the concept of a queen is similar to king, except for a difference in gender. Thus, these word2vec vectors are able to capture more contextual meaning than word frequency or order.

I use word2vec and update the vectors with the text of the SEC disclosures. I then take these vectors, and average them across documents. Doing so creates a document-level vector that is built upon the tuned word vectors. These document-level vectors then become features in the downstream classification task, which is predicting firms that are likely to suffer cybersecurity incidents.

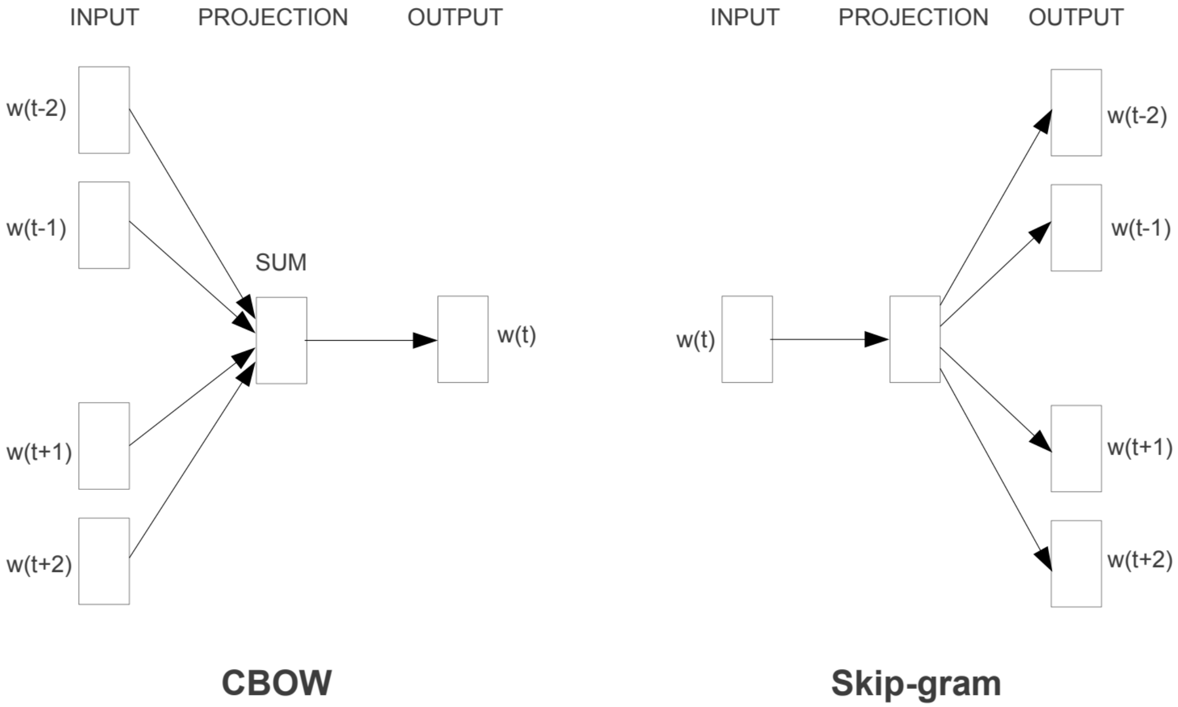


Figure 2.11: Illustration of Continuous Bag-of-Words and Skipgram, taken from Mikolov et. al. 2013

Modeling

Constituent Models

I use several constituent models before fitting an ensemble model. Importantly, each of these models is well-suited to classification tasks, though some can be used for regression as well. In machine learning terms, a classification task is distinguished from a regression task by the nature of the target variable (the variable that we are trying to predict). Classification is predicting which class label an observation belongs to. Binary classification predicts a target that can take one of two class labels, whereas multi-class classification predicts targets with many labels. In contrast, regression predicts continuous target variables. In this case, predicting whether a firm is breached or not is a binary task.

Logistic Regression

Logistic regression (logit) is a common algorithm in the social sciences, and is especially popular for binary classification tasks. Most social science applications of logistic regression report coefficients on the features (independent variables or covariates in social science language) for causal estimates. These coefficients are generally reported as log-odds, though sometimes are exponented to odds ratios. Critically, in a prediction context, the coefficients are not the object of interest for analysts. Rather, only the predicted probabilities for the target in the test set are relevant for the analysis. In a prediction setting, the causal interpretation of various coefficients is not especially relevant because a policymaker does not need to understand the precise relationship between the outcome and a feature to make a decision.

Poisson Regression

Poisson regression is a generalized linear model that is popular for modeling count data. Generally, poisson models are not used for binary outcome data, but I use one here because of poisson's strength in modeling rare events. In this case, poisson is akin to using a linear probability model. Essentially, these approaches use a linear model to estimate a binary outcome. The main disadvantage of these approaches is that without restrictions, it is possible to predict values outside the range [0,1], which would be invalid for a truly binary outcome.

Classification Tree

Decision tree learning is a machine learning approach that predicts a target value through a series of decision rules. Trees can be used for both classification and regression problems. The basic idea behind a decision tree is that learns the relationships between features and targets by growing a tree that encapsulates various decision rules. The tree starts at an initial node, and then makes a split into two new nodes based on some decision rule. At each of these nodes, the tree then splits again based on a new rule. This process iterates until the tree cannot make any more splits. A frequently used example to illustrate this concept is a classification tree that predicts whether a passenger on the titanic would survive given the rules for who was allowed to board a lifeboat (See Figure 2.12).

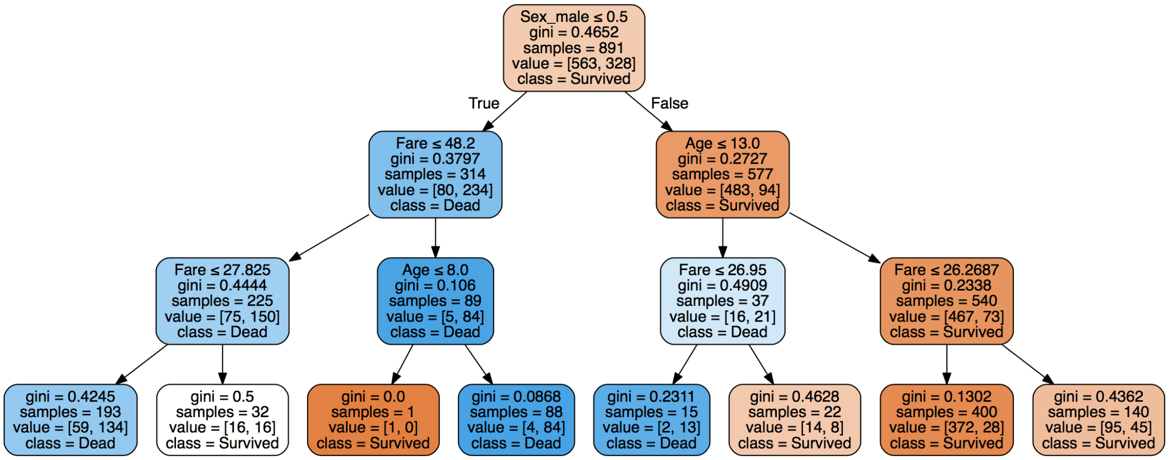


Figure 2.12: Titanic Survival Classification Tree

Random Forest, Gradient Boosting Classifier, and Adaptive Boosting

Classification trees have a few drawbacks, however. Without pruning (reducing the depth of a tree), trees tend to overfit the data, thus achieving poor performance out-of-sample. Trees also initialize from a randomly chosen feature, and make probablistic splits. Thus, any given tree may be overfit to idiosyncrasies in that particular random sample. To address these problems, classification trees are frequently combined in ”ensemble” methods.

One approach to solving these problems is using a ”bagging” technique such as a random forest. A random forest grows many classification trees in parallel, and then has each tree vote for the outcome. The prediction with the majority vote is the final prediction for the random forest. Random forests are popular because they reduce the tendency of single trees to overfit, and can be trained quickly with parallel processing.

Another ensembling approach for trees are ”boosting” algorithms. Whereas bagging grows trees in parallel, boosting instead iteratively combines weak classifiers (classifiers that do slightly better than a coin toss at predicting an outcome) to create a strong classifier (a classifier that has close to 0 error). Boosting takes longer to train than bagging because it is iterative, but has the advantage of having each sequential model learn from the mistakes of the previous models. In this case, I use gradient boosting and adaptive boosting, which primarily differ in how they combine weak learners. Gradient boosting learns from the errors (pseudo-residuals) in the previous iteration of the algorithm. Adaptive boosting learns by upweighting data points that were incorrectly classified in the previous iteration, thus forcing the algorithm to learn how to deal with more difficult decisions.

Ensembling trees is especially attractive in an imbalanced dataset setting. In this case, the cases of ”no breach” far outnumber the ”breach” observations. Ensembles are better at predicting minority class observations because they reduce noise from overfitting, and are built in a way that focuses on harder cases.

Soft Voting Ensemble Learner

Finally, I take all of the constituent algorithms, and combine them into a soft-voting ensemble classifier. Using the predicted probabilities from each model, the ensemble takes the average of these probabilities and makes a decision based on that average. This approach can be contrasted with hard-voting classifiers where each model takes a vote, and the majority vote is the ensemble's decision. The soft voting ensemble takes advantage of the fact that each of these models outputs predicted probabilities, and combines them into a meta-learner.

Ensemble classification is helpful primarily because it ameleriorates idiosyncrasies that may plague any individual model. Moreover, knowing the "correct" model a priori is impossible, and ensembles help approximate the best possible model by averaging constituent models. Ensembles take advantage of the fact that if each model is more likely than not to make the correct prediction, combing their predictions will boost the accuracy because it is less likely that idiosyncratic errors in one model will turn into incorrect predictions.

Temporal Cross-Validation

A potential problem in building machine learning models on temporal data is the tendency for future information to leak into the training process. In typical machine learning modeling, the analyst splits the data into train and test sets (sometimes adding a "validation" set as well). The train/test split is done at random in most applications, and the model is then trained on the training data, and its predictions are compared to the true observations in the test data. However, this framework quickly breaks down with temporal data. If the splits are done randomly with temporal data, the machine learning algorithm learns patterns from a future time period, and its performance will be artificially boosted when it is tested on data from a previous time period. For instance, imagine if the training set randomly included the Target 2013 breach outcome in its training data, and the test set included the 2011 and 2012 financial disclosures timeframes. When testing the algorithm, it will almost assuredly predict a breach because it borrows information from a future year. This would make the algorithm seem accurate, but would not reflect actual deployment conditions, as a regulator will not have advance notice of a breach (indeed, such information would obviate the need for an algorithmic approach).

Instead, I utilize a "temporal cross-validation" approach. The intuition here is that the model is built sequentially so that it never borrows information from the future. Using a k-fold approach, each fold will represent a sequential year. Each successive fold is a superset of the previous fold, thus ensuring that only past information is used. For instance, for each entity, we might aggregate features in 2011 and 2012, train on outcome data from 2013, and then validate/test on outcome data from 2014. (for Data Science and at the University of Chicago, 0). Figure 2.13 illustrates the basic logic of temporal cross validation.

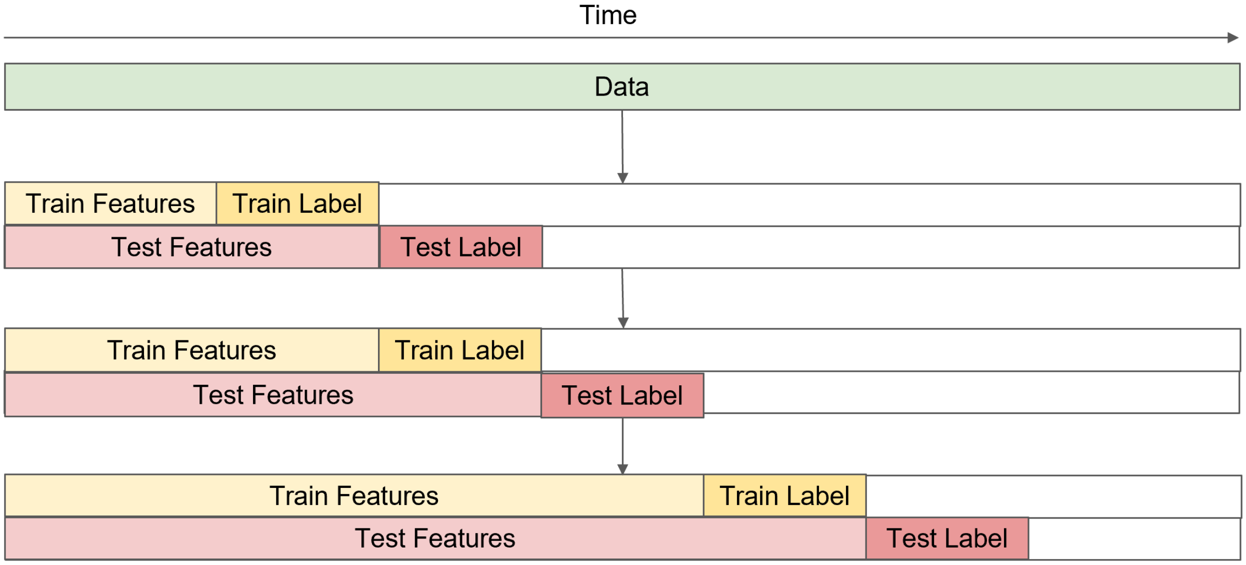


Figure 2.13: Illustration of Temporal Cross-Validation

Over and Undersampling

The major problem with predicting cybersecurity incidents is that although they are costly, they are relatively rare. In machine learning terms, this translates to an imbalanced learning problem. Essentially, instances of the majority class (”no breach”) vastly outnumber instances instance of the minority class (”breach”). Thus, if an algorithm was trained to optimize only for accuracy, it would do quite well by simply picking the majority class every time. From a policy perspective, optimizing for accuracy alone is not always fruitful because regulators are oftentimes concerned with detecting and preventing rare but significant events.

One way to overcome this problem is to utilize over- and under-sampling techniques. Oversampling takes instances of the minority class and upsamples them in the training process, whereas undersampling takes instances of the majority class and downsamples them. Oversampling comes at the cost of potentially overlearning idiosyncrasies in minority class, and thus generalizing poorly. Undersampling comes at the cost of throwing away potentially relevant and useful information, thus reducing the algorithm’s overall accuracy.

In this application, I combine over- and under-sampling together. Combining both helps capture some of the benefits of each, while ameliorating the disadvantages of each. In future iterations, I may look to other techniques such as Synthetic Minority Oversampling Technique (SMOTE) and Random Oversampling Examples (ROSE) instead of a simple oversample. SMOTE in particular may yield benefits as it avoids some of the overfitting problems of simple oversampling.

2.7 Results

Although more work is necessary before deploying a model in this context, early results are promising. Compared to random audits or no audits, algorithmic predictions more successfully target risky companies and industries. In this section, I present the results of various model configurations. I present a baseline model with just firm-level information, a model trained only on text, a model that combines both firm-level information and text, and a final model that selects the most predicitve firm-level features, and discards unimportant features. In general, combining firm-level data and text features works best for prediction, and reducing model complexity aids with improving precision.

Baseline Results

The first model I present is a baseline model that uses only firm level features. These include the features described in Section 2.4. Figure 2.14 shows the results for this baseline model without any additional text features. I use logistic regression, poisson regression, classification tree, random forest, gradient boosting classifier, adaptive boosting methods. Although logistic regression performs quite well on recall (the ratio of firms predicted to be breached over the firms that were actually breached), this performance comes at the expense of accuracy (ratio of correct predictions to incorrect predictions) and precision (ratio of correct predictions to correct plus incorrect predictions). Essentially, the logit model here too aggressively guesses the positive class (”breach”) in this case. The tree-based methods trade off some of this recall for more precision, though still are not as precise as we might hope for in a policy application. At best, the tree-based methods achieve around a .1 precision. While recall is more important in this application, too low a precision score implies that the SEC would erroneously flag too many audit candidates. Given limited resources, enough to conduct about 75 audits per year, flagging too many candidates potentially misses some risky targets.

Figure 2.15 shows feature importances from the random forest model. As suggested by the exploratory, industry riskiness is an important feature for predicting breaches. Proxies for firm size such as market value and tax liabilities are somewhat predictive, as are measures of stock volatility. Notably, only a few indicators for industry (software, retail) and geography (New York and Chicago) are predictive, with other dummy variables for these values taking on 0 or very low feature importance.

Text Only Results

Next, I show models with only text features in Figure 2.16. These models exclude any other firm-level information, as well as the manual feature-engineering of key terms corresponding to SEC interpretive guidance for describing cybersecurity risks. These results underperform the baseline models, regardless of the particular model chosen. In general,

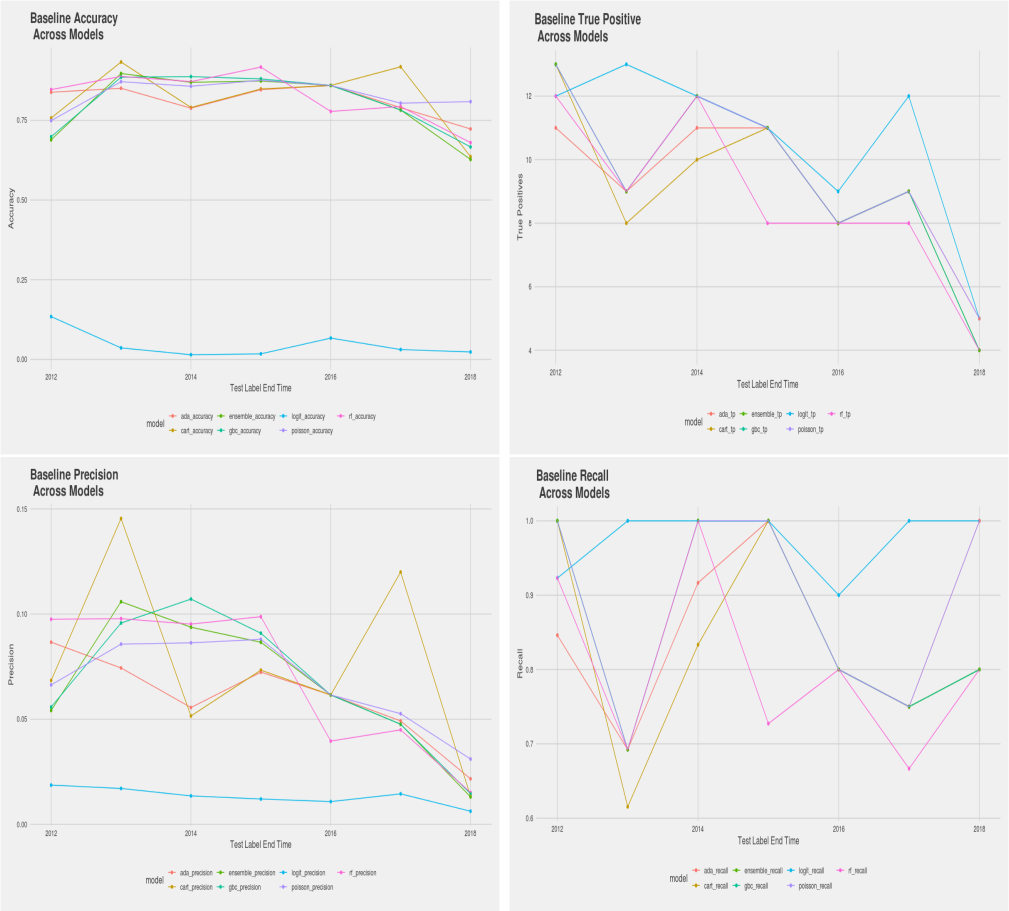


Figure 2.14: Baseline Results

this result is not surprising. That being said, the text alone does seem to be somewhat informative.

Baseline + Text Results

Combining the baseline features with text features performs similarly to the baseline alone. While some models make some gains on recall, this may just be noise. Precision also seems to be a bit lower across models. Figure 2.17 illustrates these results. Again, these results are driven by including all possible features in the model, potentially leading to overfitting.

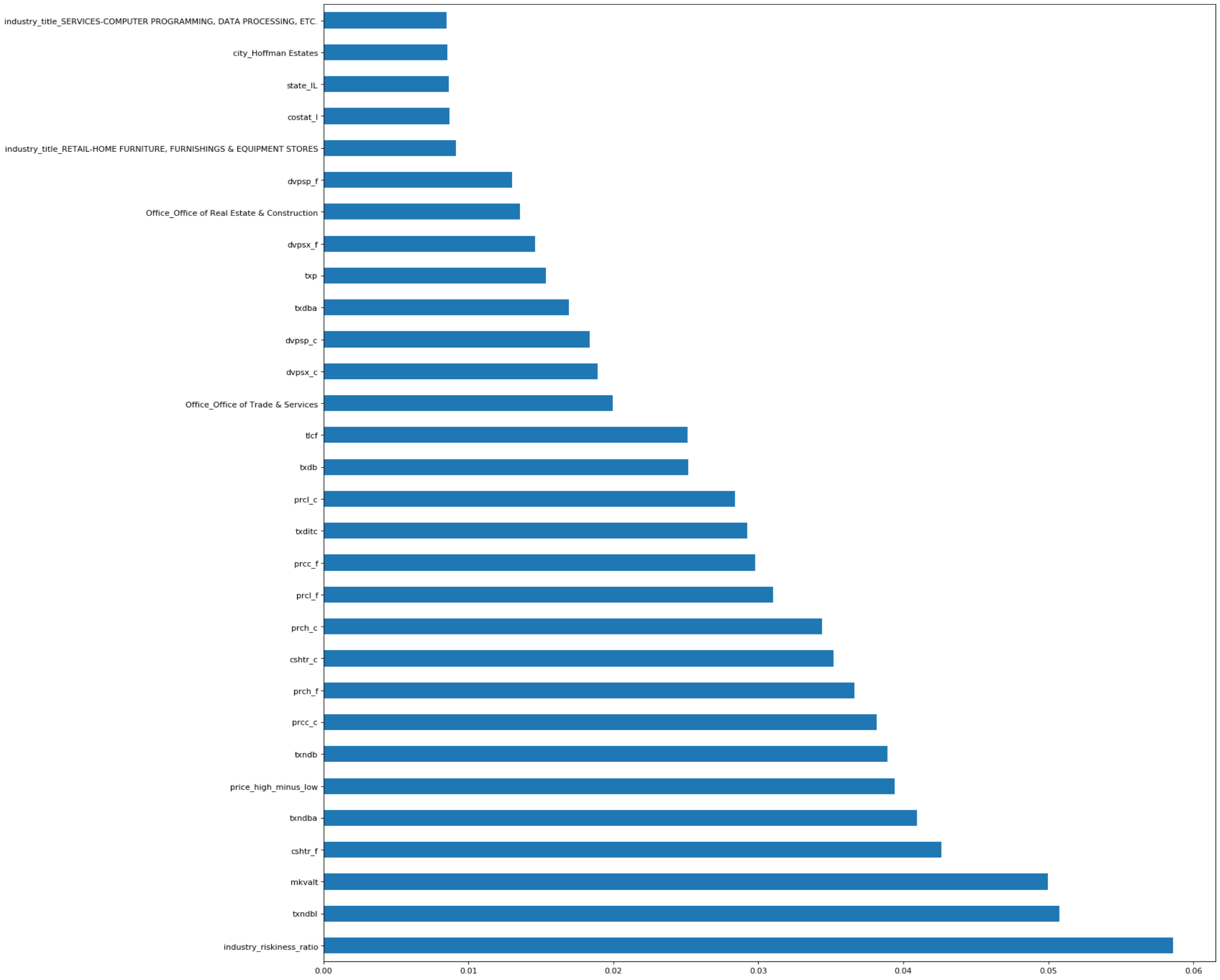


Figure 2.15: Feature Importance in Baseline Random Forest

Selected Features Results

For the final models, I remove the unimportant firm level features from the baseline features and retrain each model. Removing these features and retraining the models considerably improves precision, though at the cost of some recall. Figure 2.18 illustrates this tradeoff. Looking at random forest, gradient boosted classifier, and adaptive boost, precision improves to about .4 in most years, though recall drops to about .5. In this context, this tradeoff is probably worthwhile as the higher precision suggests that the models are more judiciously picking good candidates for audits, rather than flagging a broad range of possibilities that exceed regulators' auditing capacity.

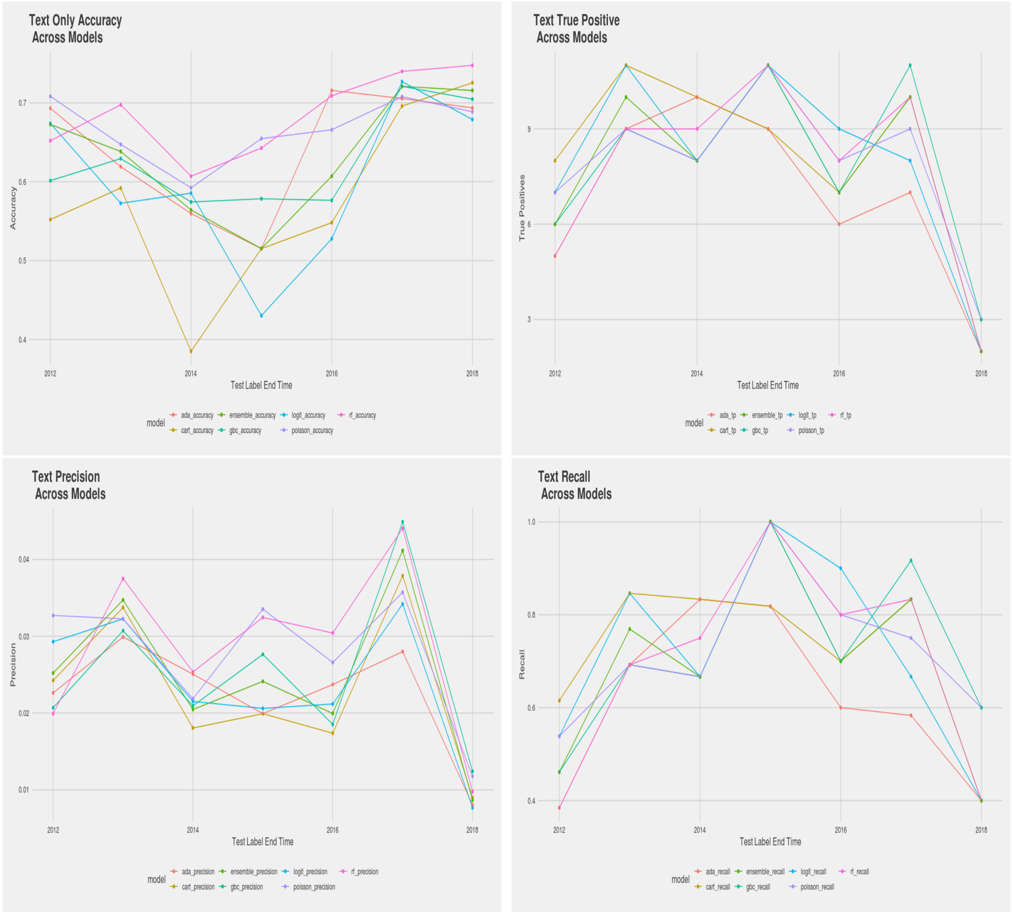


Figure 2.16: Text Only Results

2.8 Discussion

Precision-Recall Tradeoff in Predictive Auditing

While this study is specific to cybersecurity, it speaks to a larger problem in law regarding government auditing to detect rare events. Governments frequently employ auditing as a tool to ensure that private actors are complying with regulations. In U.S. federal law, some common examples include Internal Revenue Service tax audits, Department of Labor fair labor standards audits, and Federal Emergency Management Agency disaster relief audits. These audits commonly target underlying activities that occur infrequently among legitimate activities. Most people adequately report and file their tax liabilities, most employers comply with fair labor standards, and most recipients of FEMA funds properly administer those funds. Indeed, a tiny percentage of each of these activities constitutes the sort of fraud

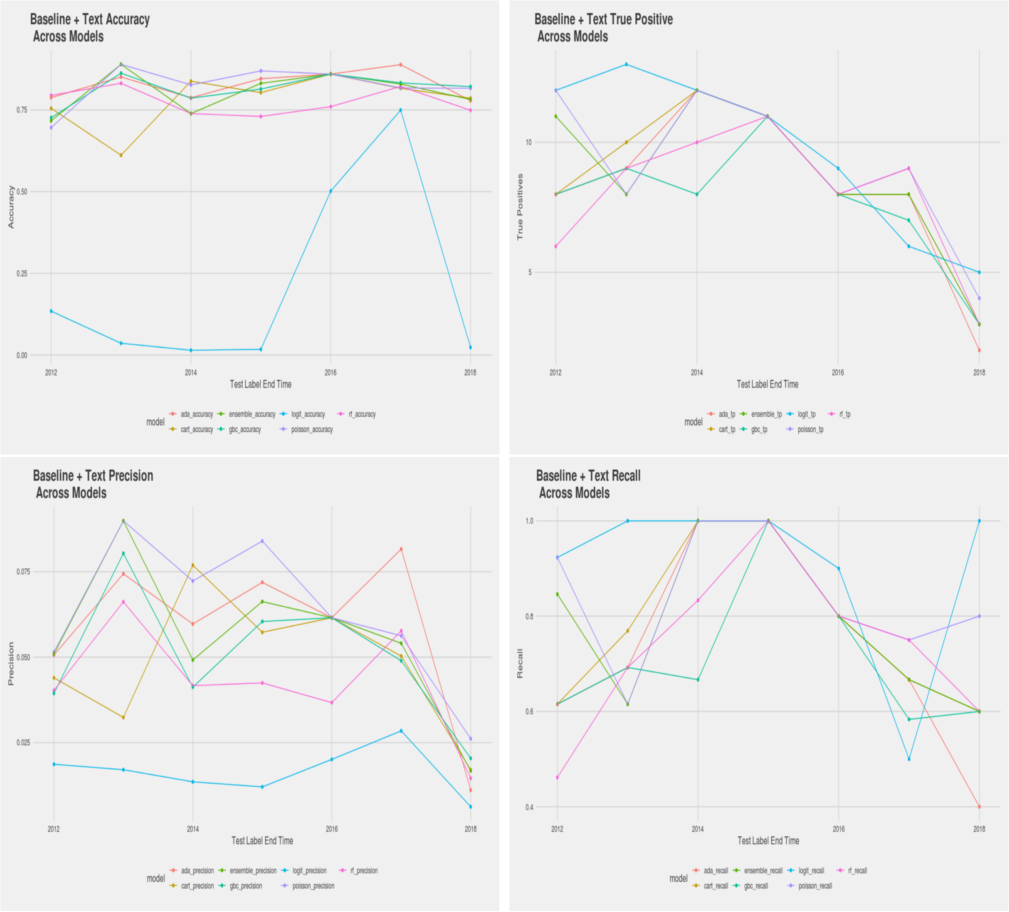


Figure 2.17: Baseline + Text Results

or vulnerability that these audits are designed to uncover. Detecting these rare events is a problem because the government has limited resources to conduct audits. Given these constraints, governments may be concerned with ensuring that auditing activity is directed towards undesirable activities.

In machine learning terms, this problem is best conceptualized as an imbalanced learning problem. Imbalanced learning refers to imbalance in the outcome variable of a dataset. In this cybersecurity context, the negative class (”no breach”) swamps out the positive class (”breach”), as approximately only 2% of firms experience a breach each year. The core problem with imbalanced learning problems is that accuracy can be optimized simply by guessing the dominant class every time. However, when used to make to an actual decision, this type of model would not be useful. There are technical approaches to imbalanced learning problems, such as random over- and under-sampling as employed in this study. Thinking more broadly about how to map metrics to a policy context is also an important

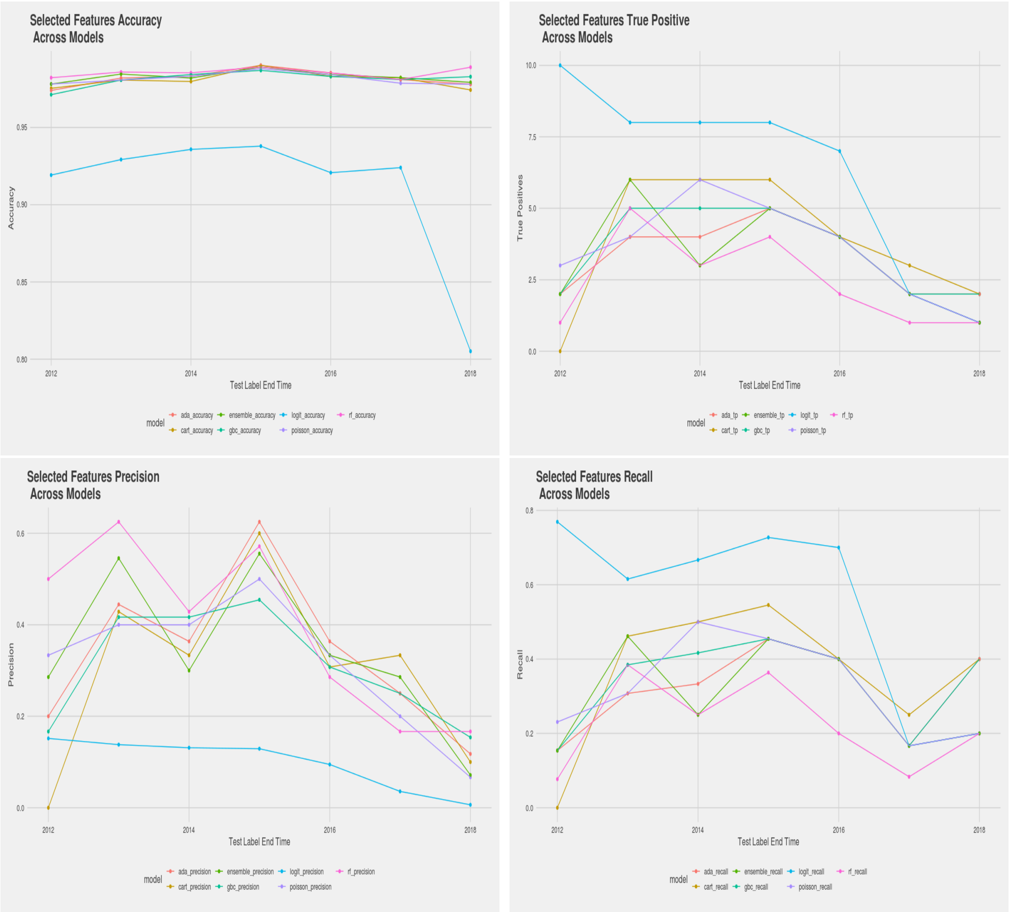


Figure 2.18: Selected Features Results

step though.

In policy contexts, precision and recall become relevant measures, but there is a tradeoff between them. One could achieve a perfect recall (finding all possible breaches) by assuming that every observation is a breach. However, the precision of this model would be quite poor, and if a government agency had the resources to audit every firm then an algorithmic approach would not be necessary. Similarly, a model could be very conservative and only make one guess about firms likely to be breached, and if that guess is correct, it could stop making predictions. While this approach would yield a perfect precision, it would miss many relevant cases, and again not be helpful for regulators who are trying to find the riskiest companies. This concept holds outside the cybersecurity context as well, as regulators frequently are implicitly optimizing the precision-recall tradeoff when targeting their auditing activities.

Framing the precision-recall tradeoff as part of a policy decision can help a decisionmaker determine the optimal amounts to trade off on each metric. In this cybersecurity context, a policymaker may prioritize maximizing true positives, maximizing recall, and minimizing false negatives, while tolerating weaker precision and a high number of false positives. These priorities are plausible because false negatives (failing to detect a breach) are more costly than false positives (auditing a firm that was not going to be breached). Similarly, recall (finding all potential breaches) may be more important than precision (the fraction of flagged firms that are actually breached). That being said, this tradeoff does not suggest optimizing these quantities by totally sacrificing precision for recall. Rather, contextualizing the tradeoff within the SEC's actual auditing program can help illuminate how policymakers should use these metrics.

To illustrate, assume that the SEC's auditing capacity is fixed at 75 audits per year. It will not conduct fewer than 75 audits even if doing so would be cheaper, nor does it have the resources to conduct more in a given year. Within these constraints, the SEC must optimize where to place these 75 audits to attempt to successfully detect companies that will be breached. Given that the number of audits is fixed, the precision is somewhat irrelevant. If a model flags 20 potential breaches, but only 5 were actually breached, the precision would be .25, but the audits of the 15 non-breached firms do not represent any marginal cost to the agency. Thus, the SEC may prioritize recall instead because it wants to make sure most of the riskiest firms do end up in the audit pool, as it will not have additional resources to audit those firms if they are not flagged. In situations where an agency wishes to conserve resources by reducing the number of audits, or the number of audits it makes is unbounded, prioritizing precision may be more sensible.

Figre 2.19 shows this tradeoff in the cybersecurity context. Using the predicted probabilities from the gradient boosting classifier model, it plots the precision-recall tradeoff. The "Random Audit" model guesses firms to pick for audits at random, and this model does quite poorly on precision. The GBC model on the other hand correctly flags several breached firms before guessing incorrectly. Importantly, while precision drops considerably once recall reaches about .5, auditors need not stop at that point. With 75 audits, the SEC could conduct audits up to a recall of about .71 before running out of resources. Thus, while auditors would tradeoff a considerable amount of precision with additional audits, doing so is not necessarily fatal to the enterprise as there are resources to spare in this case.

Simulation on Real-World Outcomes

I conclude with an illustration of how an algorithmic auditing approach improves upon a randomized approach. Figure 2.8 shows how many breached firms would be detected in advance for the 2015 fiscal year across 100,000 simulations. Occasionally, a regulator picking firms at random might find one breached firm, and rarely would find two. In most simulations, a random search would not yield any members of the positive class.

Figure 2.20 demonstrates the utility of an algorithmic approach over a randomized one. Using the assumption that SEC audits are totally effective at deterring a potential breach, it illustrates the potential reduction in breaches each year. Assuming 75 audits are conducted

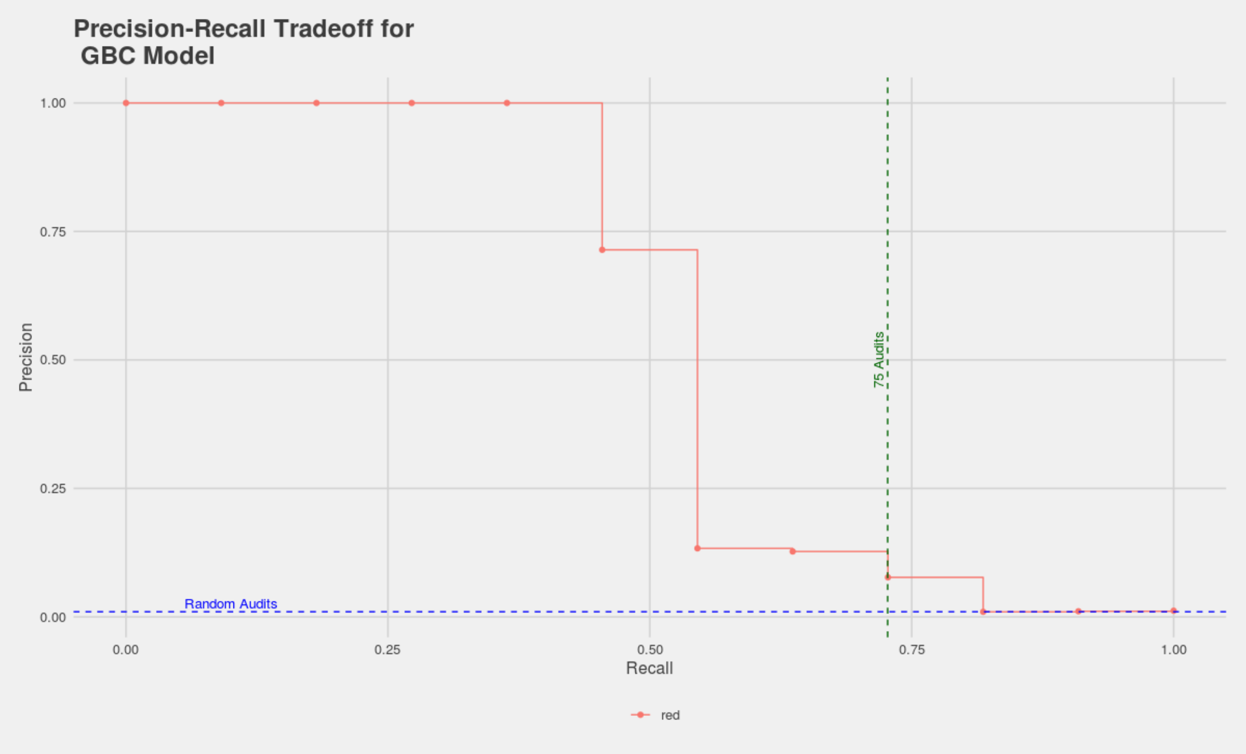


Figure 2.19: Precision-Recall Tradeoff

in each given year, we see an average reduction in breaches of about 18%. In the 2015 fiscal year, of the 24 breaches in the dataset, 5 are flagged in advance.

As seen in Figure 2.18, using the final models that select out unnecessary features, in most years the models achieve both recall and precision in the neighborhood of .4. While a poor precision score would generally be a problem in most machine learning applications, these results are actually quite promising when contextualized as a public policy problem. Although regulators would need to sift through several companies that are unlikely to be breached, the high recall suggests that they will eventually find companies that would have been breached and can act to bolster their cybersecurity practices. Most importantly, the algorithm eliminates a huge number of companies that it is confident will not be breached, thus saving regulators time and allowing them to focus their regulatory efforts on a smaller subset of companies. Table 2.5 illustrates this point with sample results from the ensemble algorithm's 2015/2016 predictions. A regulator could be furnished with a list that safely eliminates several companies from consideration, while allowing them to focus on the likeliest breach targets.

The main normative takeaway for legal scholarship as a whole is that there is value to prediction. There is currently a live debate within law and law-adjacent literatures about the use of machine learning and prediction in legal contexts. Much of the attention thus far

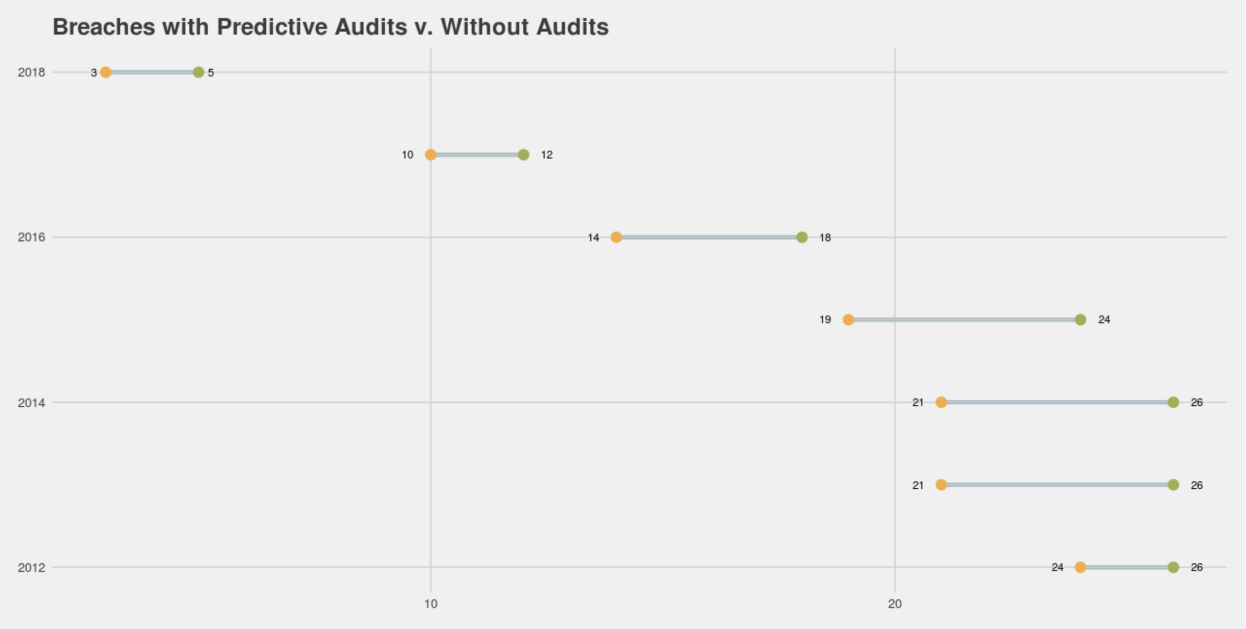


Figure 2.20: Breaches With and Without Predictive Audits

has understandably been placed on applications where decisions involve vulnerable populations and legally protected classiciations like race and gender. Thus, many of the examples focus on areas like employment, housing, and criminal law. This scholarly debate would be enriched by considering applications that do not implicate the same equity concerns. In this case, predicting the cyberriskiness of corporations shares little similarity with the aforementioned examples on equity and fairness grounds. Instead, improving auditing efforts only improves efficiency, and is beneficial to regulators, corporations, and the public alike. Audits themselves are not costly for audited firms. While some firms may bear more of the costs of precaution, this allocation is sensible if they carry more of the risk. [Cite barocas/selbst, coglianese, lehr]

Simplifying Decisionmaking

Regulators may also choose to deploy simpler models that are more easily explained to outside stakeholders. Certain firm-level features are more predictive than others. For instance, a firm's industry's riskiness, location (New York, California, or Illinois), and stock volatility can be used to construct simple decision rules. These models can also incorporate flags for whether a firm's disclosure contains elements from the SEC interpretative guidance, and build a simple model to guide auditing decisions. For complex policy decisions, simplifying models can help with conveying the reasoning behind a legal decision. Simple models may sacrifice performance on certain metrics, but the added advantage of interpretability

COMPANY NAME	filing date	breach	pred
hyatt hotels corp	2/18/2015	yes	yes
target corp	3/11/2016	yes	yes
chevron corp	2/25/2016	yes	yes
microsoft corp	7/31/2015	yes	yes
tennessee valley authority	11/20/2015	yes	yes
apple inc	10/28/2015	yes	yes
monster worldwide inc	2/11/2016	no	yes
iron mountain inc	2/26/2016	no	yes
quest diagnostics inc	2/26/2016	no	yes
commercial metals co	10/30/2015	no	no
medallion financial corp	3/11/2015	no	no
marriott international inc	2/19/2015	no	no

Table 2.5: Sample Results for Predicting 2016 breaches. The ”breach” column indicates firms that were actually breached in 2016, ”pred” indicates firms that were predicted to be breached in 2016. Blue indicates a ”breach” value and red indicates a ”no breach” value.

and ease of construction could be worthwhile. Jung et. al. detail this logic in depth. They argue for this ”select-regress-and-round” approach. They advocate for a pipeline where the analyst builds a complex model that serves as a benchmark, and then create simple rules to test against both this benchmark and human decisions. They highlight the use of simple rules in judges making bail decisions, and note that simple rules both outperform human judges and come close to complex models like random forests (Jung et al., 2017).

In the cybersecurity context, we can see the value of this framework by using a decision tree and benchmarking it against the ensemble model. Figure 2.21 illustrates a classification tree built with these features alone on the same 2015-2016 period used above. The basic logic of the tree makes splits based on market value and industry riskiness to make predictions about whether a particular observation is a ”breach” or ”no breach.” Although the model does slightly worse on recall than more complex models, it still does relatively well and is much simpler to visualize. In lieu of the more complex models used earlier, the SEC could choose to use a simple classification tree with some manually engineered features to achieve comparable results. Importantly, this simple model still avoids the accuracy trap of flagging everything as ”no breach,” and the recall trap of flagging everything as ”breach,” as seen in Table 2.6.

One way to approach this problem would be to start with the more complex models described above, and then map their complex decision rules to simple ones for deployment in practice. The exploratory analysis and complex modeling helped surface insights into which features were genuinely informative, the types of mistakes that different modeling choices would lead to, and the best possible performance of a model in this context. From these complex models, it is possible for a regulator to narrow down the features to prioritize, and focus on creating a decisionmaking pipeline that utilizes that simpler information (as I do here with a classification tree). This simpler model can be used to explain the process and justification for important legal decisions, even if more complex models were fit first. Critically, these models should not be static. Observing real-world outcomes, adjusting regulations, and retraining models should be a dynamic process that informs human decisionmaking in policy contexts, not replaces it.

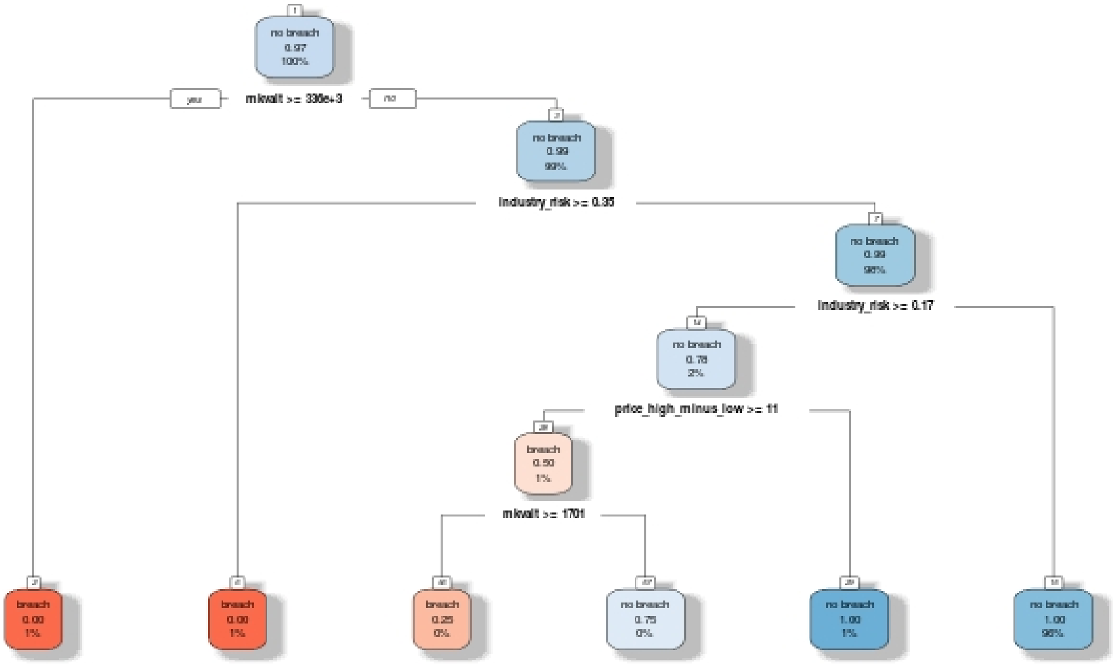


Figure 2.21: Classification Tree

	observed breach	observed no breach
predicted breach	9	13
predicted no breach	9	872

Table 2.6: CART Confusion Matrix

2.9 Future Work

There are several areas of improvement for future work to iterate upon these results. These results are drawn from matching outcome data from public databases to publicly traded companies. However, this construction is not complete. For instance, the SEC flagged 87 breaches in 2017, compared to the approximately 20 breaches I found for the same year by manually cross-checking publicly reported breaches to companies in the dataset. Resolving these inconsistencies would help bring the models closer to the ground truth, and likely help the class imbalance problems as well.

A qualitative component that includes discussions with SEC auditors, firm managers, and in-house cybersecurity personnel would also be helpful. Many of the assumptions about how managers word their cybersecurity disclosures and report their cyberrisks are based on theoretical reasoning and second-hand sources. Gaining more insight into how disclosures are actually crafted, and how companies think about their own cybersecurity postures, would be tremendously helpful in building better models. Moreover, speaking to regulators and learning what their priorities are would help determine which metrics to prioritize, and how to target audits. In particular, gaining more insight into the exact mechanism underlying the current choice of firms to audit would help establish a realistic baseline beyond random audits.

Finally, a field experiment that validates the modeling would be invaluable. While the temporal cross-validation provides some evidence of how the model would have worked historically, this is not a guarantee of performance in the future. Randomizing firms flagged by the model and observing differences in breach rates would validate the model's predictions. Creating an interplay between training new models and real-world testing will ensure that the models stays up-to-date and usable. Most importantly, targeting interventions at the firms most likely to benefit from them creates an opportunity to assess the causal effect of the audits themselves, and reevaluate the SEC guidelines and audits in light of quantitative evidence.

2.10 Conclusion

Like with many policy areas, privacy and cybercrime scholarship has traditionally focused on the theoretical underpinnings of causation. This study looks to expand the traditional scholarship by reframing cybersecurity as a prediction policy problem. Predicting incidents before they occur gives policymakers and organizations many more opportunities to prevent the privacy harms that stem from massive data losses. Prevention would be more effective than restitution, and tools that can aid in this goal would reshape the current discourse around data protection law that focuses mainly on harms.

If successful, this study could also bolster current efforts to incorporate artificial intelligence and data science into regulatory efforts. Mandatory disclosure is a commonly used and powerful legal mechanism for ensuring better institutional behavior. Scholars and policymakers have extolled the virtues of disclosure for decades. New computational tools potentially allow us to harness not only the fact that a disclosure is made, but the actual content of a disclosure. Incorporating data science into the framework of disclosure law could spur a flurry of innovative scholarship. Tools that make sense of the massive amount of text generated by mandatory disclosure can improve regulatory efforts, increase consumer information, and promote healthier corporate behavior.

Chapter 3

The Effect of Data Breach Notification Laws on Medical Identity Theft

3.1 Introduction

Identity theft and data breaches are becoming more common, and consequently, cybersecurity is quickly coming to the forefront of policy discourse. Data breaches can compromise consumer information related to things such as transaction history, payment information, health data, and personally identifiable information (PII). The financial and reputational harms that stem from these kinds of losses can be large in magnitude, but also difficult to detect because the consequences of identity theft do not materialize immediately.

Despite the growing incidents of data breaches and identity theft, there is no single federal law that regulates how an organization must respond to a data breach once it has been discovered. There are sectoral regulations that affect certain industries, however. One such sectoral regulation is the Health Insurance Portability and Accountability Act (HIPAA) that regulates healthcare information. One of its provisions requires that medical providers and health insurance companies provide notice to the department of Health and Human Services (HHS) and their data subjects when unencrypted data is breached.

States may adopt stricter requirements on top of the federal requirements. Beginning in the mid-2000s, nearly every state adopted data breach notification laws for all organizations that maintain unencrypted data from various sectors. In 2016, California amended its existing data breach notification law to mandate disclosure of breaches even when medical data is encrypted. Between 2016 and 2019, Illinois, Nebraska, New York, Rhode Island, Maryland, Delaware, and New Mexico explicitly included protected health information (”PHI”) in their definitions of personal information. Some states are now moving toward creating private causes of action following data breaches.

Despite the popularity and continued adoption of state data breach notification laws, there is relatively little evidence on their efficacy. The two main theories underlying data breach notification laws are that they will encourage organizations to invest in better cybersecurity practices so they can avoid making damaging disclosures, and that they will allow consumers the opportunity to guard against identity theft once a disclosure is made. Concretely, the goal of these laws is to minimize identity theft, yet the exercise of determining whether they work is fraught with serious methodological challenges. A simple research design that looks at the number of breaches before and after the passage of a data breach notification law would be intractable because the pre-treatment figure is difficult to estimate prior to the passage of such a law. Analyzing the effect of laws on large corporations with a nationwide presence becomes tricky when such organizations are likely to default to the strictest state's standards rather than tailor different notices to citizens of different states (Bradford, 2020). [1] Identity theft is also notoriously underreported, and many victims may not discover the crime until months or years after it occurs.

To address these challenges, I use a panel dataset containing medical identity theft complaints, and analyze the effect of breach notification requirements using an augmented synthetic control approach. Specifically, I examine the effect of California's 2015 amendments that expanded breach notification requirements to include encrypted medical data. In adopting this expanded breach notification standard, California considerably raised the bar for organizations that hold critical health data. More broadly, the effects these laws have on healthcare organizations could potentially be generalized to other types of data, including financial, educational, and consumer data.

3.2 Overview of Data Breach Notification Laws

Law & Economics of Data Breaches

The classic law and economics theory of crime suggests that the law should minimize the social cost of a crime, which is equal to the sum of the harm the criminal activity causes, and the costs of preventing that activity (Cooter and Ulen, 2016). This economic rule provides insights into the optimal level of punishment that the state should set to deter crime. In contrast to tort law, criminal sanctions are justified primarily as a mechanism for deterring criminal behavior, rather than forcing perpetrators and victims to internalize the costs of their own actions.

Cybercrime vexes policymakers because cybercriminals' benefits are large, the costs of perpetrating cybercrimes are low, and the costs of punishment are high. The economics of crime suggests that a rational criminal will choose to commit crime when the utility of the crime exceeds the utility of not committing the crime. Formally:

[1] As with many other regulatory areas, privacy law may be subject to a "California Effect." The California Effect (also called the Brussels Effect when discussing the European Union) essentially describes the phenomenon where a large market is able to unilaterally impose its regulatory preferences on other markets. In this case, California privacy laws that regulate large businesses will change privacy regulation across U.S. states. See Anu Bradford's, "The Brussels Effect" for more on the general theory.

$$(1-p)\mu_s + p\mu_f \geq \mu_{nc}$$

Where:

- μ_s is the utility of successfully committing a crime
- μ_f is the utility of failing to commit a crime
- is the probability of being caught and punished
- μ_{nc} is the utility of not committing a crime

In this framework, the probability of being caught and punished, p, is vanishingly small. Cybercriminals frequently operate anonymously, outside of the local jurisdiction, and in complex networks. These facts make it difficult for law enforcement to identify, much less apprehend, cybercriminals. Thus $p\mu_f$ becomes very small because p is close to 0, meaning even large punishments (μ_f) will be ineffective deterrents. Deterrence through punishment is therefore an ineffective strategy, even if the social costs of cybercriminal activity are high.

Cardenas et. al. provide a typology and economic analysis of these dynamics in "An Economic Map of Cybercrime" (Cardenas et al., 2010). They explicate various types of cybercrime techniques including malware, botnet herding, phishing, Distributed Denial of Service (DDoS), and identity theft, among others. They also argue that estimating the social costs of cybercrime is difficult because there is little reliable data about the costs borne by companies that are the victims of cybercrime. More disclosure could help alleviate this problem. Similarly, estimating the benefits accrued by cybercriminals is difficult because of the uncertainty surrounding the monetization of cybercrime.

Identity theft illustrates some of the difficulty of estimating the benefits of cybercrime. Creating the necessary infrastructure to properly impersonate someone for these purposes is quite laborious, and mitigates the expected benefit of identity theft. Conducting identity theft at scale becomes extremely difficult because of the elaborate process involved with impersonating even one person. Thus only a fraction of the records compromised in a data breach will be successfully used for conducting identity theft. Although the benefits to criminals are ameliorated by the complexity of the apparatus involved, the problem of estimating the benefits is still difficult because of the lack of consistent data.

These dynamics make the law and economics of cybercrime slightly different than conventional accounts of the law and economics of crime in that law focuses more on the benefits than the costs of the crime. In particular, with such a low probability of punishment, estimating the optimal level of punishment to achieve deterrence would lead to an implausibly high figure for legal sanctions. In practice, cybercrime law instead focuses on deterrence by denial, specifically by reducing the benefits of cybercrime. For instance, if breached companies provide their consumers with identity protection services, the benefits of identity theft will be reduced, thus making cybercriminal activity less profitable.

This focus on reducing benefits motivates data breach notification laws. Because sanctions against perpetrators are ineffective, regulatory attention is instead placed on organizations that collect and hold data. By requiring disclosures from breached companies, policymakers aim to nudge organizations to invest in cybersecurity and give consumers adequate opportunity to safeguard their identities. These goals both serve to reduce the potential benefits that might be realized by a cybercriminal by making stolen credentials useless for impersonating someone.

Legal Background

As of 2018, every U.S. state now has a data breach notification law. These laws share several similarities, but there are also some key variations that are important to note. Generally, data breach notification laws contain the following elements:

- **Definition of Personal Information**: Examples of personal information that is covered by the law. Some example include email addresses and passwords, health information, driver's licenses, federal identification numbers, and biometric data.

- **Covered Entities**: Entities that must comply with the law. Typically, all government agencies, businesses, and non-profits are covered.

- **Encryption Safe Harbor**: Whether organizations are exempt from disclosure requirements if the breached data was encrypted. Almost every state provides a safe harbor for encrypted data.

- **Notification Trigger**: The threshold that triggers a notification obligation. Examples include "substantial harm to individuals," "reasonable likelihood of harm," or "awareness of breach." Under the first two standards, organizations only need to provide notice if they believe there will be harm to their data subjects, whereas standards closer to "awareness of breach" remove this discretion.

- **Content**: The content of the breach notification. Most states do not mandate any specific content, while others regulate the information that must be provided, including things like descriptions of the incident, types of personal information compromised, and toll-free numbers for consumers to call.

- **Timing of Notification**: How long an organization has to provide notice once it has discovered a breach. Common requirements are that notice must be provided within 30, 45, 60, or 90 days, though some states do not specify a timeframe at all.

- **Penalty** The civil penalty associated with failure to comply with the requirements of the law. Some states penalize by days over the time limit, while others penalize by number of individuals affected.

- **Cause of Action**: Whether consumers have the right to bring a cause of action following a breach notification.

- **Notice to State AG or CRAs**: Some states include provisions that require additional notice to the state Attorney General and/or consumer credit reporting agencies.

While states may differ on some of these specific requirements, much of the legislation shares common language and policies. These similarities may indicate some degree of policy diffusion across state jurisdictions. This diffusion is helpful primarily because states share a remarkably similar common baseline for breach notification requirements, and researchers can analyze the differences that arise from particular policy choices that are layered on top of this common baseline.

In 2016, California amended its existing breach notification standards to change the "content" aspect of its law. An example disclosure that conforms to this law can be seen in Figure 3.1. In particular, it requires that the notice has the following headings:

- Subject Line that says "NOTICE OF DATA BREACH"

- **What Happened**: A description of when and how the breach occurred.

- **What We Are Doing**: Actions that the organization is taking to mitigate potential harms.

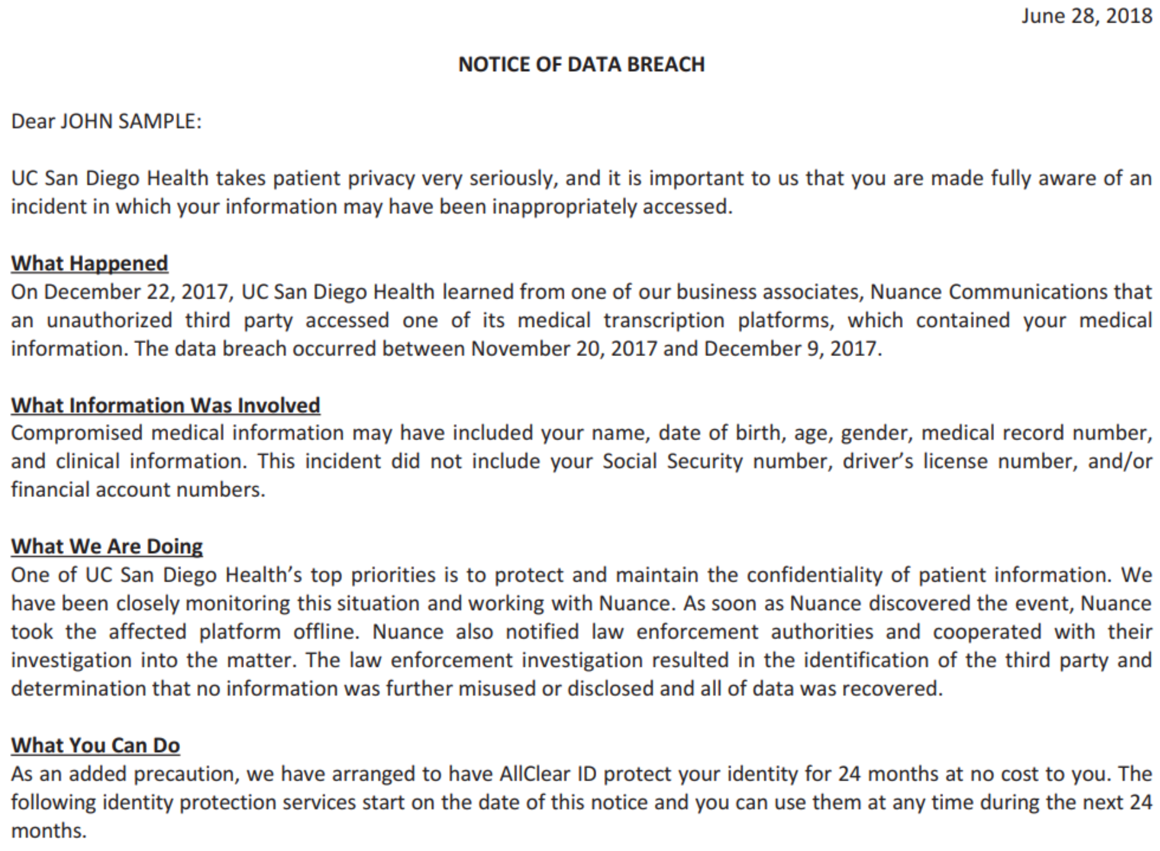
- **What You Can Do**: Suggestions for how data subjects can safeguard their identity.

- **If You Have Questions**: Contact information and toll-free numbers for consumers to call with concerns.

These requirements essentially mandate that all disclosures provide certain content and are organized in a specific way - there is little room for an organization to change the style mandated by law. Thus, this amended law provides a good vehicle for exploring the effect of mandated disclosure on reported medical identity theft.

3.3 Literature Review

Data breach notification legislation has been explored in a few different pieces, but is largely an understudied area of law. In part, this may be because data breaches are a relatively new phenomenon. Furthermore, despite their costliness to victims, they are still somewhat rare events. Moreover, companies may not even be aware that they have been the victims of a cyberattack, and the consumers who lost data may be unaware of identity theft for months of years after the breach. Finally, absent legal mandates, they are likely to be underreported, as victimized companies have strong incentives to not report or delay reporting serious incidents.



June 28, 2018

NOTICE OF DATA BREACH

Dear JOHN SAMPLE:

UC San Diego Health takes patient privacy very seriously, and it is important to us that you are made fully aware of an incident in which your information may have been inappropriately accessed.

What Happened

On December 22, 2017, UC San Diego Health learned from one of our business associates, Nuance Communications that an unauthorized third party accessed one of its medical transcription platforms, which contained your medical information. The data breach occurred between November 20, 2017 and December 9, 2017.

What Information Was Involved

Compromised medical information may have included your name, date of birth, age, gender, medical record number, and clinical information. This incident did not include your Social Security number, driver's license number, and/or financial account numbers.

What We Are Doing

One of UC San Diego Health's top priorities is to protect and maintain the confidentiality of patient information. We have been closely monitoring this situation and working with Nuance. As soon as Nuance discovered the event, Nuance took the affected platform offline. Nuance also notified law enforcement authorities and cooperated with their investigation into the matter. The law enforcement investigation resulted in the identification of the third party and determination that no information was further misused or disclosed and all of data was recovered.

What You Can Do

As an added precaution, we have arranged to have AllClear ID protect your identity for 24 months at no cost to you. The following identity protection services start on the date of this notice and you can use them at any time during the next 24 months.

Figure 3.1: Example of a Breach Notification Under the 2016 California Amendments from UC San Diego

The major obstacle to empirically studying identity theft is lack of data. Chris Hoofnagle describes this problem as one of making known unknowns known (Hoofnagle, 2007). One issue with studying the extent of identity theft is that much of the collected data is drawn from surveys that are fraught with sampling errors. Without comprehensive data on identity theft, studying the effect of interventions becomes difficult. In another piece, Hoofnagle addresses this issue by using FTC Consumer Complaint Data (Hoofnagle, 2008). Here, he argues that identity theft victim data provides evidence that different financial institutions have different identity theft protection practices, and therefore different footprints. This evidence suggests that there is a plausible market for identity theft and fraud protection, and that government interventions can be targeted at a handful of firms with outsized identity theft footprints. Unfortunately, as discussed later, the full FTC Consumer Complaint Database is no longer available to researchers.

One such intervention is the ”data breach notification law.” States have adopted breach notification laws under the theory that breach notification will create incentives for companies to preemptively invest in cybersecurity measures. The idea is that a company that is forced to disclose a cybersecurity incident will face consumer backlash (Romanosky et al., 2011). Hoping to avoid such a consequence, companies will invest in cybersecurity beforehand so as to avoid the possibility of hurting their image. Additionally, should a breach actually be disclosed, and consumers are made aware of it, then consumers may pro-actively take steps to protect their identities, thus reducing the overall identity theft rate.

Sasha Romanosky and colleagues have offered the most extensive study of these data breach notification laws to date. In "Do Data Breach Disclosure Laws Reduce Identity Theft?," Romanosky et. al. directly address this issue. Using a panel data regression approach, they find that states that adopted breach notification laws experienced about a 2% decrease in per capita rates of identity theft. They included controls for state adoption of a breach notification law. They also checked for potential endogeneity issues (states adopted laws because they were experiencing a lot of cybercrime), but found this was a negligible factor with robustness checks (Romanosky et al., 2011). However, such checks are never able to provide absolute guarantees.

Another strand of the literature is concerned with how markets react to the announcement of a breach. This focus tests the notion that negative consequences stem from a breach announcement. The absence of an effect here would imply that firms do not have a strong incentive to invest in cybersecurty measures when faced with breach disclosures. Sanjay Goel and Hany Shawky, in two separate pieces, examine these market effects. In "Estimating the Market Impact of Security Breach Announcements on Firm Values," they use event study methodology to examine the effect a breach disclosure has on a firm's stock market value. In this piece, they find about a 1% decline in market value immediately following the disclosure of a cybersecurity incident (Goel and Shawky, 2009). In "The Impact of Federal and State Notification Laws on Security Breach Announcements," they use a similar event study method and conclude that after the passage of data breach notification legislation, the negative effect on stock prices is somewhat mitigated. They conclude that this implies that the legislation is effective in forcing firms to mitigate the potential damage from a cyberattack (though it could also be evidence that firms are mitigating the message sent to investors) (Goel and Shawky, 2014).

More recently, Joshua Mitts and Eric Talley in an upcoming Harvard Business Law Review piece entitled "Informed Trading and Cybersecurity Breaches," examine whether there is evidence for insider trading prior to a breach announcement. Using matched sampling to compare breached and unbreached firms, they find systematic evidence that arbitrage occurs prior to a breach announcement. This implies that there are individuals who have prior knowledge of a breach before the market does. They argue that this raises normative concerns that go beyond run-of-the mill insider trading because the harms a hacker causes a company are endogenous to the company's cybersecurity practices, and thus creates an opportunity for sophisticated arbitrage based on knowledge of a company's security vulnerabilities. This is distinct from an informed trader using exogenous information because allowing insider trading on cyberattacks effectively subsidizes hacking activity, whereas traditional insider trading is generally an exercise in price discovery (Mitts and Talley, 2018).

In the legal literature, much of the attention toward data breach notification laws has been targeted at whether a federal standard is appropriate, and what the contours of mandatory notification should look like. In ”Federal Security Breach Notifications: Policy and Approaches,” Priscilla Regan gives an overview of congressional debates around adopting a federal breach notification law during the 2000s. She notes that procedurally, the U.S. Congress faces higher barriers to enacting legislation than many states do because of its extensive committee structure, and various veto points within both Houses. Substantively, the Democratic and Republican parties had bitter disagreements about the extent to which regulators or companies themselves should maintain discretion in when a notification is necessary. At the time the piece was written (2009), Regan expressed some optimism that a federal law would be enacted, but as of 2020, no such legislation has passed yet (Regan, 2009).

Sara Needles, in ”The Data Game: Learning to Love the State-Based Approach to Data Breach Notification Law,” explicitly argues against adopting a federal standard. In her view, the fact that the states have quite easily adopted their own breach notification standards while the federal government has struggled is strong evidence that the current state-by-state approach is working well. Moreover, she notes that the various different intricacies in state law (what types of data are covered, how a breach notification should be worded etc.) will be difficult to reconcile in a federal law, and the result may be unsatisfactory (Needles, 2009).

Overall, the literature reaches a few important conclusions. First, there is tentative evidence that breach notification laws do reduce identity theft. Second, one mechanism by which this occurs, namely that consumers and investors punish breached firms, seems somewhat plausible in that there are negative effects on stock valuations in the immediate aftermath of a breach disclosure. However, the lack of enduring effects on stock prices may indicate that this market mechanism does not work well, particularly if there is insider trading. Finally, the debates surrounding breach notification span back to at least the early 2000s, but are perhaps receiving renewed attention because of recent incidents.

3.4 Data

Collecting data that on incidents of data breaches before and after the passage of breach notification law suffers from endogeneity issues. Prior to the enactment of a breach notification law, firms presumably have little incentive to report data breaches. Indeed, Romanosky et. al. showed that the number of reported breaches increases quite rapidly immediately after a state breach notification law is adopted, suggesting that there are a large number of unreported breaches that previously occurred. Assessing the effectiveness of breach notification laws by looking at the number of reported breaches would be a fruitless endeavor as the measurement of the outcome is confounded with treatment.

That being said, a plausible way forward is to use victimization rates, in particular with regards to identity theft. Since the late 1990s, the U.S. government has tracked identity theft through various law enforcement agencies, both at the federal and state levels. The gold standard for this data is the Consumer Sentinel, which is a Federal Trade Commission database that contains over 20 million self-reported identity theft complaints collected from a variety of different agencies and non-profit organizations. Unfortunately, the Consumer Sentinel is only available to law enforcement agencies. Some organizations and scholars have used Freedom of Information Act (FOIA) requests to retrieve some of this data. However, a recent case in the District Court for the District of Columbia (Ayuda v. FTC (2014)) ruled against a general FOIA request for detailed records. Because these data are no longer available, I instead use the Consumer Financial Bureau's (CFPB's) consumer complaint database. Consumers who have disputes with companies can file complaints to the CFPB, and the named companies are obligated to respond to the complaints. One such category of complaints is related to ”identity theft,” and within that, there are a number of ”medical debt” complaints. Although there will necessarily be a good deal of undercoverage, the CFPB is a major contributor to the Consumer Sentinel, and the data contain rich information about the complaints. The data go back to 2013, which precludes studying the effects of breach notification laws passed between 2003 and 2008.

The CFPB database offers the raw text of a complaint, its category, subcategory, and state that it took place. I specifically subset to medical identity theft by using the ”identity theft,” ”debt collection,” and ”medical” filters. In total, there are approximately 12,000 records. Figure 3.2 gives a sample of what these data look like, with a full example available here.

Date Received	Product	Sub-Product	Company	State
10/26/2015	Debt collection	Medical	Collection Information Bureau, Inc.	FL
10/12/2016	Debt collection	Medical	Diversified Consultants, Inc.	FL
8/15/2018	Debt collection	Medical debt	Phoenix Financial Services LLC	TX
8/22/2018	Debt collection	Medical debt	Credence Resource Management, LLC	VA
9/15/2016	Debt collection	Medical	Rash Curtis and Associates	CA

Figure 3.2: Sample CFPB Medical Identity Theft Data

Within the CFPB, I focus on medical identity theft primarily because doing so avoids Stable Unit Treatment Value Assumption (SUTVA) violations. Organizations with a multistate presence may not create fifty different disclosures to comply with each state's individual laws. Rather, they will tend to default to the strictest disclosure law because of the California Effect. Some exceptions to this general pattern may exist. For instance, Massachusetts forbids organizations from disclosing ”How it Happened.” Outside of these idiosyncrasies though, it is common for requirements adopted in one state to leak into the notices given in other states, thus violating SUTVA. Thus focusing on identity theft as a whole is unlikely to yield a credible causal estimate.

Medical identity theft differs from other identity theft in that the institutions responsible for handling data - hospitals, health insurers, etc. - typically either do not have an interstate presence or localize protected health information data storage. Indeed, policymakers and health data stakeholders are more concerned with facilitating health data transfers across state lines than preventing such transfers, because these sorts of transfers are difficult to make under the current regulatory patchwork. The Center for Medicare & Medicaid Services (CMS) has baseline rules for how stakeholders (clinics, hospitals, pharmacies, etc.) in states can exchange information across state lines and with the federal government. The federal government is piloting ”State Health Information Exchanges” where local, regional, and state governments harmonize their health data protocols to allow for easier transfer between organizations and across state boundaries. These programs may make state lines less meaningful for health data in the future, but their existence suggests that transferring medical records across those lines is quite difficult. While the inability to easily move data across state lines is a problem for medical service providers, patients, insurers, and governments, it does come with the advantage that state health privacy laws can be more plausibly analyzed without violating SUTVA.

3.5 Exploratory Data Analysis

Data breaches and medical identity theft have both grown in number over the last decade. Using data drawn from the HIPAA data breach portal maintained by the U.S. Department of Health and Human Services (HHS), I examine incidents of medical data breaches over time. U.S. federal law requires that breaches of unencrypted medical data be reported to HHS. These data thus provide a useful baseline for examining the growth in data breaches over time.

In Figure 3.3, we see the growth in reported breaches over time. While the number does seem to be declining since 2018, these figures may be subject to lag as organizations may not discover breaches for months or years after the breach occurred. More organizations may also be encrypting their data, which would reduce the amount of unencrypted data that could be stolen. Broken down by state (see Figure 3.4), there are some clear patterns that emerge. California, Texas, and Florida lead the nation in reported data breaches, and several less populous states oftentimes do not report any breaches in some years.

Examining the sources of medical breaches, we can see that medical data breaches resemble data breaches more broadly. Hacking, theft, and accidental loss are the major categories contributing to data breaches. Interestingly, hacking seems to grow as a share of the overall number of breaches as time goes on. This may indicate that while incidents of employees stealing data or losing it continue in similar numbers, the number of incidents of malicious external attacks grow over time.

As breaches become more common, identity theft consequently rises as well. Figure 3.6 shows the number of medical identity theft reports to the CFPB over time. Again, reports to the CFPB represent a sample of the total extent of medical identity theft in the country.

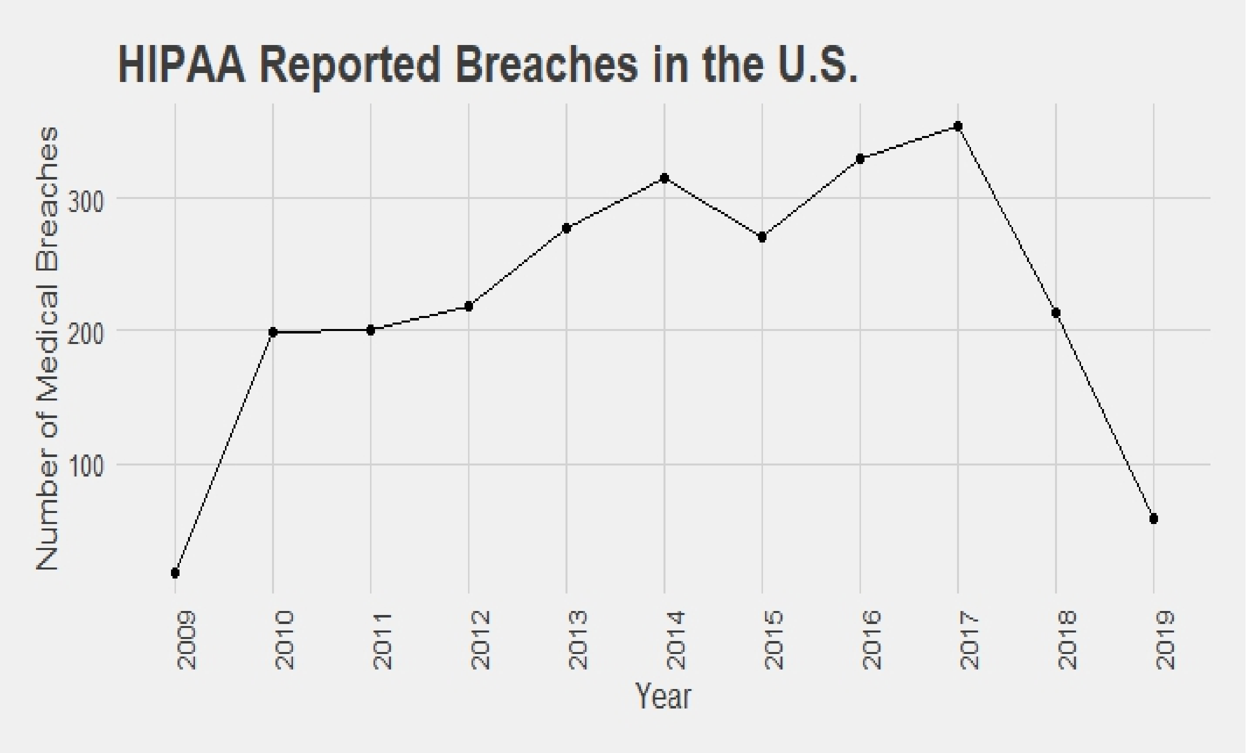


Figure 3.3: HIPAA Reported Breaches Over Time

Victims of identity theft may never report this fact to the CFPB or law enforcement. They may report it to local law enforcement, the Federal Trade Commission (FTC), state attorneys general, or credit reporting agencies instead of the CFPB. That being said, the amount of reported identity theft has grown each year, from fewer than 500 in 2013 to about 2500 in 2019.

Again, breaking down this information by state, some interesting trends emerge. Figure 3.7 shows the number of medical identity theft reports by state, scaled per 100,000 population (according to the 2010 U.S. Census). As with the HIPAA breach data, the most populous states unsurprisingly also have the most medical identity theft reports. Interestingly, this pattern holds even when scaling to identity theft theft per 100,000 people.

Virtually every state has seen the number of reported thefts increase over time, but some regional patterns emerge as well. In particular, as seen in Figure 3.8, Florida and Texas each have around 400 reports in 2019, far more than the 200 or so that California, New York, and Illinois all have. Scaled by population (Figure 3.9), the South as a whole outpaces the rest of the country. Florida and Georgia each have victimization rates close to 2 reports per 100,000 people, compared to California, New York, and Illinois with closer to .5 reports each. One possible explanation for this trend could be Florida's large share of senior citizens, and therefore Medicare recipients, however this hypothesis is undercut by Georgia's similarly

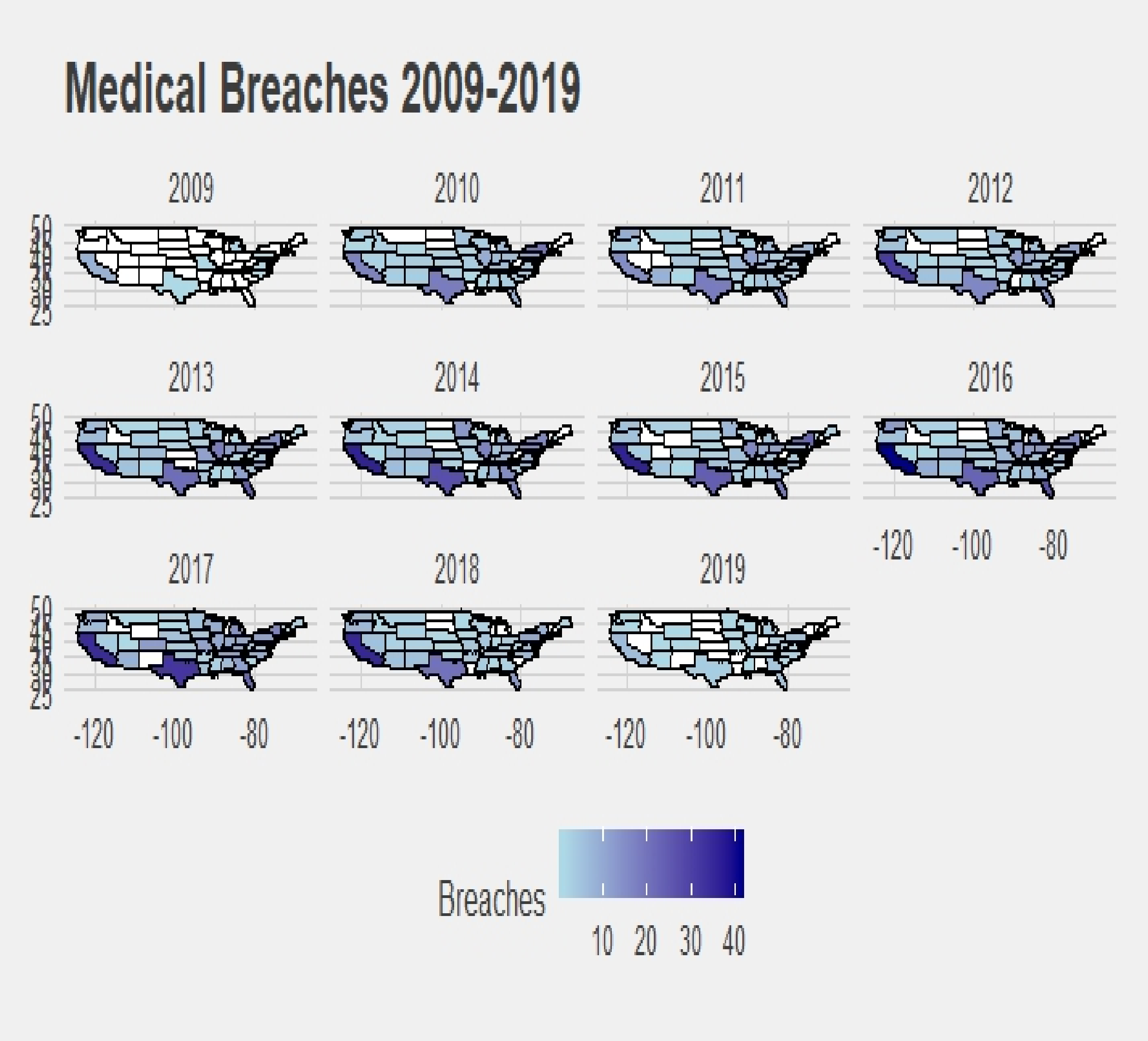


Figure 3.4: HIPAA Reported Breaches Over Time By State

high victimization rates and low proportion of senior citizens. It is also possible that people in Southern states are more likely to report medical identity theft than people in other states, however the theoretical reason why this might be the case is unclear.

These regional patterns are important primarily because they suggest that there are genuine state-level differences. Some Southern states were relatively late adopters of data breach notification laws (for example, Alabama was the last state to enact one in 2018), however this does not tell the whole story as many were early adopters and their laws shared nearly the exact same provisions as laws in other states.

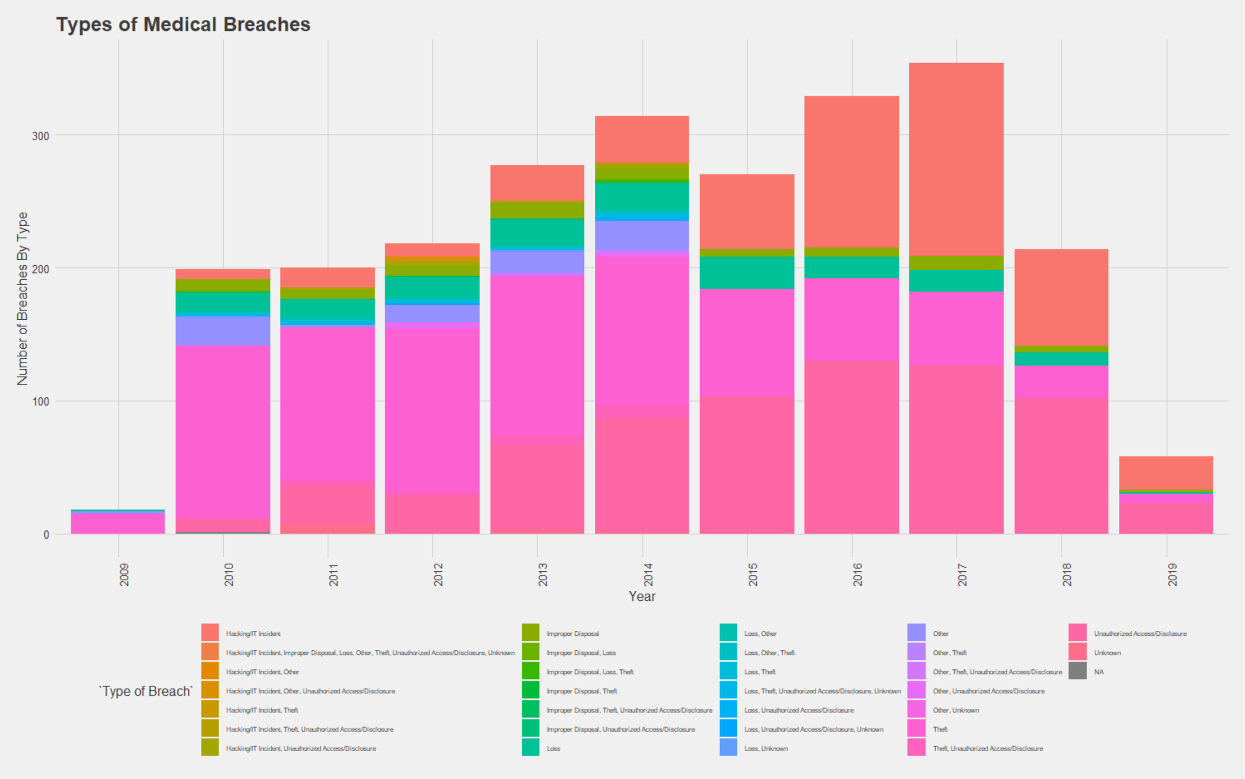


Figure 3.5: Types of HIPAA Reported Medical Breaches

3.6 Identification & Methodology

I specifically ask what is the effect of California's 2016 breach notification requirements in a data breach notification law on rates of medical identity theft. By 2008, 47 states already adopted some form of a breach notification law that covered businesses, thus making data generated after that time unhelpful in assessing a treatment effect. Also in 2008, California explicitly included health information under its definition of personal information. The U.S. Department of Health and Human Services (HHS) also has a long-standing regulation that requires processors of medical data to report breaches of unencrypted data. California's 2016 amendments expanded the notice requirements to require that they follow a particular format, be posted on an organization's website, and follow updated definitions of personal information. Thus, the mechanism that I am examining is whether a clear disclosure affects reported identity theft. Theoretically, data subjects who become aware of a breach following a disclosure may take precautions to safeguard their identities, and organizations may also be more careful about their cybersecurity practices if they know that data subjects will have such clear information about how incidents occurred.

To address this question, I employ an augmented synthetic control. Synthetic control

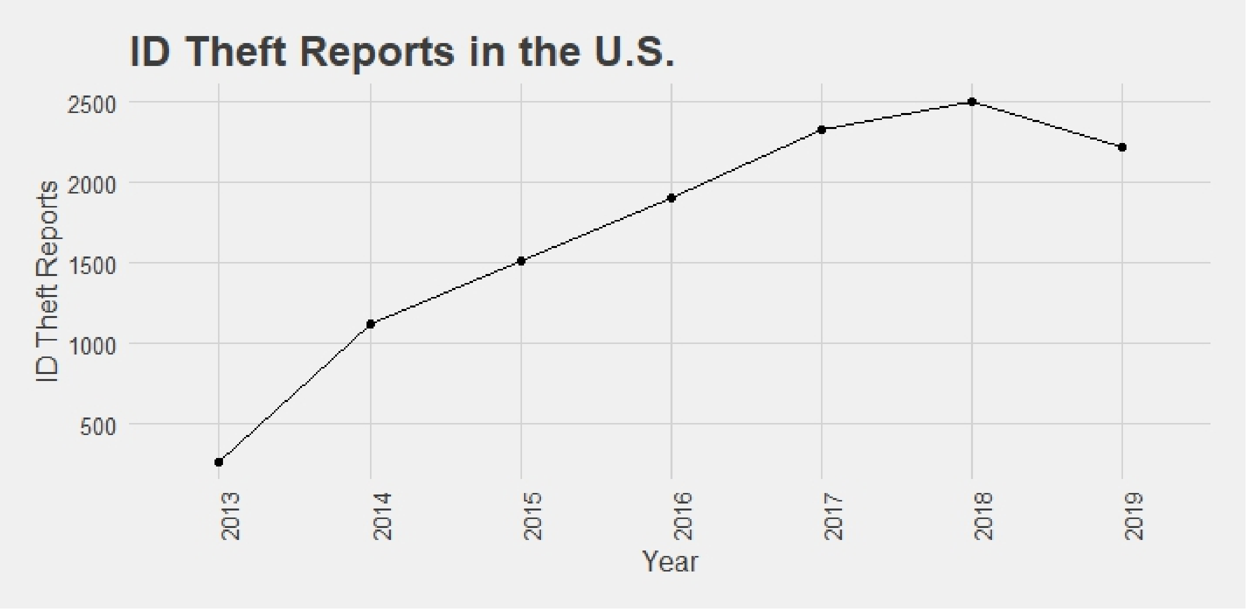


Figure 3.6: U.S. Identity Theft Over Time

was introduced by Abadie and Gardeazabal (Abadie and Gardeazabal, 2003) in 2003 where the authors studied the effect of terrorist conflict on the Basque region's GDP. It was further explored by Abadie, Diamond, and Hainmueller in 2010 where the authors examined the effect of California's cigarette tax on cigarette consumption (Abadie et al., 2010). The method is useful for comparative analyses, particularly when evaluating policies at the level of a state or country where there are relatively few units in the dataset (Athey and Imbens, 2017). The basic logic underlying synthetic control is that a real-world "treated" unit is compared to a synthetic control of itself. The synthetic control is created by borrowing covariates from other units. The key to creating a successful synthetic control is making sure that the synthetic unit matches the real world observed unit in the pre-treatment period. Once a successful synthetic unit is created, it is straightforward to calculate the treatment effect by subtracting the post-treatment control outcome from the post-treatment treatment outcome. It is essentially an extension of the difference-in-difference method that allows for the comparison of the treated unit against a hypothetical controlled version of itself, rather than an observed control unit.

[Note: Might scrap the Hainmueller notation and make this match what comes below] Formally, the authors motivate the model as follows. Imagine that there $J + 1$ regions of interest. Allow Y_{it}^{N} to be the outcome for region i at time t for units $i = [1 : J + 1]$ and time periods $t = [1 : T]$. Let T_0 be the number of preintervention periods, with $1 \leq T_0 < T$. Let $Y_i^I t$ be the outcome if unit i at time t was exposed to the intervention.

In this case, we wish to estimate the effect of the updated data breach notification law on California's medical identity theft rates. This relationship can be expressed as:

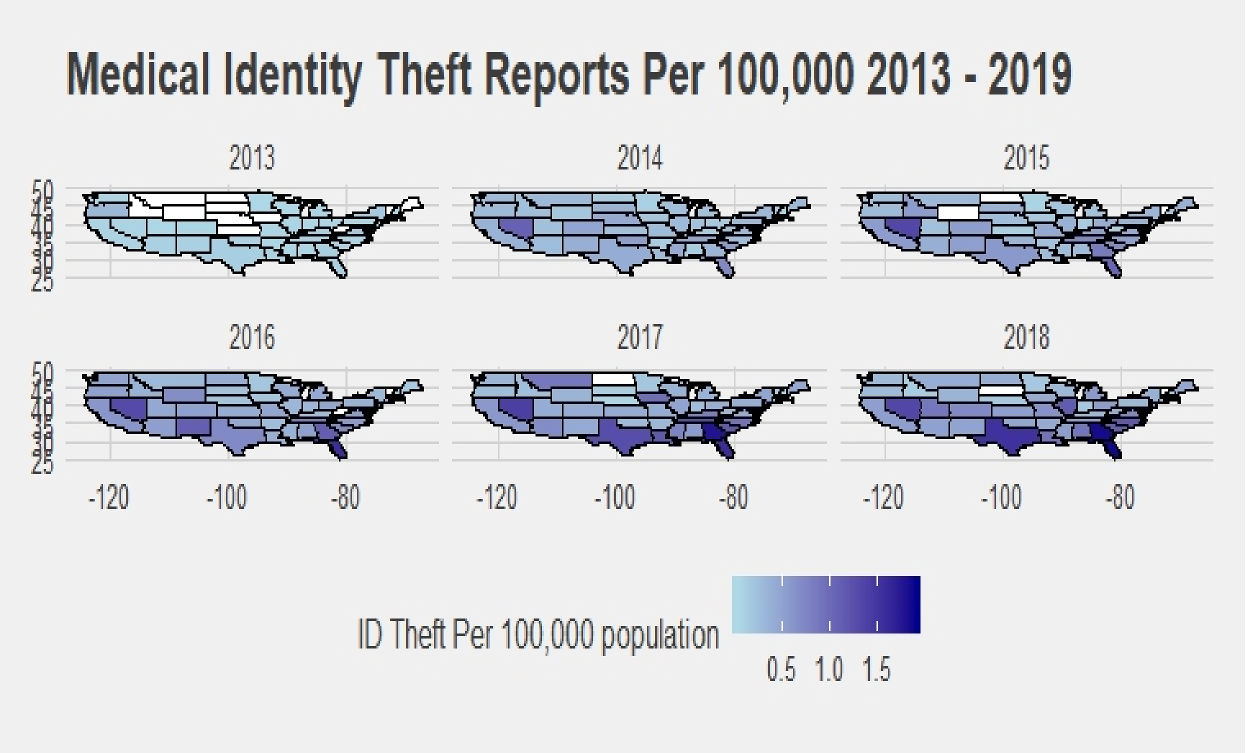

Medical Identity Theft Reports Per 100,000 2013 - 2019
2013
2014
2015
2016
2017
2018
50
45
40
35
30
25
-120
-100
-80
ID Theft Per 100,000 population
0.5
1.0
1.5

Figure 3.7: Medical Identity Theft by State Per 100,000 Population

$$ATT = Y^I - Y^N$$

Where ATT_t is the treatment effect[2], Y^I is California's medical identity theft rates with the law, and Y^N is California's medical idenitity theft rates without the law. This setup is analogous to the Neyman-Rubin potential outcomes framework. The fundamental problem of causal inference is that we can never observe the same unit under both treatment and control conditions (Rubin, 1974). Addressing this problem requires estimating the potential outcome for a unit under the counterfactual condition (i.e. estimating the value under treatment for a control unit, or estimating the value under control for a treated unit). The synthetic control method handles the fundamental problem of causal inference by creating an estimate of a counterfactual treated unit through a weighted combination of other untreated units.

One issue with the synthetic control method is that it is only valid when the synthetic unit matches the observed unit in the pre-treatment period. Ben-Michael, Feller, and Roth-

[2] ATT = Average Treatment Effect on the Treated. ATT is used in applications like difference-in-differences and synthetic control because the effect is being estimated for units that received treatment. This concept can be distinguished from Average Treatment Effect (ATE) which is an estimate of a treatment effect in a randomized control trial.

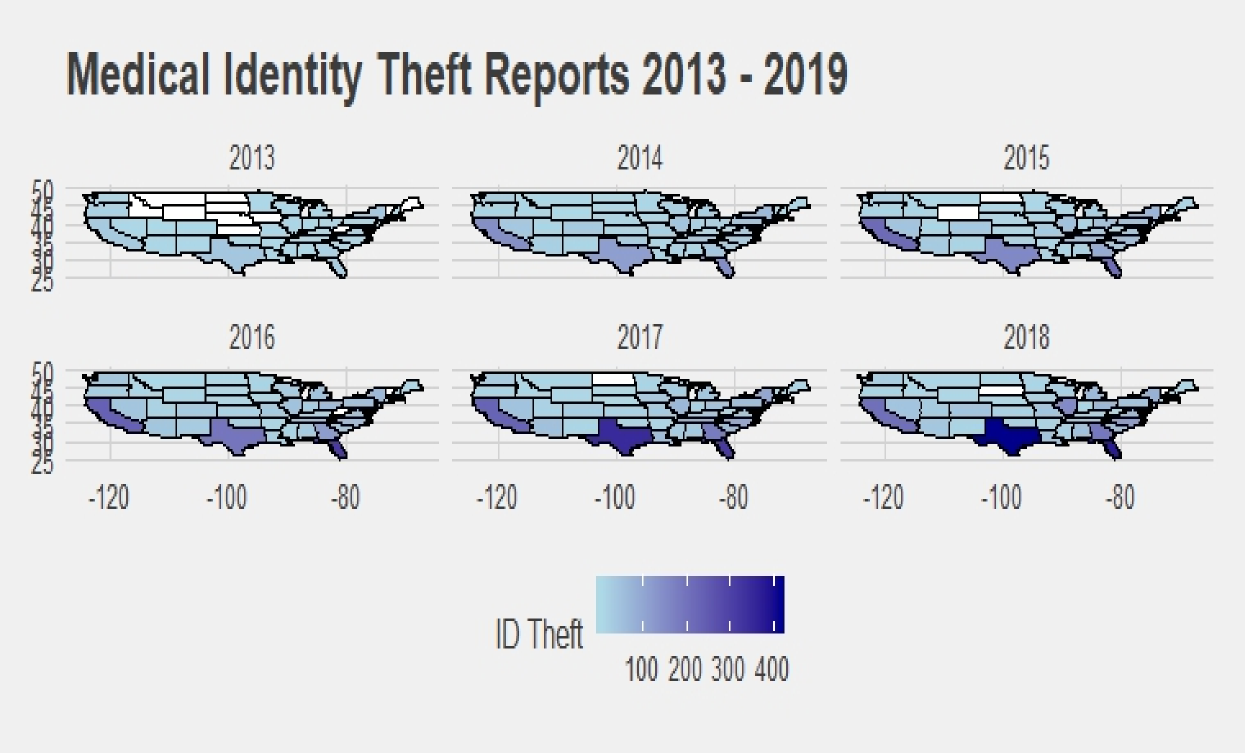


Figure 3.8: Medical Identity Theft by State

stein proposed an augmented synthetic control method that offers bias correction tools in situations where such pre-treatment matching is infeasible (Ben-Michael et al., 0). They propose using an outcome model to estimate bias in the pre-treatment fit of the synthetic control, and then debias the original synthetic control estimate. The authors specifically recommend using a ridge regression, and also provide random forest and matrix completion methods (Athey et al., 2017), among others [3]. Augmented synthetic control essentially has all of the transparency advantages of a standard synthetic control, but provides additional options in situations where perfect pre-treatment fits are not possible.

Extending the basic framework to this problem, imagine there are a set of states $i \epsilon S = 1 : 50$, and a set of time periods, $t \epsilon T$. The general problem of estimating the treatment effect for a unit, i at time t can be described as:

[3] Athey et. al. argue that ”The [Neyman-Rubin] unconfoundedness approach estimates patterns over time that are assumed to be stable across units, and the synthetic control approach estimates patterns across units that are assumed to be stable over time.” These different assumptions impose different restrictions on the missingness of the outcome. The authors suggest matrix completion methods that use regularization to estimate missing data, and relax the assumptions of either the uncoufnoundedness or synthetic control approaches that are popular in econometrics.

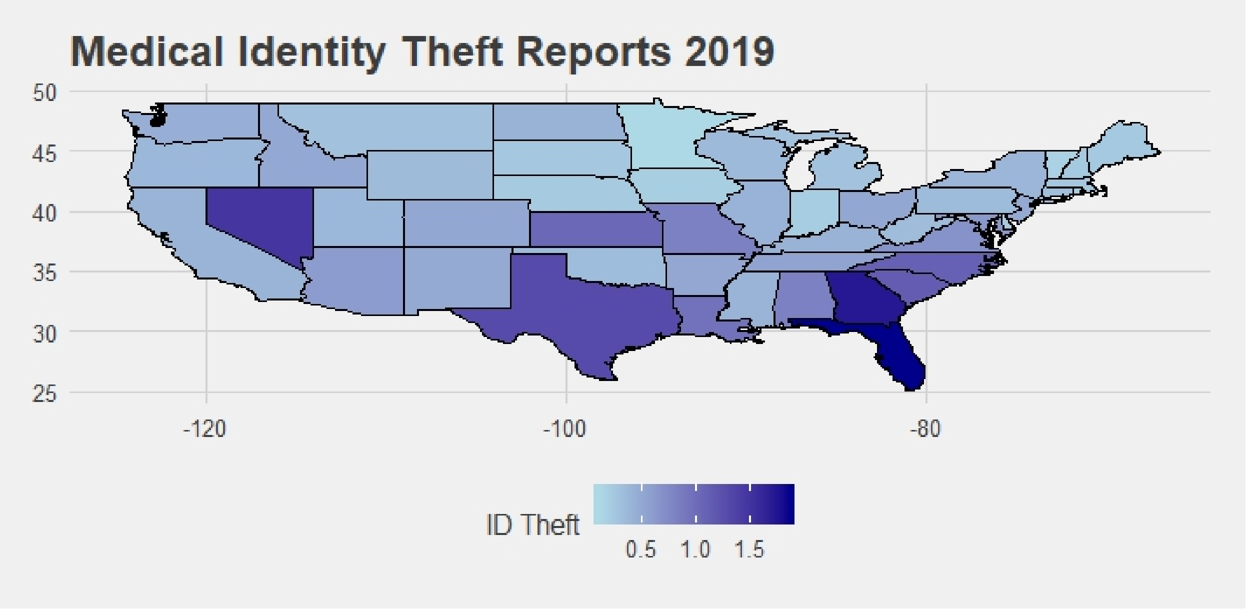


Figure 3.9: 2019 Medical Identity Theft Reports per 100,000 Population

$$ATT = Y_{it}^{I} - Y_{it}^{N}$$

In this specific instance, assume that California is unit 1. Thus for each time period, the estimate is:

$$ATT = Y_{1t}^{I} - Y_{1t}^{N}$$

Where:

- ATT = Average Treatment Effect on the Treated. Reduction in medical identity theft rates per 100,000 people.
- Y_{1t}^{I} = Observed medical identity theft rate in California
- Y_{1t}^{N} = Synthetic estimate of medical identity theft rate in California

Where Y_{1t}^{N} is estimated by constructing a synthetic control. For an overall treatment effect, I average the ATTs across each post-treatment time period. I also fit various models, both with and without augmentations.

3.7 Results

Difference-in-Differences Baseline

I start with a demonstration of a simple difference-in-differences estimate that compares California to the U.S. average (minus California) for medical identity theft rates. Figure 3.10 shows the synthetic California estimate, the observed California, and the U.S. average (by state) for medical identity theft reports. The U.S. average does approximate California fairly well in the pre-treatment period, it tends to underestimate rates in California, and therefore is not an ideal comparison unit in the diff-in-diff framework. The synthetic control, which is a weighted sum of U.S. states, lessens this problem somewhat and tends to match California in the pre-treatment period, particularly in the periods immediately preceding treatment.

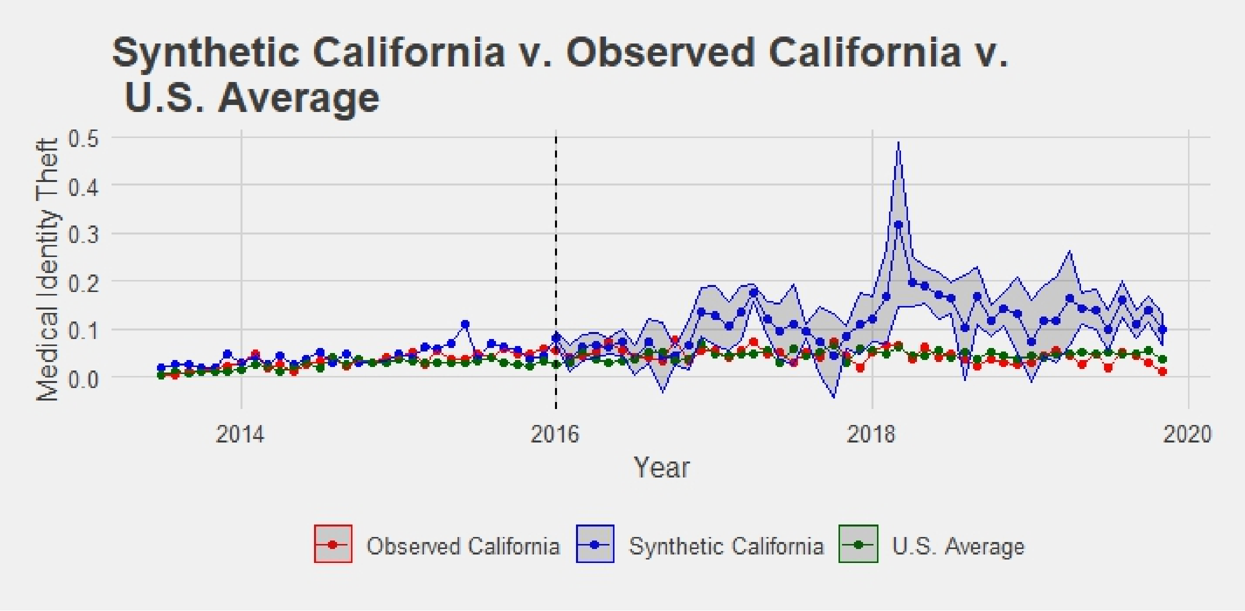


Figure 3.10: Synthetic California, Observed California, and the U.S. Average

Figure 3.11 illustrates the U.S. average differing from California rates in the pre-treatment period more clearly. Diff-in-diff relies on the "parallel trends" assumption that requires that the difference between the treated and control units is constant over time. The U.S. average tends to underestimate California, but sometimes exceeds it, and varies from time period to time period.

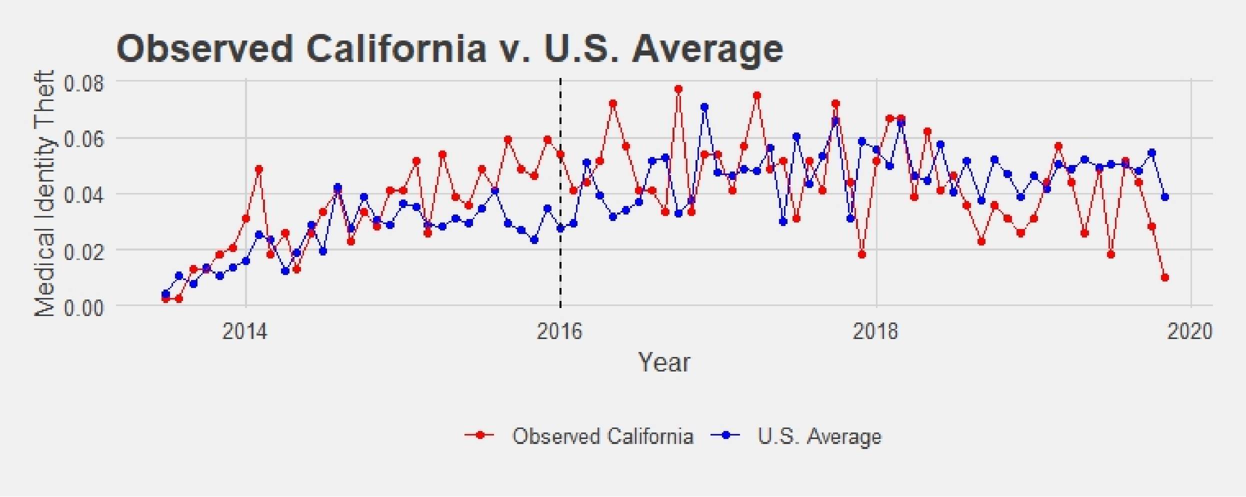


Figure 3.11: Difference in Differences Observed California v. U.S. Average

No Augmentation

Next, I show synthetic control results without augmentation. Augmenting a synthetic control is mainly useful in situations with poor pre-treatment fit. In this case, the fit is still fairly good without augmentation. Figure 3.12 shows the standard synthetic control estimate with no augmentations or additional covariates.

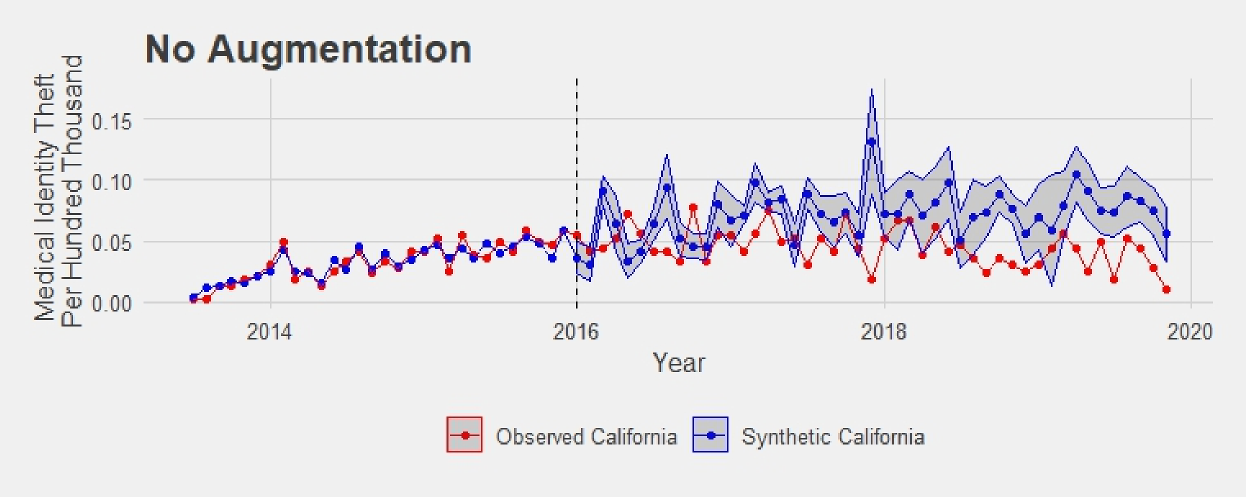


Figure 3.12: Standard Synthetic Control

Ridge Augmentation

Moving to the augmented estimates, I present results from the Ridge-augmented synthetic control. Preliminary results suggest that there is a small effect of these expanded notice standards on reported medical identity theft. I employ a Ridge-augmented synthetic control on reported medical identity theft. I provide estimates both per year and per month, and estimate total number of reports and reports per 100,000 people. The synthetic control also adds number of HIPAA reported breaches and the number of individuals affected by medical breaches in each state in a given year as additional covariates.

Figure 3.14 shows synthetic control estimates broken down by month. Table 3.1 shows the relative contributions of each state to the synthetic control, and Figure 3.13 visualizes these weights. The advantage of using synthetic control over the U.S. average is that the synthetic control attaches weights to each state so construct a more appropriate control unit. Scaled to victimization rates per 100,000 people, the average treatment effect on the treated is approximately .065 fewer reports of medical identity theft per 100,000 people (see Figure 3.14). The effect size grows over time, with the final estimate being close to .1 fewer reports per 100,000 people. This suggests that expanded disclosure requirements have a modest effect that potentially grows over time, though caution is advised against extrapolating too far into the future with relatively few pre-treatment periods.

	State	Weight
1	AL	-0.03
2	AR	-0.05
3	AZ	-0.00
4	CO	0.00
5	CT	-0.05
6	DC	-0.01
7	DE	-0.08
8	FL	0.20
9	GA	0.08
10	IA	-0.05
11	ID	-0.02
12	IL	0.66
13	IN	-0.02
14	KS	-0.04
15	KY	0.01
16	LA	-0.03
17	MA	0.01
18	MD	0.01
19	ME	-0.04
20	MI	0.01
21	MN	0.00

22	MO	0.00
23	MS	-0.06
24	MT	-0.00
25	NC	0.04
26	ND	-0.05
27	NE	-0.05
28	NH	-0.05
29	NJ	-0.00
30	NM	-0.04
31	NV	-0.06
32	NY	0.21
33	OH	0.06
34	OK	-0.03
35	OR	-0.03
36	PA	0.09
37	RI	-0.04
38	SC	-0.03
39	SD	-0.08
40	TN	0.03
41	TX	0.64
42	UT	-0.04
43	VA	-0.00
44	VT	-0.06
45	WA	-0.01
46	WI	-0.05
47	WV	-0.05
48	WY	0.06

Table 3.1: Weights Generated by Synthetic Control

The L2 imbalance (square root of the sum of the sum of squared vector values) is .11, with an average estimated bias of .05, and an average ATT estimate of -.069. The estimated treatment effect for November 2019 is about .1 fewer reports per 100,000 people. However, because there are so few reports in any given month, the estimates and observed values can be quite noisy. Thus, we should be careful about interpreting a large treatment effect from these estimates. That being said, pre-treatment fit is generally good regardless of the chosen model. Figure 3.15 shows synthetic control estimates across no augmentation, ridge, matrix completion, and gsynth (linear factor model). In each case, the overall pre-treatment fit is similar, as are the estimates.

The expanded breach notification requirements does imply a modest effect on medical identity theft reports. Both the synthetic estimates and the observed values of medical

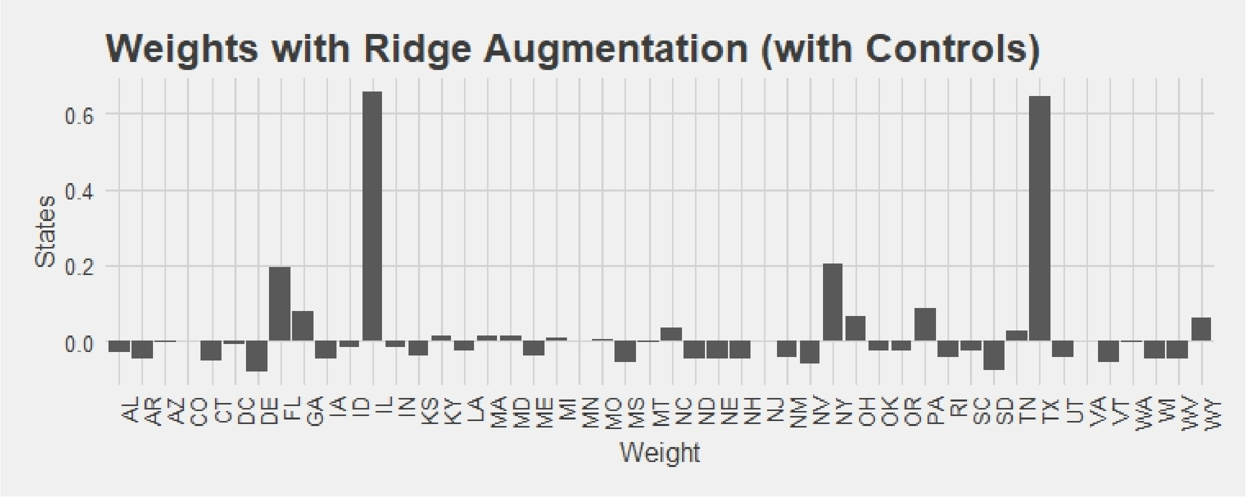


Figure 3.13: Barplot of Ridge Augmented Weights

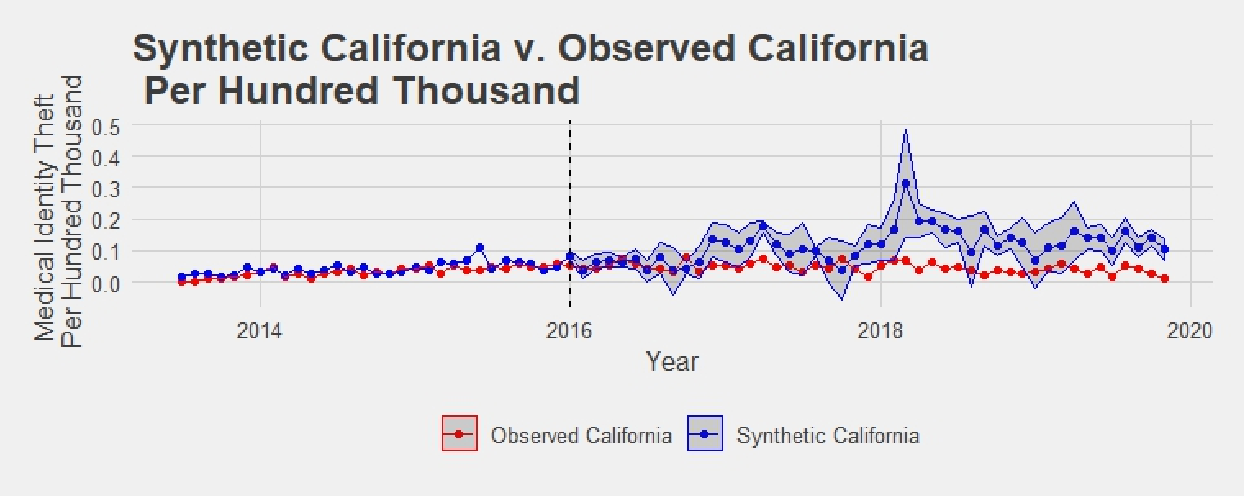


Figure 3.14: Estimates by Month

identity theft reports represent lower bounds. The CFPB medical identity theft reports are (likely unrepresentative) a sample of all reported identity theft reports, and not all identity theft is reported at all. More data on identity theft reports would be helpful for reducing the noise in the estimates, particularly at the monthly level.

One main takeaway from these results is that the effect of the law grows over time. Figure 3.16 shows the estimated treatment effect on medical identity theft reports per 100,000 people per month. In the months immediately following the law going into effect, there is little change from a null effect, and matches the pre-treatment outcomes. Between January

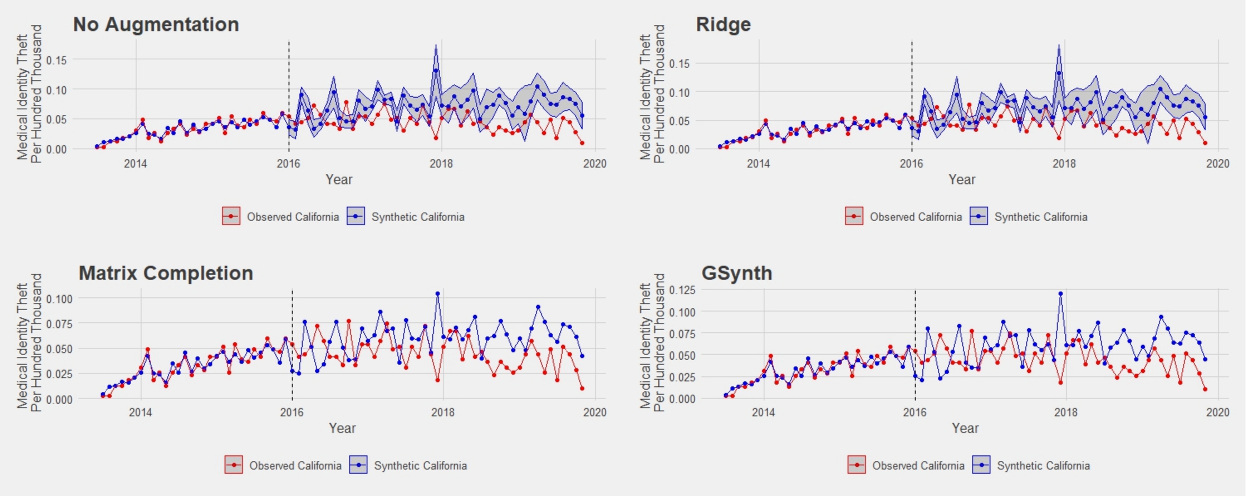


Figure 3.15: Augmented Synthetic Control Across Outcome Models

2016 and November 2019 however, there is downward movement. Again, this pattern holds regardless of the specifications chosen, and suggests that the law has a modest but real effect on reported medical identity theft. Table 3.2 shows the results from various specifications, both with controls for state medical infrastructure and without.

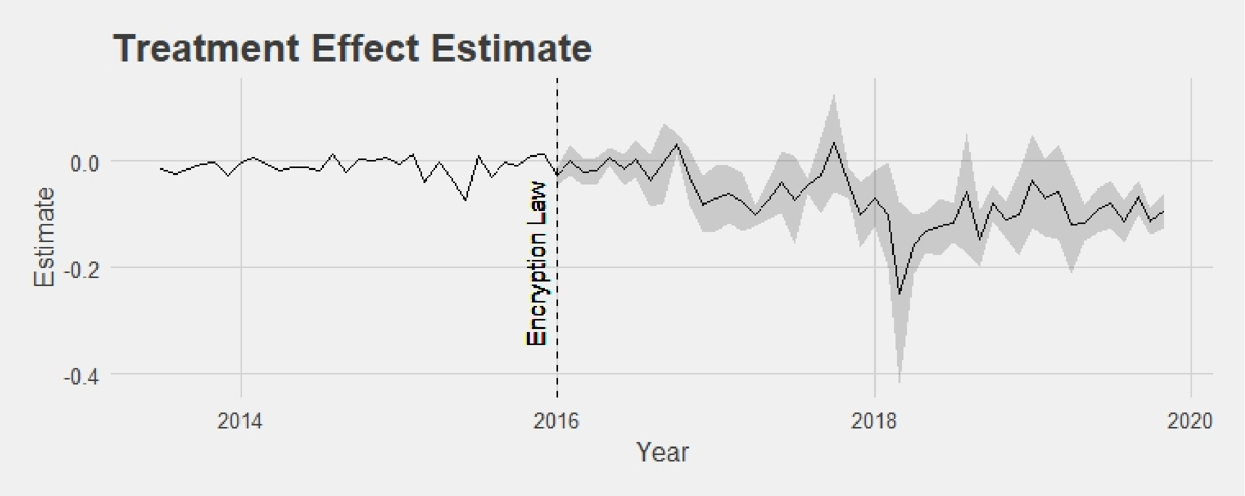


Figure 3.16: Average Treatment Effect on The Treated Over Time

Otherwise, a question worth probing further based on these results is whether previous estimates of the effects of data breach notification laws potentially understated the magnitude of these effects. Previous work that estimated these effects using a panel regression found

	Outcome Model	L2 Imbalance	Average ATT
1	None	0.03	-0.03
2	None With Controls	0.03	-0.02
3	Ridge	0.03	-0.03
4	Ridge With Controls	0.11	-0.07
5	Matrix Completion	0.03	-0.01
6	GSynth	0.03	-0.02

Table 3.2: Outcome Models with L2 Imbalances and Average ATT

an effect of about 2% on reported identity theft (Romanosky et al., 2011). This work evaluated laws from the mid-2000s, and looked at identity theft rates across all kinds of identity theft. These estimates potentially understate the effect of state laws on state identity theft rates because of leakage of notifications across state lines for commercial breaches, underinclusiveness in the definition of "personal information" in the mid-2000s wave of laws, and omitted variable bias in the regression specification. In contrast, the synthetic controls imply an average treatment effect on the treated that corresponds to between a 2/100,000 a 7/100,000 decline in reported identity theft in California. In the final periods, the estimated percentage effect is closer to 40% [4]. Again, there are many potential sources of bias in the synthetic estimates as well, and smaller and null treatment effects are within one standard error of the point estimates in the periods immediately following adoption of treatment. That being said, the synthetic control estimate suggests that data breach notification may be more effective than previously thought.

3.8 Policy Discussion & Future Work

Policymakers are increasingly paying attention to privacy and cybercrime issues, and the data breach notification law remains the most popular and widespread tool used by U.S. states, federal agencies, and the European Union (EU). Despite its prevalence, there is little evidence about its efficacy, especially in recent years. As states rapidly and frequently adopt and update their data breach notification laws, understanding their effects will be of paramount importance.

One ongoing debate in privacy law is whether disclosure is an effective regulatory mechanism. Beyond data breach notification, governments are actively creating various notification requirements pertaining to individuals' privacy. The EU's GDPR contains several provisions

[4] For the most part, I avoid expressing the effect in these percentage terms because of the low baseline rates of reported identity theft. Even a reduction of reports on the order of dozens or hundreds can have a sizable percentage effect, even in a large state like California. To effectively compare to previous literature I provide the percentage effect here, but with the understanding that it is highly susceptible to swing because of low base rates.

that require companies to make their privacy agreements visually appealing and intuitive. The California Consumer Protection Act (CCPA) requires that companies that collect consumer information disclose what information is being collected about them, and whether it is sold. The theory underlying all of these regulations is that disclosures will help consumers control their information, and make rational decisions about their market participation. Both the GDPR and CCPA are recent developments, but indicate that policymakers continue to look to disclosure as the best regulatory option in this space.

This study provides an empirical examination of whether disclosure works. In the 15 years since California implemented the first data breach notification law in the U.S., every state has adopted some version of one. Previous attempts to study these laws were hampered by lack of access to data about the number of breaches and identity theft reports. By using CFPB data, this study overcomes some of those previous challenges. That being said, more comprehensive and publicly available data on identity theft reports would enhance researchers' ability to empirically answer important questions about privacy law and policy.

With regards to implications for privacy law, these results tentatively suggest that California's data breach notification updates in 2016 had an impact on the reported medical identity theft in the years after its adoption. There are a few important pieces to note here before generalizing to all data breach notification laws in all states across time. First, California already had a fairly strong data breach notification law in place, with a fairly expansive definition of "personal information," requirements that breached organizations notify consumers and the Attorney General, and penalties for failure to comply. The 2016 amendments required that notices use a particular format, provide clear information, and be labeled clearly. Thus, the 2016 amendments were more focused on the style and substance of the disclosure, rather than changing the types of disclosures that needed to be made. These results therefore point to the effect of mandating that disclosures look a particular way, not the effect of a generic data breach notification law. Moreover, the results may not generalized well beyond California; the exploratory data analysis showed that there are clear state and regional patterns in medical breaches and identity theft, so another state that adopts California's requirements may not enjoy the same benefits. Medical identity theft may also be different from other kinds of identity theft, and the law could be especially good or bad at deterring that category of cybercrime and not others. Keeping this caveats in mind though, the results suggest that disclosure does matter, and more importantly, that clear, well-designed disclosures matter.

Data breach notification laws continue to evolve, and these changes should provide researchers with ample opportunities to study the effects of various aspects of disclosure regulation. For example, one sources of variation between state laws is the presence or absence of private causes of action following a disclosure. Some states allow individuals to sue organizations following a breach notification, while others only allow the Attorney General to make that determination. Various states have different rules regarding who must be notified (consumers, attorneys general, and/or credit reporting agencies). Exceptions to breach notification requirements when data is encrypted are being reexamined. Differences in requirements for "likelihood of harm" analysis may also produce divergent outcomes. While breach notification laws share many similarities, these differences could provide a rich set of questions for more empirical work.

3.9 Conclusion

Data breach notification will likely continue to be a popular tool for policymakers regulating cybercrime, thus making evidence of how well the current regime works important for future policy decisions. Legislators and regulators are actively debating whether disclosure is an effective mechanism for protecting consumers without implementing heavy-handed market interventions. Quantifying the harms stemming from privacy invasions is a notoriously difficult problem, making it difficult for policymakers to know which policy levers to pull. Estimating the potential effect of disclosure on identity theft is a first step in understanding whether data breach notification laws are the most effective tools for protecting privacy. Using an augmented synthetic control approach, I estimate the effect of California's 2016 amendments to its breach notification law. These results tentatively suggest that breach notification does reduce identity theft, but more work is needed to get a complete picture.

Chapter 4

Deterring Cybercrime: Focus on Intermediaries

[1]

4.1 Introduction

Businesses that sell illegal pharmaceuticals, counterfeit goods, or offer computer attacks online have similar goals and needs as ordinary firms. These enterprises must acquire new customers, have a supply chain, maintain a web presence, collect payments, deliver a product or service, and, finally, cultivate a positive reputation to encourage repeat sales. In pursuit of profit, the legitimate and illegitimate alike depend on many third parties, including web hosts, payment providers, and shipping companies.

Licit businesses are deterred from illegal acts by punishment, through fines, threats, and regulatory actions. But enforcers often cannot use traditional deterrence against financially–motivated cybercriminals because law enforcement is limited by scarce resources, competing enforcement priorities, and jurisdictional challenges. As a result, enforcers—both public and private—have turned to deterrence by denial approaches. Such approaches attempt to deter conduct by spoiling, reducing, or eliminating the benefits of computer crime. Frustrated by attempts to reach actual illicit actors, enforcers focus on third parties that are critical to business operation, thus denying cybercriminals' access to banking, web resources, or even shipping services. Cybercrime, often presented as ephemeral and stateless, can be reined in through attacking dependencies critical to its operation. Much of the legal academic scholarship on Internet intermediaries focuses on intermediaries' general immunity from state law actions under the Communications and Decency Act Section 230 (CDA 230) or the provisions of the Digital Millennium Copyright Act (DMCA). CDA 230 creates broad immunity for Internet intermediaries, insulating them from the illegal acts of their users; intermediaries, even when given notice of noxious content, are not required to remove it

[1] Coauthored with Chris Hoofnagle and Damon McCoy

(USC, 2012c). The DMCA can shield providers from liability for user's infringing activities if certain steps are taken to receive and respond to takedown requests by intellectual property (IP) owners (USC, 2012b).

This Article turns away from the CDA 230 and the DMCA procedures to focus on mechanisms that force intermediaries to address alleged user misbehavior. Specifically, this Article focuses on three mechanisms that are used to cause intermediaries to take or refrain from some action related to financially–motivated cybercrime. Parts II and III set the stage for the survey. Part II canvasses the literature on cybercrime and intermediaries. Part III discusses the business constraints of three kinds of actors: botnet operations, sites that offer illegal and infringing goods, and the contests among intermediaries and intellectual property owners. Part IV covers the use of Rule 65 of the Federal Rules of Civil Procedure (FRCP) and its allowance for broad forms of injunctive relief, and the Domain Name Service takedown procedures that use the U.S. government's ability to target infringing websites and make them inaccessible. Part V covers administrative remedies focused on financial services intermediaries. Finally, part VI looks at self–regulatory procedures that intermediaries have established to allow IP owners and governments to block user activity.

An intermediary–focused approach raises due process and fairness concerns because intermediaries may not be privy to criminal activity, and enforcement mechanisms affect consumers and other licit firms. Cybercriminals may mask their behavior by commandeering ordinary users' accounts and computers for attacks and monetization of crimes. Thus, when an enforcer investigates and makes interventions, legal demands may fall upon third parties, individuals, and businesses that were merely used as conduits by the suspect. These intermediaries themselves may have been hacked or otherwise be cybercrime victims themselves. Additionally, compliance may impose costs on intermediaries and to civil society in the form of censorship or erosion of Internet anonymity as intermediaries are asked to know their customers [2] and make requirements that ordinary users provide documentation of their identity. Interventions are done ex parte, with surprising speed, raising the risk that others' interests may not be fully considered by a court. Finally, there is always the problem of claimant abuse—claims of wrongdoing may be motivated by anticompetitive interests or simple censorship.

In sum, this Article offers an exploratory look at an understudied area of intermediary interventions. Intermediary liability conversations typically surround CDA 230 and the DMCA, but our survey reveals that intermediaries can be subject to costly, broad interventions in cybercrime contexts. This Article highlights the current legal practices in this space, and evaluates their merits and demerits.

[2] Anti–money laundering customer identification requirements are known as "Know Your Customer" regulations. See USC (2012a); CFR (2016)

4.2 Literature Review

his Part highlights relevant literature from both a theoretical and legal perspective, starting with a brief overview of the literature on the economics of financially–motivated cybercrime. These pieces link the economics of cybercrime to the economics of crime more broadly, and identify the features of cybercrime that make it amenable to intermediary interventions. The focus then shifts to the legal theory concerning the extent to which intermediaries should be held liable, before turning to the implementation of legal rules and interventions. Some cybersecurity literature focuses on intermediaries' centrality in Internet activity. Authors detail the both private and government policies that aim to thwart cybercrime and create secure systems. Goldman and McCoy set up the motivation for our inquiry in their paper, Deterring Financially Motivated Cybercrime (Goldman and McCoy, 2016). For instance, they address how some cybercriminals are dependent upon a handful of payment processors, which empowers those payment processors to effectively combat criminals (Goldman and McCoy, 2016). They argue that mainstream payment processors adopt policies that help thwart and deter cybercrime, in large part because payment companies want to maintain the integrity and the reputation of their own systems (Goldman and McCoy, 2016). This is a boon for the government and potential victims of cybercrime because payment processors can interrupt a large portion of cybercrime without imposing direct costs on consumers.

Cybercrime is an increasingly professional endeavor that implicates activities involving American companies, or are otherwise subject to U.S. courts' jurisdiction. In The Economics of Online Crime, Moore et al. elaborate on how cybercrime professionalization lends itself to well-understood policy fixes (Moore et al., 2009). They explain that criminal firms have emerged that specialize in botnet creation, phishing, and identity theft Moore et al. (2009). They argue, "[w]ith this new online crime ecosystem has come a new profession: the 'botnet herder'—a person who manages a large collection of compromised personal computers (a 'botnet') and rents them out to the spammers, phisher[s], and other crooks."9 Because cybercrime has become increasingly professionalized, it has started to look more like conventional crime that has been explored at length in the economics of crime literature (Moore et al., 2009).

Beyond payment processors and criminals themselves, another area of this literature focuses on internet infrastructure and its relationship to cybercrime. In The Turn to Infrastructure in Internet Governance, the authors look at the fundamental building blocks of the Internet as sources for governance and, consequently, security (Musiani et al., 2016). For instance, authors discuss the role of the Domain Name System (DNS) and the Internet Corporation for Names and Numbers (ICANN) as the backbones of the Internet (Musiani et al., 2016). This is important for cybersecurity because of the U.S. government's ability to directly seize domain names and take control of infringing websites, as discussed, in more detail. Essentially, the authors point out that even though the Internet was designed to usher in diffused and ground–up governance, in actuality, governance structures can reduce access to certain resources needed by cybercriminals (Goldsmith and Wu, 2006).

Similarly, in Holding Internet Service Providers Accountable, Douglas Lichtman and Eric

Posner argue that Internet Service Providers (ISPs) can be essential nodes in cybercrime networks, and should be held to higher legal standards (Lichtman and Posner, 2006). They argue that the move toward granting immunity to ISPs is ill–advised because it underestimates ISPs' ability to deter cybercrime, and gives them license to allow dangerous behavior (Lichtman and Posner, 2006). This argument is in line with standard law and economics theory, predicting that always assigning liability to one party (in this case, victims) will cause the other party (in this case, ISPs) to take inefficient levels of precaution (Cooter and Ulen, 2016). Lichtman and Posner map the theory of indirect liability onto the actions taken by ISPs and conclude that ISPs should share some responsibility for cybercrime (Lichtman and Posner, 2006).

Moving from theory to policy, Operation Seizing Our Sites raises criticisms of an overbroad intermediary–centered approach (Kopel, 2013). In the article, Karen Kopel discusses federal programs that aim to take down copyright and trademark infringing websites Kopel (2013). In particular, she critiques "Operation In Our Sites," a major government initiative for enforcing stricter IP protections (Kopel, 2013). The program's main mechanism allows the Department of Justice and Immigrations and Customs Enforcement (ICE) to seize domain names by ordering intermediaries to reassign them, and make these resources inaccessible to users who try to access a website through its alphanumeric name (Kopel, 2013). She argues that the process largely circumvents normal procedural safeguards, and grants the government wide discretion in pursuing potential infringers (Kopel, 2013). She also notes the risks associated with the approach, namely that the government has taken legitimate websites offline and offered them few due process protections to appeal the decision (Kopel, 2013). In practice, hardly any websites are able to recover their domains after a government seizure (Kopel, 2013).

The next Part turns to the business constraints that financially–motivated cybercriminals face, and examines those actors' various activities, including their dependence on intermediaries. It then summarizes the legal processes used in situations where enforcers—both public and private—attempt to deter financially–motivated cybercrime by interfering with intermediaries.

4.3 Business Constraints, Relevant Actors, and Activities

Financially–motivated cybercriminals face many of the same business constraints and challenges that legitimate enterprises do. A paradigmatic example comes from an illegal goods business called Silk Road, which provided a marketplace for drugs, fake identification documents, and materials for credit card fraud (Christin, 2013). Another comes from the infringing goods space, where sellers, often using quickly seized ephemeral domains, market knock–off designer bags and other cheap–to–produce but high–priced items. In the illegal or infringing goods businesses, a successful enterprise needs a prominent web presence, similar to the mainstream brands. Businesses gain such prominence by having easy–to–recognize domain names and search–optimized sites. The website has to be reasonably well–designed and available to users. One also needs to be able to collect payments from users and to deliver the product to the consumer. Even illicit businesses, such as counterfeit pharmacies, care about reputation because they earn up to thirty–seven percent of their gross revenue from repeat purchases (McCoy et al., 2012). Thus, customer reviews are important (Krebs, 2014). Turning to botnets, operators face business–like costs too. Cybercriminals have specialization and expertise, as do many other actors in the broader economy, creating a complex market for services (Cardenas et al., 2010). Cybercriminals in these markets must advertise their services, deliver them reliably, collect payment, and (in the case of botnets) maintain a collection of compromised computers. In this last function—botnet maintenance—bot herders, who conscript vulnerable machines into botnets, are in constant conflict with both nation states and sophisticated technology companies. To turn a profit, like ordinary businesses, illegal and infringing good sellers must make many sales.

In both the illegal goods and infringing goods contexts, each critical function to monetizing the crime relies on third party intermediaries. Sellers and marketplaces need domain names, hosting services, access to payment systems, banking services, access to postal or shipping networks, and so on. Many of these intermediaries are probably unaware of misconduct (Alrwais et al., 2017). For various practical reasons, the nature of the web causes businesses to concentrate their services, making entire enterprises dependent on single intermediaries in some contexts. For instance, Levchenko and collaborators found that just three banks processed transactions for ninety–five percent of the goods advertised by spam in their study (Levchenko et al., 0). In another study, author Hoofnagle showed that among the most prominent online pharmacies, many shared the same shopping cart and same telephone services for sales (Hoofnagle et al., 2017). It also appears that the rewards from such activities inure to a small number of actors. For instance, McCoy and colleagues performed an in-depth study of the customers and affiliates associated with three online pharmacy networks (McCoy et al., 2012). The group observed that affiliate marketers are major purveyors of web spam to promote online pharmacies and that a small number of advertisers in the affiliate network captured the most revenue.33 In particular, the largest earner of commissions was a company that specialized in web spam, and it made $4.6 million.34 McCoy and colleagues also found that twenty to forty percent of sales from email spam arise from users who actively open their spam folder and click on links to pharmacy sites (Chachra et al., 2014).

At the root of this discussion are actors who are perpetrating a wide variety of cybercrimes. These crimes are as diverse as illegally distributing copyrighted content, hacking, and engaging in the trade of illicit goods and services (i.e. drugs, sex trade, human trafficking, etc.). These criminals are exceptionally difficult to pin down because they operate with complex social networks that often span international borders.

The next Sections turn to some of the key actors that depend on or attempt to interfere with intermediaries: botnets, sellers of counterfeit goods, and intellectual property owners.

Botnets

Botnets are networks of infected computers (the “bots”) that are used to conduct illegal operations. In particular, botnets can be used to forward communications (i.e. spam emails, viruses, etc.) to other computers and grow the network, and to execute Distributed Denial of Service (DDoS) attacks that can disrupt all internet use. DDoS attacks were a principal tactic in the first nation–state cyberattacks, thus making botnet mitigation a concern for both businesses and nations.36

As targets of legal interventions, botnets are tricky to pin down because of their international and self–propagating nature. Individual bots could be in the homes of consumers all over the world, and could be in the form of the computers and software embedded in internet–connected cameras and even routers.37 Bots take their direction from remote command and control servers that are generally operated by bot herders. Skilled botnet herders mask these systems, and even distribute control to new domains on predefined schedules. Presumably, the botnet herder knows what new domains will be selected and can compromise them in time to issue new instructions to the bots.38

The handoff of the command and control infrastructure offers an opportunity to disrupt botnets—in effect by rustling them from the herder. Technically and legally sophisticated actors such as Microsoft Corporation can use legal processes to seize the domains that the botnet will next connect with. Once seized, a company can issue instructions to the bots to update their software and stop new attacks. For example, a “sinkhole” tactic routes the associated domain names to a new DNS server, which then assigns non–routable addresses to the domains. Basically, this prevents anyone from actually accessing the website where the individual bots receive their instructions, rendering the botnet impotent (Cooke et al., 2005).

Illegal and Infringing Goods Sellers

People intent on selling illegal and infringing goods use their own websites and online marketplaces to do business and point buyers to other internet resources where infringing content may reside. Similarly, they may use social media and other pages to boost infringing services’ prominence in search engines (McCoy et al., 2012).

These actors pose a challenge for law enforcement because it is difficult to discern legal operations from illegal ones. For example, it is difficult to tell whether a handbag sold on eBay is stolen, counterfeited, or simply from a legitimate owner trying to resell an expensive fashion item. In other cases, sellers set up networks of websites that are obviously in the business of knockoffs. Domain Name Server (DNS) seizure is a common tool leveraged against these easier–to–identify sellers. Because the sellers are typically outside the United States, they are difficult to physically track down, and therefore enforcers have an easier time directly seizing the infringing web domain and other services. The next Section details the prototypical procedure. The basic notion is that the enforcer can take over a domain and prevent anyone from accessing it, therefore shutting down the illegal operations that were being carried out.

Intermediaries and Intellectual Property Owners

This Article focuses on intermediaries' capacity to deter and combat cybercrime, and therefore highlights key players, such as technology companies that provide products and act as online platforms, and IP owners that take actions against infringers. These actors are important because when they invest in cybersecurity, they can produce positive externalities for end users and smaller firms (Mulligan and Schneider, 2011). That being said, there are key distinctions between companies involved with combating botnet activity and companies involved with IP enforcement. The former set of actors is intertwined with the governance and security of Internet infrastructure, and therefore regularly cooperates with public and private institutions to maintain a secure Internet (Eichensehr, 2017). The latter set is mainly concerned with preventing the sales of counterfeit, physical goods over online platforms. But in some cases, the two interests mix—as detailed below, botnets are sometimes used to sell counterfeit pharmaceuticals. Although Internet security and IP enforcement are distinct policy areas, they are discussed in tandem because courts employ similar toolkits in approaching both issues.

On the botnet issue, this Article emphasizes Microsoft's role in cybercrime deterrence because of the company's dominance in operating systems and its centrality in cybersecurity. Microsoft's signature product is its Windows operating system, and protecting the integrity of that product is a major goal for the company. Over the course of several years, Microsoft's reputation suffered as various viruses infected machines running Windows, and for some time, almost by definition a botnet was comprised of Windows machines (WIRED, 2002). In 2002, Microsoft announced a major security rethink. It aggressively invested in cybersecurity infrastructure and participated in legal proceedings aimed at taking down the most expansive botnets, thus rehabilitating its product's reputation (WIRED, 2002). Microsoft has gone so far as to use this mechanism—a kind of privately–waged lawfare—against the "Fancy Bear" hacking group suspected to have aided President Trump in his contest against Hillary Clinton (Poulsen, 2017). Microsoft's litigation activity is an illuminating example of private activity that leads to more public cybersecurity, because Microsoft's actions arguably had spillover effects for consumers, businesses running Windows machines, and virtually anyone who was impacted by these botnets.

In terms of IP owners, companies that specialize in goods such as fashion products, rather than music and DVD piracy, are more relevant when discussing intermediary–driven approaches to cybercrime. Generally, fashion products can either be found in brick–and–mortar stores or through online retailers, and are susceptible to being undercut by knockoffs. This is particularly true for fashion products that trade on exclusive, European labels but are actually made in China rather than Italian or French workshops. The exclusive branding of these products drives high prices, but counterfeits often exhibit identical or good enough indicia of quality. Throughout this Article, companies such as Tiffany, Kate Spade, Gucci, and the like provide examples of enforcers that go after infringers. These retailers face challenges in addressing counterfeit sales, which occur in American markets, such as Amazon and eBay, and non–American markets, such as China's Taobao online marketplace. The ease of creating and distributing counterfeit goods in these domestic and foreign marketplaces invites IP infringement.46 Furthermore, jurisdictional issues leave American courts with few options to directly deter this sort of activity.

As such, IP owners have developed a toolkit for dealing with counterfeit goods that simply circumvents the CDA 230 regime and its immunities. Enforcers bring lawsuits using Rule 65 of the FRCP to quickly obtain equitable relief. Enforcers also join professional alliances and organizations, cooperate with payment intermediaries (Bridy, 2015), and work directly with online marketplaces to remove infringing products. Because their products are so easily counterfeited, these companies have a strong incentive to invest in lawsuits as well as technological infrastructure that detects and prevents this activity. In turn, consumers presumably benefit from not being duped through online marketplaces. However, for many consumers, cheap knock–offs or higher quality "factory counterfeits" (those created by employees of the authorized factory during a secret, "fourth shift") might be a perfect substitute for the real thing (Houk, 2016).

4.4 Judicial Interventions

Rule 65 Interventions

Whether enforcers are attempting to police intellectual property rights or fight botnets, they rely on obtaining equitable relief through Rule 65. For all practical purposes, as soon as an enforcer obtains a Temporary Restraining Order (TRO), it has legal authority to order intermediaries to deny services to identified suspects and their internet resources. This deterrence by denial approach is intended to block the defendant from enjoying the gains of their alleged cybercrime.

LEGAL STEP	**LEGAL DESCRIPTION**	**COMMENTS**	**TIMELINE**
Motion for a Temporary Restraining Order	Plaintiff(s) files a motion (often sealed) in District Court requesting a Temporary Restraining Order against one or more Defendants. In the motion, the Plaintiff lists the trademarks that were infringed upon, the websites involved in the alleged activity, and the requested relief.	At this stage, the Plaintiff demonstrates harm, reasons why injunctive relief is necessary, and lists the domain names they would like to seize, along with exhibits with screenshots of offending sites. The TRO is sought without notice to the defendant.	Preparation for this action presumably takes some time because of the need to document infringements and domain owners.

Court issues Temporary Restraining Order	Court issues the TRO. The TRO typically includes a Temporary Injunction, a Temporary Transfer of the Defendant Domain Names, and a Temporary Asset Restraint, among others. At this point, the Plaintiff has achieved the most important legal intervention to deny the benefits of cybercrime to the suspect.	Not only does the Order enjoin the Defendants from further infringement, it also extends to intermediaries that are served with it. For instance, domain name registries are required to either change the infringing pages' registrar or make them inactive and untransferable, and registrars must transfer Defendants' domain names to a registrar account of the Plaintiff's choosing.	Approximately 1 week. Under Rule 65, TROs are to be temporary, thus courts assign short durations to them and prioritize a follow-up hearing for preliminary injunction.
Preliminary Injunction	Court restrains Defendants from operating their allegedly infringing websites	Interventions from the TRO stage are sustained until trial. However, in practice, enforcers typically obtain a default judgment.	4 weeks

Summons served	Clerk of the Court issues summons to Defendants. If Electronic Service is granted, e-mail and posting notice on websites serves as sufficient notice	Defendants are put on notice to respond to claims	Within days of granting of preliminary injunction
Motion to Enter Default Judgment/Final Judgment Order	Finalizes the actions undertaken with the TRO and Preliminary Injunction	At this stage, intermediaries are directed to ensure that the Plaintiff gets permanent control over the infringing domain names	Approximately 2 weeks

Table 4.1: Outline of Legal Steps in Rule 65 Interventions

A TRO is an extraordinary measure because it can be obtained entirely ex parte. The plaintiff bears the burden to show “specific facts in an affidavit or a verified complaint clearly show that immediate and irreparable injury, loss, or damage will result to the movant before the adverse party can be heard in opposition” and the plaintiff must both certify its efforts to give notice to the adverse party and explain why notice should not be required [3]. The purposes of this remedy are to preserve the status quo and prevent irreparable harm until a hearing can take place.

As explained below, once an enforcer obtains a TRO, it has a powerful remedy to use against intermediaries. The court order commands top–level domain name systems to replace the names of the infringing authoritative name servers with a new name controlled by the plaintiff. In some cases, IP enforcers operate the newly acquired domain and notify the public that illegal goods were once sold on it [4]. In botnet cases, enforcers can use the TRO to direct the domain into a sinkhole, which prevents anyone from accessing it [5]. Thus,

[3] Federal Rules of Civil Procedure 65

[4] Luxottica Grp. S.p.A. v. The Partnerships and Unincorporated Associations Identified on Schedule ”A,” No. 1:16-cv-08322, 2016 WL 8577031 (N.D. Ill. Aug. 25, 2016

[5] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF. No. 32.

the intermediary, in cooperation with the Registry, routes the names to the sinkhole, then prevents anyone from accessing the names once they have been successfully placed there.

Rule 65 interventions can occur with incredible speed, relative to ordinary litigation in the notoriously overburdened federal court system. Figure 1 gives an overview of the basic steps and timeline that parties can expect in an expeditious Rule 65 intervention. Plaintiffs enjoys a statutory privilege to get a hearing and relief quickly, often with key documents filed under seal. In the next Section, this Article gives context to these steps in a trademark infringement case.

A TRO is not necessarily a "silver bullet" and it can have some negative repercussions. When cybercriminals are attacked through their intermediaries, the intervention can cause fragmentation—a turn to smaller intermediaries, where substitutes are available. In the case of botnets, operators might move their command and control systems to DNS provided by a bulletproof host, switch to a peer–to–peer architecture, or cloak more of their systems using Tor or I2P. Also, the intervention may be overbroad, negatively affecting innocent third parties (Romanosky and Goldman, 2016).

All judicial interventions are also subject to claimant abuse—situations where one invokes the procedures in order to engage in censorship or anticompetitive behavior. Such abuse comes about both intentionally and unintentionally. Yet, Rule 65 interventions have two checks built into them that are not present in private–sector remedial schemes (discussed in Part VI below). First, Rule 65 requires that movants file a security bond to pay the costs and damages of any party "wrongfully enjoined or restrained." [6] Notice of this bond is provided to intermediaries served with the court's order. Second, the intervention is court supervised. Thus, lawyers, as officers of the court, will presumably avoid abusive applications of TROs and preliminary injunctions lest they attract negative judicial attention.

Examples from Trademark Infringement

In Figure 2, we visualize the typical case flow for a Rule 65 intervention. We chose the Luxottica case as it illustrates several of the most notable features of this intervention, namely the rapid pace from the filing of the lawsuit to the final order, and the massive scope given to the IP enforcer for seizing and controlling assets, coopting intermediaries into compliance, and recovering damages.

Luxottica, a company that owns many sought–after brands of eyewear, provides a paradigmatic example of employing Rule 65 in the IP enforcement context, one that is troubling in scale and presages a kind of automation of litigation. As outlined in Figure 2, in a single 2016 case, Luxottica sued 478 defendants that were allegedly infringing marks in operating 1,024 domains and 52 marketplaces (most of which were "stores" on eBay) [7]. The case caption is so long that it occupies five pages in print, and in the electronic filing system, the defendants are listed as "The Partnerships and Unincorporated Associations Identified on Schedule A."

[6] Fed. R. Civ. P. 65(c)

[7] Amended Complaint, Luxottica Grp. S.p.A. v. The Partnerships and Unincorporated Associations Identified on Schedule "A," No. 1:16-cv-08322 (N.D. Ill. Sept. 1, 2016), 2016 WL 8577031

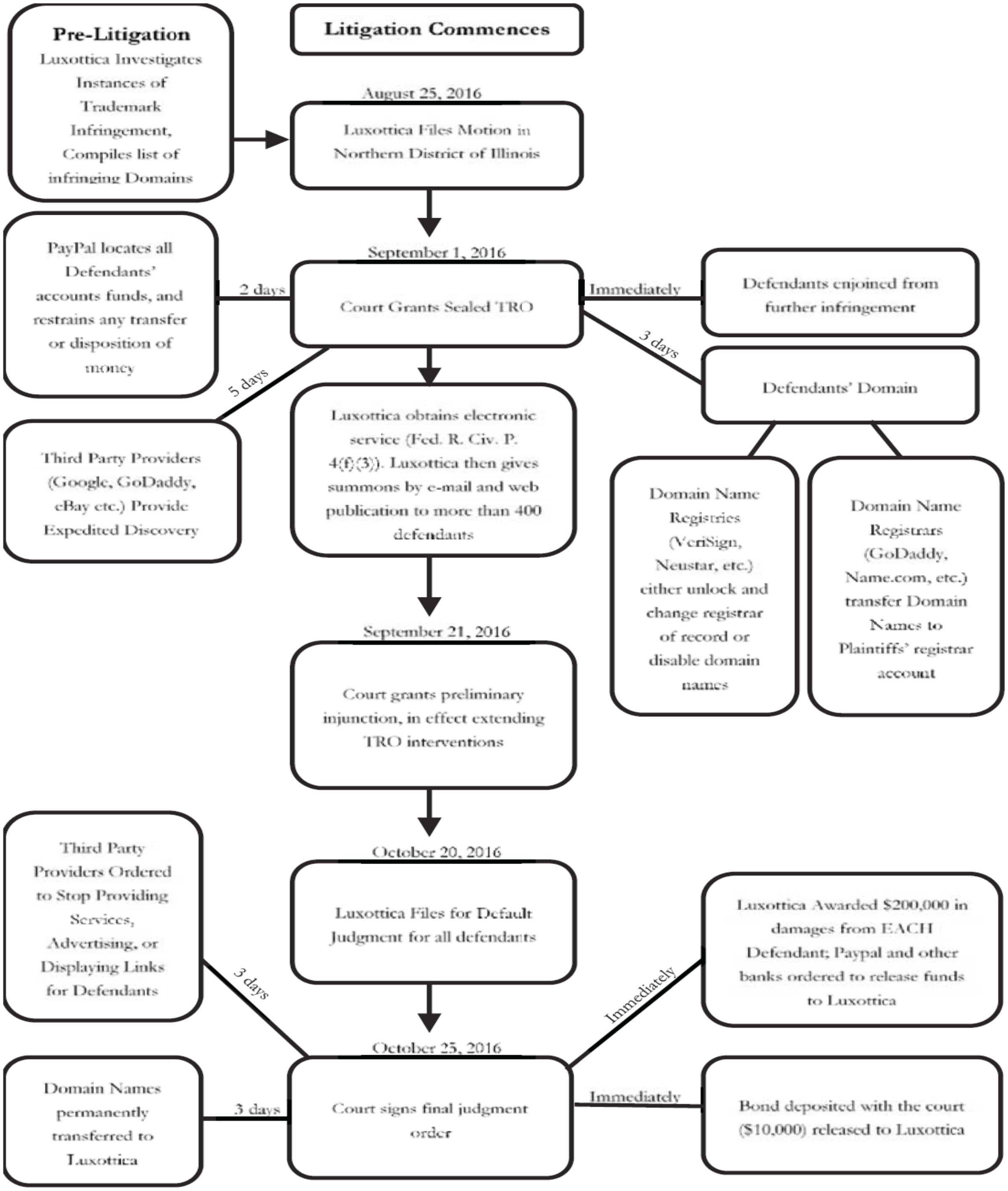

Pre-Litigation
Luxottica Investigates Instances of Trademark Infringement, Compiles list of infringing Domains
Litigation Commences
August 25, 2016
Luxottica Files Motion in Northern District of Illinois
September 1, 2016
Court Grants Sealed TRO
2 days
PayPal locates all Defendants' accounts funds, and restrains any transfer or disposition of money
Immediately
Defendants enjoined from further infringement
3 days
Defendants' Domain
5 days
Third Party Providers (Google, GoDaddy, eBay etc.) Provide Expedited Discovery
Luxottica obtains electronic service (Fed. R. Civ. P. 4(f)(3)). Luxottica then gives summons by e-mail and web publication to more than 400 defendants
Domain Name Registries (VeriSign, Neustar, etc.) either unlock and change registrar of record or disable domain names
Domain Name Registrars (GoDaddy, Name.com, etc.) transfer Domain Names to Plaintiffs' registrar account
September 21, 2016
Court grants preliminary injunction, in effect extending TRO interventions
Third Party Providers Ordered to Stop Providing Services, Advertising, or Displaying Links for Defendants
October 20, 2016
Luxottica Files for Default Judgment for all defendants
Luxottica Awarded $200,000 in damages from EACH Defendant; Paypal and other banks ordered to release funds to Luxottica
3 days
Immediately
October 25, 2016
Domain Names permanently transferred to Luxottica
3 days
Court signs final judgment order
Immediately
Bond deposited with the court ($10,000) released to Luxottica

Figure 4.1: Illustration of TRO Procedure with Luxottica Case

Luxottica filed the case on August 24, 2016, and received a TRO nine days later against all the defendants. [8] Luxottica argued that relief without notice was necessary because the targeted domain owners would likely move their operations if told that an enforcement action was afoot. The lack of notice gave Luxottica another advantage—Rule 65 requires that TROs lacking notice receive a hearing as soon as possible, and so Luxottica received a preliminary injunction less than a month from the date the complaint was filed.

Figure 2 outlines the basic steps taken by Luxottica to obtain the TRO and the many varied entities bound by it. Luxottica filed a required form with the PTO to indicate it was about to enforce its trademark. It obtained a $10,000 bond filed with the court per Rule 65 [9]. That amount was proposed by Luxottica and approved by the court, but presumably could have been raised or lowered to prevent abuses raised by the facts of the case. Luxottica prepared the motions for equitable relief, and in the process, filed straightforward exhibits, thousands of pages long, with screenshots of websites clearly showing Luxottica's product trademarks. But it did not engage in test purchases, which are required to identify the merchant processing account(s) used by a website to accept credit card payments. Its proof that the targeted websites were infringing was based on an in–house investigator's deductive reasoning: the websites were not in Luxottica's approved channel list, the suspect websites had lower prices, and the websites offered shipping to the United States [10].

Luxottica moved for and obtained approval to give adverse parties electronic notice. It gave notice via email and by posting a notice of the lawsuit on the very web properties it seized with the TRO. But none of the defendants answered the summons within twenty–one days. Thus, just two months after filing the complaint, Luxottica had a final, default judgement in the case—for $200,000 per defendant (in theory, up to $95 million) [11].

Luxottica's relief is also typical of cases in the field,[12] and this relief is broad. The Luxottica court found the defaulting defendants liable for willful trademark infringement and counterfeiting, false designation of origin, cybersquatting, and for violating a state consumer protection law. The final judgment gave Luxottica permanent transfer of the 1,024 domains, and seizure of the defendants' PayPal accounts. It also ordered broad categories of unnamed

[8] Amended Complaint, Luxottica Grp. S.p.A. v. The Partnerships and Unincorporated Associations Identified on Schedule "A," No. 1:16-cv-08322 (N.D. Ill. Sept. 1, 2016), ECF No. 30 (granting temporary restraining order in a minute entry).

[9] Luxottica Grp. S.pA. v. The Partnerships and Unincorporated Associations Identified on Schedule "A," No. 1:16-cv-08322 (N.D. Ill. Sept. 7, 2016), ECF No. 31 (reflected posted bond of $10,000

[10] Amended Complaint, Luxottica Grp. S.p.A. v. The Partnerships and Unincorporated Associations Identified on Schedule "A," No. 1:16-cv-08322 (N.D. Ill. Aug. 25, 2016), 2016 WL 8570031.

[11] Amended Complaint, Luxottica Grp. S.p.A. v. The Partnerships and Unincorporated Associations Identified on Schedule "A," No. 1:16-cv-08322 (N.D. Ill. Aug. 25, 2016), 2016 WL 8570031.

[12] We found no cases with the number of defendants in Luxottica, but others follow a similar procedure and have even shorter times to relief. For instance, in one case, a plaintiff received a TRO in one day. See Kate Spade, LLC v. Zhou, No. 1:14-cv-05665 (S.D.N.Y. Aug 28, 2014), ECF 9 (granting TRO in a minute order). We cannot assess the frequency of these suits but they appear to be quite common. A search in Bloomberg Law's Dockets search for civil suits where Chanel was a plaintiff and the keywords "trademark infringement" and "domain" were present returned 163 results. The cases date back to 2001 and were initiated in federal courts all over the country. Twenty-six of the cases were "open" as of May 3, 2017.

businesses not to service the defendants when they were displaying Luxottica's marks. Luxottica's order covered marketplaces (such as eBay and Alibaba), web hosts, sponsored search engine and ad–word providers, credit cards, banks, merchant account providers, third party processors, payment processing service providers, search engines, and domain name registrars. These are all the intermediaries critical to operating a web business.

Examples from Hacking and DDoS

TROs are also the legal tool of choice for public and private enforcement against botnets (Eichensehr, 2017). Botnets are notoriously difficult to police with traditional deterrence by punishment because botnet herders are likely to operate outside the United States [13]. Although botnets are a different security concern, the TRO procedure is remarkably similar to the IP context illustrated earlier. The landscape includes public and private sector collaboration, and the use of civil and criminal mechanisms to obtain information and to seize assets. The basic sequence of events is that either the U.S. government or a sophisticated technology company such as Microsoft files for a TRO in District Court, the TRO is granted under seal, the command–and–control servers are either physically or remotely seized, and finally Microsoft issues a software update that commands infected bots to disengage from the network and cease malicious behavior. The following Sections highlight the use of TROs in the Coreflood, Rustock, and Kelihos cases to illustrate the efficacy and issues to consider.

a) The Coreflood Botnet

Coreflood was a Russian–based botnet that infected computers across the public sector, as well as other critical systems belonging to hospitals, businesses, etc. At its peak, it infected over two million machines, and it could repurpose these computers for several different tasks—to attack other computers with denial–of–service attacks, to provide an anonymous platform for hackers for multi-stage attacks, and to capture user keystrokes, thereby enabling the botnet controllers to discover credit card numbers and bank login information [14]. The privacy and security implications of Coreflood and other botnets are profound, making those infected vulnerable to many different kinds of wrongs.

The government averred that a single Coreflood command server "held approximately 190 gigabytes (GB) of data, recorded from 413,710 infected computers while unsuspecting computer users were browsing the Internet."[15] The government claimed that Coreflood was used to steal six– figure sums from a number of small businesses, even ones that had used two–factor authentication to carefully protect banking accounts.[16]

In a civil complaint, the government obtained a TRO from a district court, along with several search warrants in different districts. The TRO sought by the Justice Department

[13] In 2009, the Federal Trade Commission, using its powers to obtain injunctions for unfair and deceptive practices, took down 3fn, which was regarded as among the last US-based "bulletproof" hosts. Fed. Trade Comm'n v. Pricewert LLC, No. C-09-2407 RMW, 2010 WL 2105614, at 1 (N.D. Cal. Apr. 8, 2010).

[14] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF No. 32.

[15] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF No. 32.

[16] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF No. 32.

authorized the Internet Systems Consortium (ISC), a nonprofit, to swap out privately owned command– and–control servers and turn them over to the government.[17] Once this happened, Microsoft released a patch through its Malicious Software Removal Tool, which instructed machines infected with Coreflood to remove the program.[18]

In requesting the TRO in the civil case, the government argued that obtaining a search warrant was impracticable, explaining that botnet situations justified use of the special needs exception to the general preference that the government obtain a warrant for a search or seizure. The government assured the court that it would not collect any protected information or communications from the computers infected with Coreflood. The court granted the TRO but prohibited the agencies from storing, reviewing, or using information unrelated to the data needed to battle the botnet.[19] Interestingly, although the government obtained the TRO through a civil procedure, the Department of Justice also announced that it would pursue a criminal prosecution.[20] The line between civil and criminal procedure blurs as TROs are used as tools to combat criminal activity, which is why this case reflects the basic due process concerns at play when the government, intermediaries, and nonprofits cooperate on operations that implicate constitutional and statutory interests.

b) The Rustock Botnet

Rustock was a self–propagating botnet that was responsible for a large portion of spam emails worldwide. This botnet used a Trojan virus to infect machines that received spam communications, and was difficult to detect. Several previous attempts to bring down Rustock failed due to its ability to quickly restore its capacity after any partial attack.[21]

Microsoft, in cooperation with Pfizer (which suffered potential reputational and financial harm because Rustock sent spam emails for knock–off Viagra), the U.S. government, and the University of Washington, finally disabled the botnet through Operation b107. Microsoft brought suit in the Western District of Washington, and obtained a TRO that authorized the implementation of the operation under seal.[22] Accompanied by U.S. Marshals, Microsoft seized equipment used in Rustock, performed forensic analysis on it, and concluded that the evidence pointed to a Russian–based operation (Bright, 2011).

Microsoft gained standing to pursue this action under a combination of the CAN-SPAM Act and the Lanham Trademark Act, in part because Microsoft's trademarks are used to propagate malware.[23] Pfizer's involvement was key for invoking the Lanham Act, and for

[17] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF No. 32.

[18] Press Release, U.S. Dep't of Justice, Department of Justice Takes Action to Disable International Botnet (Apr. 13, 2011), justice.gov.

[19] United States v. John Doe 1, No. 3:11-CV-00561 (D. Conn. Apr. 13, 2011), ECF No. 51.

[20] Press Release, U.S. Dep't of Justice, Department of Justice Takes Action to Disable International Botnet (Apr. 13, 2011), justice.gov.

[21] Microsoft Corp., Battling the Rustock Botnet: Security Intelligence Report 7 (2011), https://lammgl.files.wordpress.com/2011/03/battling-the-rustock-threat $_{e}nglish.pdf(\backslash Rustockchecksforthepresenceofkerneldebuggers...and...alsotriestomaintaincodeintegritybyconstantlychecki$

[22] Complaint, Microsoft Corp. v. John Does 1-11 Controlling a Computer Botnet THereby Injuring Microsoft and Its CUstomers, No. 2:11-cv-00222 (W.D. Wash. Mar. 1, 2011), 2011 WL 921612.

[23] Microsoft Corporation's Application for an Emergenct Temporary Restraining Order, Seizure Order,

triggering a sense of urgency—the drugs sold via Rustock were passed off as real, but in test purchases, some proved to differ from those sourced from Pfizer's supply chain. Moreover, Microsoft ensured that the court order was under seal until the operation was complete, so as to avoid tipping off the botnet herders in advance. As in the Coreflood proceedings, the plaintiffs justified their actions by noting that Microsoft would respect due process concerns and that this intervention was the narrowest possible (Bright, 2011). It also filed a $170,000 bond. Microsoft updated the court weeks after it seized IP addresses and domain names to report that it had received no requests to reinstate these resources.[24]

c) The Kelihos Botnet

One ongoing example of government and intermediary efforts to thwart a botnet is the Kelihos case.[25] Kelihos is a botnet that functions in a similar fashion to Rustock by using spam to infect peer computers with malware.[26] In this case, the program is able to conduct a range of operations including DDoS attacks and stealing cryptocurrency wallets. On April 10, 2017, the Justice Department announced that it was undertaking actions to dismantle the botnet. Unlike in the Coreflood case however, the government invoked the 2016 Amendments to Rule 41 of the Federal Rules of Criminal Procedure (FRCrP), instead of Rule 65 of the FRCP.[27] Under the new language, the federal government is able to seek a warrant to search a computer that is hidden through the use of technology (such as anonymizing software like Tor or I2P), and sue in just one jurisdiction in cases where devices in five or more districts are implicated (as opposed to all districts).[28] This is an important step because previously the government struggled to remotely search anonymized criminals, and faced high litigation costs arising from the requirement to sue in multiple districts.[29]

As indicated earlier, the government generally used a combination of TROs from civil procedure and criminal investigations to cooperate with intermediaries in botnet cases. In this case, the government relied solely on criminal procedure. However, despite using the FRCrP instead of the FRCP, the technical procedure used looks to be the same as previous botnet cases. The government got authorization to take control of command–and–control

and Order to Show Cause Re Preliminary Injunction, Microsoft Corp. v. John Does 1-11 Controlling a Computer Botnet Thereby Injuring Microsoft and Its Customers, No. 2:11-cv-00222 (W.D. Wash. Feb 9, 2011), 2011 WL 1193746

[24] Microsoft Corporation's Status Report Re Preliminary Injunction at 2, Microsoft Corp. v. John Does 1-11 Controlling a Computer Botnet Thereby Injuring Microsoft and Its Customers, No. 2:11-cv-00222 (W.D. Wash. Apr. 4, 2011), ECF No. 43.

[25] Press Release, U.S. Dep't. of Justice, Justice Department Announces Actions to Dismantle Kelihos Botnet (Apr. 10, 2017), https://www.justice.gov/opa/pr/justicedepartment- announces-actions-dismantle-kelihos-botnet-0.

[26] Kelihos, N.J. CYBERSECURITY COMMC'S. INTEGRATION CELL (Dec. 28, 2016), https://www.cyber.nj.gov/threat-profiles/botnet-variants/kelihos.

[27] Fed. R. Crim. P. 41(B)(6)B).

[28] Fed. R. Crim. P. 41(B)(6)B).

[29] Press Release, supra note 77; Leslie R. Caldwell, Rule 41 Changes Ensure a Judge May Consider Warrants for Certain Remote Searches, U.S. DEP'T JUST. (June 20, 2016), www.justice.gov/archives/opa/blog/rule-41-changes-ensure-judge-may-considerwarrants- certain-remote-searches.

servers, identified IP addresses, and then turned them over to an intermediary to sever connections between the botnet herder and the servers. As in the previous cases, Microsoft used its software updates to instruct infected computers to delete the virus that propagated the botnet. This case may signal a legal framework that courts will use going forward, but it substantially represents the same combination of the government cooperating with an intermediary to seize servers and dismantle them via a sealed court order.

Criticisms of Rule 65 Interventions

Despite the apparent efficacy of Rule 65 TRO interventions in both trademark and botnet applications, this tool is criticized for potential overbreadth. As demonstrated in Luxottica, federal courts, notorious for their slow processes, place these cases at the top of the docket. Because the invocation of Rule 65 expedites litigation, it is possible to get powerful, broad remedies in a matter of days or weeks. Electronic service further greases the wheels by eliminating the labor–intensive but salient event of being physically served with process. These court orders are also usually heard ex parte, and the restraining order is granted under seal to avoid alerting infringers and perpetrators.

In the IP context, the orders are broad in that they cover a wide variety of actors. The TRO allowed Luxottica to compel action from hundreds of defendants, domain name registrars, payment processors, search engines, online marketplaces, and advertisers. Not only did the TRO reach a massive number of actors (many of which were intermediaries), it compelled action from them within a matter of days. This breadth reflects that judicial harmony with enforcers that pursue infringers alone is inadequate, and therefore courts lean on intermediaries to undertake actions to punish and prevent unlawful behavior.

Annemarie Bridy mounts a strenuous critique of domain seizure in Three Notice Failures in Copyright Law (Bridy, 2016). Bridy argues that seizure without notice to domain owners infringes both First and Fifth amendment rights (Bridy, 2016). Her argument is at its strongest when enforcers seize domains with significant non–infringing purposes, such as file sharing systems. Non–infringing uses may not be apparent to courts, and enforcers may see these services as primarily piracy operations. Enforcers tend to target niche players, and as Bridy explains, innocent users of such systems are presumed guilty (Bridy, 2016).

The botnet context suffers from similar concerns, with the additional problem of creating collateral damage for innocent third parties. For instance, Microsoft requested a TRO to sink several computers that were generating dynamic IP addresses to conduct illegal activities. The TRO was directed at NO-IP.com, but inadvertently took down many sites that were using dynamic IP addresses for legitimate purposes.[30] Users, as well as organizations like the Electronic Frontier Foundation, criticized this action (Hiller, 2014). Again, the controversy stemmed from the sudden nature of the action because the TRO was carried out ex parte

[30] Brief in Support of Application of Microsoft Corporation for an Emergency Temporary Restraining Order and Order to Show Cause Regarding Preliminary Injunction, Microsoft Corp. v. Mutairi, No. 2:14-cv-00987 (D. Nev. Jun 19, 2014); see also Zach Lerner, Microsoft the Botnet Hunter: The Role of Public-Private Partnerships in Mitigating Botnets, 28 HARV. J.L. TECH. 237 (2014).

and under seal. Moreover, Microsoft was criticized for its outsized role in seeking and implementing the legal and technical actions necessary for the TRO. In virtually every case examined, Microsoft, in partnership with the federal government and other companies, was responsible for developing and implementing the software that disrupted botnets. Microsoft's role here is natural for the obvious reason of Microsoft's product being a dominant operating system worldwide, and indeed this is an attractive feature in terms of effectively combating large and diffuse botnets. However, this also means that Microsoft is disproportionately powerful, and can cause unintended harms by pursuing an overbroad TRO. Without any way to raise concerns before implementation, potential victims must rely on Microsoft's and a court's foresight of potential harms to innocent parties.

More generally, there is continued discussion about the extent to which preliminary injunctions may properly conscript intermediaries. Rule 65 orders can only bind certain entities, including parties, entities related to the parties (such as their servants, employees, and agents), and those in "active concert or participation" with the parties (Fed. R. Civ. P. 65(D)(2)(C)). What is the status of payment providers or domain registrars in these cases? Courts do not specify their precise role in orders. For example, in one case, CloudFlare, which provides reverse–proxy service, complied with a preliminary injunction that required it to terminate user accounts that used specific domain names.[31] CloudFlare, however, opposed obligations to filter on a continuing basis for customers using the domain name "grooveshark."[32] In this effort, CloudFlare found an ally in the Electronic Frontier Foundation (EFF), which argued that this preliminary injunction required CloudFlare to act as "enforcers" of the plaintiff's trademark and could potentially affect customers who were not using the domain name in an infringing matter. [33]

This situation contains many parallels to other cases examined here. As noted, the final judgment order in the Luxottica case compelled search engines and online marketplaces to stop serving the defendants, implying an obligation to continue monitoring their systems for infringing behavior.[34] Like with the CloudFlare example, this puts intermediaries in the position of continually enforcing another party's IP rights. While this arrangement is pragmatic, since the fact that infringers will not realistically comply with court orders means that focusing on intermediaries is more effective, intermediaries may challenge overreliance on their capacity and willingness to pursue infringers on behalf of IP owners.

Internet commerce has a different logic than offline business operations. Firms supplying infringing content probably never meet any of the third–party service providers that make their operation possible. Some of the intermediary services may be offered free, or at a very

[31] Arista Records, LLC v. Tkach, 122 F. Supp. 3d 32, 34 (S.D.N.Y. 2015) .

[32] Arista Records, LLC v. Tkach, 122 F. Supp. 3d 32, 34 (S.D.N.Y. 2015).

[33] Mitch Stoltz, Victory for CloudFlare Against SOPA-like Court Order: Internet Service Doesn't Have to Police Music Labels' Trademark, ELECTRONIC FRONTIER FOUND. (July 15, 2015), https://www.eff.org/deeplinks/2015/07/victory-cloudflareagainst- sopa-court-order-internet-service-doesnt-have-police.

[34] Luxottica Grp. S.p.A. v. Zhou Zhi Wei, No. 17-CV-05691, 2017 WL 6994587, at *3 (N.D. Ill. Sept. 12, 2017)

small cost. Additionally, the various intermediaries probably are neither aware of nor wish to be involved with infringement. For these and other reasons, Bridy recommends that enforcers should prove that "nonparty service providers . . . either expressly or tacitly agreed to act in furtherance of a common plan of infringement" (Bridy, 2016). Such a burden of proof would render Rule 65 interventions toothless.

4.5 Government-Led Interventions

This Part details two ways in which the government uses the courts and administrative powers to police intellectual property and computer hacking crimes. The first section covers seizures of websites using the Prioritizing Resources and Organization for Intellectual Property (PROIP) Act.

PRO-IP Act Domain Seizures

Government and intellectual property owners have used domain name seizures to interdict websites that host, or even simply link to, illegal content. Domain names identify things connected to the Internet, and link them to IP addresses. Domain names are considered a core component of Internet governance, and are a fundamental part of establishing property rights on the Internet.

Special legal authority for domain name seizure comes from the PROIP Act, which gave birth to interagency efforts to interdict online IP violations.[35] Basically, the federal government seizes a website accused of engaging in illegal activity, making it impossible to reach it by searching its alphanumeric name. It is still generally possible to reach it by directly entering its numeric IP address. Most consumers probably will never figure this out, so the blocked sites are, in effect, boarded up. Once seized, the government can continue its investigation of the website's alleged infringement, before pursuing further legal action (Kopel, 2013).

DNS seizures are useful because they are effective at targeting internet resources complicit in illegal behavior. For these one–off instances, DNS seizures are easy to undertake, and only require cooperation between the government and domain name registrars. Moreover, they can be used to pursue websites whose owners may be difficult to track down or may live outside the United States, without requiring a lengthy legal proceeding.

Yet, DNS seizures are controversial because of the potential overbreadth and potential lack of regard for process. They can be overbroad in that the government can identify targets for seizure that are not directly related to illegal activity. Relatedly, it can bring down the websites without the defendants showing up in court. Without strong procedural protections, the government can seize domain names that are not actually associated with illegal activity (Bridy, 2016).

[35] Prioritizing Resources and Organization for Intellectual Property Act of 2008, Pub. L. 110-403, 122 Stat. 4256.

For instance, the Rojadirecta and Dajaz1 cases reflected this flaw in PRO-IP Act DNS seizures. Rojadirecta was a website that linked to other websites illegally streaming sportscasts, and Dajaz1 was a website that offered hip-hop commentary and reviews, as well as song samples. Rojadirecta's legality was upheld in Spanish courts two years prior to the U.S. seizure (Martinez, 0). In both cases, the government seized the domain names, but the owners of the websites successfully challenged the orders and regained control of the web properties (Kopel, 2013). In the Dajaz1 case, the U.S. government never came up with the adequate evidence to justify a permanent injunction, and thus handed the domain back to its original owner (Kravets, 2012). In both cases, the domain owners were deprived of the properties for over a year. For even the most successful web businesses, a short service outage can be ruinous.[36]

DNS seizures are criticized for the same reason that civil forfeiture has become widely scrutinized: they take control of property whose owners lack the means to challenge the government's allegations. As was the case with TROs, the Rojadirecta and Dajaz1 cases were heard ex parte and were seized without notifying the owners beforehand. The owners of these domains successfully challenged their seizure, but the vast majority of the more than 1,000 domains seized never challenge the government (Kopel, 2013). The breadth and muscularity of intellectual property rights obviously raises the specter of these mechanisms being used for censorship.

Government Intervention in Financial Services Intermediaries

Aside from court–authorized actions, the U.S. government also uses administrative power to investigate and disrupt cybercrime. Multiple agencies claim jurisdiction over cybercrime because it implicates financial security, protection of critical infrastructure, criminal statutes, and intellectual property protections. As such, both the Justice Department and the Treasury Department launched financial security programs that touch on cybercrime.

Operation Choke Point was a President Obama-era Justice Department program that focused on banks and their business clients (Silver-Greenberg, 2014). Specifically, it targeted certain merchant categories that were recognized as being high– risk. For instance, it covered money laundering, consumer exploitation (scams, payday lenders, etc.), and online gambling. Essentially, the program aimed at uncovering information about exploitative and illegal practices by leveraging banks' access to unique insights about the merchants that banks connect to the payments system (Silver-Greenberg, 2014). The Justice Department, focused on payment providers, targeted banks with subpoenas and investigative attention to determine whether they were aware of or were colluding in fraud perpetrated by partner payment providers.[37] This investigatory attention caused banks to sever relationships with

[36] Puerto 80 Projects, S.L.U. v. United States, No. 11-3983 (S.D.N.Y June 20, 2011).

[37] U.S. HOUSE OF REPRESENTATIVES COMM. ON OVERSIGHT GOV'T REFORM, THE DEPARTMENT OF JUSTICE'S "OPERATION CHOKE POINT": ILLEGALLY CHOKING OFF LEGITIMATE BUSINESSES? (2014), https://oversight.house.gov/wp-content/uploads/2014/ 05/Staff-Report-Operation-Choke-Point1.pdf.

both questionable and lawful merchants, raising the ire of the business community and triggering Congressional blowback. The Trump administration ended operation Choke Point in 2017.[38]

Other examples include President Obama's Executive Order (EO) 13694,108 which was amended by EO 13757.[39] In EO 13694, the U.S. Department of Treasury was authorized to place a block on all property and property interests in the United States that are associated with cybercrime by placing individuals and entities on the specifically designated nationals and blocked persons list (SDN).[40] This intervention is similar to other Treasury holds placed in response to illegal activities,[41] and its authority stems from the International Emergency Economic Powers Act.[42] With this authority, the Treasury, in conjunction with the Justice Department can freeze bank accounts, deplete them, and generally prevent their owners from accessing them. As of this writing, the government has not placed anyone on the 13694 list.

EO 13757, adopted late in President Obama's tenure, amended the earlier order in light of Russian state–sponsored attacks on American presidential candidates Hillary Clinton and Senator Marco Rubio. EO 13757 specified that activities "interfering with or undermining election processes or institutions" trigger designation on the SDN.[43] Over forty individuals and entities have been placed on the SDN under the EO 13757 process.

These programs are useful in that they can target cybercriminals who are not physically located in the United States. In both cases, the government leverages the fact that financial institutions are central to cybercriminal operations. Because much cybercrime is financially motivated, identifying and dismantling perpetrators' financial assets is a key tool for deterring it.

Financial interventions are also more likely to effectively disrupt cybercriminal activity than DNS seizures. Since there are many registrars, DNS seizures may only temporarily take infringing websites offline. A study by Wang and collaborators found that domain name seizures did not significantly reduce the number of counterfeit online stores found in search engine results for luxury goods (Wang et al., 2014).

[38] Letter from Stephen F. Boyd, Assistant Att'y Gen., Dep't of Justice, to the Honorable Bob Goodlatte, Chair, Comm. on the Judiciary, U.S. House of Representatives (Aug. 16, 2017), http://alliedprogress.org/wp-content/uploads/2017/08/2017-8-16- Operation-Chokepoint-Goodlatte.pdf.

[39] Taking Additional Steps to Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities, 82 FED. REG. 1 (Dec. 28, 2016).

[40] Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities, 80 FED. REG. 18,077 (Apr. 1, 2015).

[41] See Transnational Criminal Organizations Sanctions Regulations, 31 C.F.R. § 590 (2017) (using the 2016 classification of PacNet as a significant transnational criminal organization pursuant to E.O. 13581 as an example of SDN interventions against intermediaries for online crime); see also Specifically Designated Nationals and Blocked Persons List (SDN) Human Readable Lists, U.S. DEP'T OF TREASURY (Feb. 2, 2018), https://www.treasury.gov/resource-center/sanctions/SDN-List/Pages/default.aspx [hereinafter Dep't of Treasury SDN].

[42] 50 U.S.C. sec. 1701 (2012)

[43] Taking Additional Steps to Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities, 82 FED. REG. 1 (Dec. 28, 2016).

There is evidence suggesting that DNS seizures are not a one–time intervention, and companies must bring a series of lawsuits to continue pursuing infringers, which may help explain why they do not significantly reduce the number of counterfeit stores in the long–run. Indeed, in the Luxottica cases, the court order also instructed PayPal to restrain payment accounts based in China and Hong Kong, indicating that simply seizing the domain names in question was not an adequate remedy.[44]

4.6 Private Remediation Procedures

Under pressure from intellectual property owners, some market platforms have developed their own takedown policies. Large platforms allow for their users to report IP infringement, and then take actions to remove the infringing listings or transactions. These interventions do not require the explicit consent of law enforcement, and rather reflect the intermediaries' effort to mitigate the harm done by cybercriminals. The following Section details some examples of these mechanisms, and then discusses the general advantages and disadvantages of self–regulation (Liu et al., 2015).[45]

eBay VeRO Program

eBay's Verified Rights Online (VeRO) Program is geared towards helping IP owners prevent sellers from illegally marketing merchandise, unauthorized copies, and other branded materials. The process largely relies on IP owners reporting infractions to eBay, but provides participants with a few different options for large–scale and chronic infringements. This program provides a concrete example of how an online marketplace takes actions against infringers. eBay is a particularly good example because virtually all of the products it sells are supplied by users, therefore breaking the channel controls that some luxury brands use to maintain vertical price fixing and exclusivity. It is thus important to understand that brand owners may be objecting to any sale of their branded merchandise in addition to items that are infringing or counterfeit.

The procedure for VeRO is straightforward and easily accessible. Anyone (including people and companies that do not have listings on eBay) who owns a product or piece of intellectual property is eligible to participate. An interested party must sign up for a VeRO account and provide links to the infringing products. Then, the user emails "vero@ebay.com"

[44] Luxottica Grp. S.p.A. v. Zhou Zhi Wei, No. 17-CV-05691, 2017 WL 6994587, at 3 (N.D. Ill. Sept. 12, 2017).

[45] The survey here includes efforts focused on large platforms that control a huge transaction space, such as eBay and the Visa payment network. However, the literature includes discussions of countless other private remediation programs. For instance, Liu et al. explore policy changes that affect domain name acquisition in the .cn ccTLD and the effects of a verification service that screened domains for illegal pharmacy activities.

to notify eBay that s/he would like to assert IP rights. Finally, the party submits a "Notice of Claimed Infringement (NOCI)" form.[46]

This process is geared toward individual violations, but naturally some parties may have larger needs. eBay provides for users to search for IP infringement through manual monitoring, setting up a "Watch List," or hiring a full–time monitoring agency. eBay imposes no fee for reporting infringement or for creating watch lists, but companies incur expenses in employee time or in hiring boutique monitoring services.[47]

Visa IP Enforcement

Visa, like MasterCard, is a payment network, an ISP–like entity for banks and merchants that exchange money in order to process consumer purchases. Visa thus can monitor aspects of transactions but it cannot track the specific items purchased by the consumer (Hoofnagle et al., 2012). However, Visa can monitor suspicious merchants and link their activity across different banks.

Visa voluntarily searches for potential IP infringement in its payment systems, and attempts to enforce IP owners' rights.[48] Visa has at least two different procedures that it uses for its IP takedown activities, one of which is an online form that victims can fill out identifying a merchant who has infringed on IP. The website claims that individuals may file five claims per month, and afterward Visa investigates each claim and arbitrates.[49]

More detailed information comes from a 2011 congressional testimony by Visa on the issue of IP takedowns. Visa explained that it deals with complaints directly via emails to "Inquiries@visa.com." One important note is that Visa and other credit card companies do not generally have direct relationships with individual merchants who accept their cards as payment. Instead, merchants have relationships with payment companies that link them to the network. After receiving a complaint, Visa does a test transaction to identify the payment company that signed up the suspected infringing merchant. Visa then instructs the payment company to investigate the merchant, and report within five business days.127 After reviewing the report, Visa has the payment company send a "comply or terminate" notice to the suspected infringer.

[46] Notice of Claim Infringement, EBAY http://pics.ebay.com/aw/pics/pdf/us/help/ community/NOCI1.pdf (last visited Feb. 2, 2018).

[47] General information about the program is available on eBay's website. See Verified Rights Owner Program, EBAY, http://pages.ebay.com/seller-center/listing/createeffective- listings/vero-program.html (last visited Feb. 2, 2018). A list of participating members is also available. See VeRO Participant Profiles, EBAY, http://pages.ebay.com/ seller-center/listing/create-effective-listings/vero-program.htmlm17-1-tb3 (last visited Feb. 2, 2018).

[48] Targeting Websites Dedicated to Stealing American Intellectual Property: Hearing Before the S. Comm. on the Judiciary, 112th Cong. 1 (2011) (statement of Denise Yee, Visa, Inc.), https://www.judiciary.senate.gov/imo/media/doc/11-2-16Yee%20Testimony.pdf.

[49] Report Intellectual Property Abuse, VISA, https://usa.visa.com/Forms/report-ipabuse- form.html (last visited Feb. 2, 2018).

International Anticounterfeiting Coalition (IACC)

The IACC is a nonprofit that brings together various actors concerned with international IP infringement (Bridy, 2016). The organization is composed of over 250 member organizations, including private businesses, law firms, security firms, and government organizations. It also hosts semiannual conferences dedicated to informing members about best practices, and coordinate efforts to clamp down on IP infringement.

The IACC offers a suite of services to its members, namely the "RogueBlock" and "MarketSafe" features. RogueBlock is a back–end network that connects IP owners to investigators, payment companies, the government, and related actors.132 When infringement occurs, the IACC processes reports and distributes them to the relevant intermediaries on behalf of its members.

MarketSafe is a direct partnership between the IACC and Alibaba to take down counterfeiting infringers on the online marketplace, Taobao.[50] It includes access to "expedited take–down procedures" that presumably guarantee members a quick turnaround on their reports of IP infringement on the website. Essentially, the IACC provides investigative and administrative services to its members by specializing in searching for infringement, producing relevant evidence, and filing the proper documentation in IP takedown cases.

Backpage.com: Private Remediation as a Scaffold for Criminal Prosecution

Backpage.com is a popular online classified ads site, similar to Craigslist. But Backpage is known for its adult escort ads, which are believed among the not–born–yesterday to be a front for organizing online prostitution and child sex trafficking.[51] Experts in human trafficking believe that Backpage does not simply provide a substitute for offline child sex markets, but rather contributes to an explosive growth in reports of child sex trafficking: astonishingly, the National Center for Missing and Exploited Children claims that 73% of the child sex trafficking reports it receives involve Backpage.

Years ago, law enforcement agencies pressured credit card networks to stop accepting payments initiated on Backpage.[52] By July 2015, American Express, Visa, and MasterCard all agreed to stop such payments.[53] In a December 2016 criminal complaint, the State of California charged Backpage's operators with money laundering and conspiracy for creating

[50] IACC MarketSafe®, INT'L ANTICOUNTERFEITING COAL., http://www.iacc.org/ online-initiatives/marketsafe (last visited Feb. 2, 2018).

[51] S. COMM. ON HOMELAND SEC. GOV'T AFFAIRS, BACKPAGE.COM'S KNOWING FACILITATION OF ONLINE SEX TRAFFICKING 1 (2017) [hereinafter REPORT ON BACKPAGE.COM].

[52] Rebecca Hersher, Backpage Shuts Down Adult Ads in the U.S., Citing Government Pressure, NPR (Jan. 10, 2017, 11:23 AM), https://www.npr.org/sections/ thetwo-way/2017/01/10/509127110/backpage-shuts-down-adult-ads-citing-governmentpressure.

[53] Michelle Hackman, Backpage Files Suit Against Cook County Sheriff Over Credit Card Service, WALL ST. J. (July 21, 2015, 2:35 PM), https://www.wsj.com/ articles/backpage-files-suit-against-cook-county-sheriff-over-credit-card-service- 1437496670.

fake e-commerce sites to evade American Express' payment ban. The state alleges that the defendants instructed escorts and pimps on how to buy "credits" on these third party sites that were actually destined for Backpage's escort business.

The State charged the Backpage operators with financial crimes because an earlier attempt to prosecute them ended in failure—a state court judge held that under CDA 230, the operators we not liable for the classified ads posted by third parties.[54] The State brought the new charges just weeks after the failed prosecution.[55]

Comments on Voluntary and Self-Regulating Procedures

Voluntary procedures avoid use of the courts, thereby avoiding costs and delays. Moreover, they allow for the victims of infringement and fraud to directly deal with the infringement. These are important advantages because they avoid the costs associated with lawsuits, and encourage victims to take advantage of these policies. Because of these features, these platforms establish credibility in their services. As seen in the Backpage.com example, voluntary procedures can also lay the groundwork for government enforcement actions.

On the other hand, the lack of transparency obscures actual practices and subtle shifts in policy. Self–regulatory procedures hide the actual penalties levied by intermediaries on various actors. They also make it possible for the intermediary to weaken its posture over time, perhaps by reducing penalties once scrutiny from enforcers eases. Self–regulatory procedures can hide awful practices that indicate the most noxious uses of the platform—for instance, Backpage.com was filtering terms that indicated child sex trafficking such as "Amber Alert." Finally, a lack of transparency obscures how self–regulatory systems distribute seized proceeds from suspected cybercrime.

Another major disadvantage to these approaches is that they all rely on self–reporting from the victim. Although eBay and Visa allow for some automation in their services, they are not inherently designed to deal with large–scale fraud or theft. By default, they are designed to deal with individual complaints, which means they are probably more effective at isolated incidents involving smaller victims.

This feature makes voluntary efforts difficult to rely on when dealing with botnets, large crime networks, and systemic fraud. Although organizations like the IACC attempt to clean up marketplaces like Taobao, self–regulation on its own does not necessarily alleviate the structural problems with these platforms. Perhaps this is why some enforcers have pursued litigation or administrative enforcement actions.

[54] Trial Order, People v. Ferrer, No. 16FE019224 (Cal. Super. Ct. Nov. 16, 2016), 2016 WL 6905743.

[55] Felony Criminal Complaint, People v. Ferrer, No. 16FE024013 (Cal. Super. Ct. Dec. 23, 2016), 2016 WL 7884408.

4.7 Summary and Conclusion

Cybercrime is often presented as an intractable problem because it can be committed by users under a cloak of anonymity and committed from jurisdictions without effective rule of law. Intermediaries are presented as being broadly shielded from liability for their users' actions. This Article explains that these frames obscure the reality of deterring financially–motivated cybercrime: such cybercrime shares characteristics of ordinary businesses. Like ordinary businesses, financially–motivated cybercrime is an activity of scale, not a jackpot activity such as robbing a bank. Criminals need to optimize their processes, make sales, and critically, they rely on many different intermediaries for everything from marketing, to web hosting, to delivery of products. Reliance on intermediary service providers gives enforcers the opportunity to disrupt these networks. While CDA 230 provides intermediaries great cover for demands to take down some material, anti–botnet and IP enforcers have found some success using FRCP Rule 65 to compel intermediaries to hand over or block resources used by cybercrime networks, typically within days of filing suit.

Intermediaries are in a tussle among law enforcement, powerful brands, legitimate users, and rogue users. Enforcers have found effective technical fixes (sinkholes, delisting a website's alphanumeric name, etc.), yet there is no one simple solution that works across all classes of crime.

Narrower gateways offer more powerful interventions. For instance, a prior study by author McCoy and collaborators found that payment platforms, because of their breadth and oligopoly status, have more power over cybercriminals than interventions in the DNS (McCoy et al., 2012). There is more competition in domain name administration, and far too many top–level– domains (e.g. .com, .net, and so on) to control the entire space (Liu et al., 2015).

Moreover, these types of interventions can cause serious collateral damage that disrupt legitimate operations or otherwise impose costs on legitimate users. Considering that these interventions require intensive cooperation among the government, nonprofits, and corporate actors, the motivations of these actors must be balanced as to not interfere with other public policy goals like fair market competition, strong privacy protections, and encouragement of innovation.

It is unlikely that enforcement approaches focused on intermediaries will cause decentralization and turns to harder–to–disrupt technologies, such as cryptocurrencies. This is because financially–motivated cybercriminals need to appeal to a mass consumer population. For these consumers, PayPal and similarly–mainstream payment mechanisms are accessible, whereas decentralized ones are difficult to use and generally go unused by ordinary consumers.(Vigna, 2017) For instance, author McCoy and colleagues found that blocking DDoS–for–hire services from PayPal caused an almost immediate, short term reduction in availability of such services. The McCoy team observed that a DDoS service that only accepted cryptocurrency Bitcoin had a two percent conversion to paid subscriber rate, while two competitors that accepted PayPal had fifteen percent and twenty–three percent conversion rates, respectively (Karami et al., 2016). At least for financially–motivated criminal enterprises that depend on sales to average consumers—the purchasers of online pharmaceuticals and counterfeit handbags discussed in this Article—profitmaking will depend on low transaction costs and simple access procedures for consumers. Skeptics may invoke the technically–shrouded, sophisticated marketplace Silk Road as a counternarrative, but Silk Road was small in comparison to the enormity of the international drug trade and there is some evidence that it served a business–to–business function for drug dealers (Aldridge and Decary-Hetu, 2014). Presumably drug dealers finding a supply for drugs to resell would be more motivated to learn the intricacies of cryptocurrencies, but many ordinary consumers cannot.

Enforcers will likely continue their focus on intermediaries to police their brands and to break up botnets. These efforts raise concerns over due process, property rights, and privacy rights. This Article shows that IP enforcers are able to take control over thousands of domain names, including those that include goods other than the infringing items. Interventions are often done ex parte, and may not require notice to the affected websites under Rule 65. In fact, attacks on botnets must omit this notice for fear that cybercriminals can avoid the attempts to sinkhole the botnet. It is important that interventions reflect appropriate humility in light of the lack of adversarial process.

Chapter 5

Conclusion

Collectively, the preceding chapters pave a promising path forward for academics and policymakers who are interested in privacy and cybersecurity issues. In particular, these pieces show that a deterrence by denial strategy is not only feasible to enact with pre-existing regulatory frameworks, but also effective in making cybercrime a manageable problem. Conventional accounts of cybercrime treat perpetrators as anonymous, diffuse, and unpredictable. However, these accounts underestimate the extent to which cybercrime is financially motivated, and thus focused on certain types of organizations. In these pieces, I show how publicly trade companies, medical providers, and online intermediaries are focal points for cybercriminal activity, and how the law deter that activity by regulating these organizations. The law & economics of crime implies that one way to deter crime is by increasing the cost of punishment and the probability of being aprehended. In contrast, these chapters show that cybercrime may be most effectively combated by instead denying cybercriminals benefits.

This dissertation also makes a contribution by advancing empirical research in cybersecurity law. Although privacy, cybercrime, and cybersecurity are active areas of inquiry within the legal literature, there is relatively little empirical scholarship in this area. Empirical legal scholars are confronted with a lack of data, and traditional social science methods struggle to adapt. Meanwhile, the computer science literature addresss a variety of technical questions with regards to privacy and security, but these results do not always seamlessly translate into policy contexts. By integrating questions and methodological approaches of both the legal and technical literatures, I highlight the potential for using unconventional data sources to answer outstanding questions in cybersecurity law and policy.

Even more broadly, these pieces illustrate various ways how interdisciplinarity that draws on law, social science, and data science can flourish within empirical legal studies. The chapters on predicting cybercrime and estimating the effects of data breach notification laws in particular demonstrate the utility of data-adaptive methods in empirical legal studies. Machine learning and text analysis open up a plethora of research questions centered around prediction, and Chapter 2 provides an example of what applied work can look like in law. The synthetic control method used in chapter 3 shows how causal inference can also be enriched, and provides an alternative to regression that is intuitive and interpretable, which are attractive features for scholarship that is oftentimes aimed at non-technical audiences. Empirical legal studies already has a rich history of drawing upon economics, pscyhology, and other quantitative disciplines to frame research questions and borrow methodological toolkits. As data science continues to transform academic disciplines, computer science and statistics will surely become core parts of the empirical legal studies family. By showing the creativity and power that a data science framework provides, these pieces will hopefully become part of the first wave of scholarship that applies data science to legal questions.

Each chapter helps build a bridge between law, social science, and data science, and in doing so establishes a foundation for future research into data protection law and policy. Looking forward, exploring different ways to build datasets that further enrich possibilities for empirical research in this area should be a top priority. This dissertation focuses on the role of various disclosure regulations, but there are other aspects of data protection law that are ripe for empirical analysis as well. Methods in machine learning, natural language processing, and causal inference continue to evolve rapidly, and these advances present fascinating opportunities for applied researchers as well. The confluence of the emergence of privacy law as a major concern in law and policy discourse and the maturation of data science within academic research will hopefully lead to a strong research agenda for data protection. I hope that this dissertation spurs further work at these intersections, and becomes an early example of what will eventually be an exciting and dynamic body of scholarly work.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*.

Abadie, A. and J. Gardeazabal (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review 93(1)*, 113–132.

Aldridge, J. and D. Decary-Hetu (2014, May). No and 'Ebay for Drugs': The Cryptomarket 'Silk Road' As a Paradigm Shifting Criminal Innovation. *SSRN Working Paper*.

Alrwais, S., X. Liao, X. Mi, P. Wang, X. Wang, F. Qian, R. Beyah, and D. McCoy (2017). Under the Shadow of Sunshine: Understanding and Detecting Bulletrpoof Hosting on Legitimate Service Provider Networks. *IEEE Symposium on Security and Privacy*.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*.

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2017). Matrix Completion Methods for Causal Panel Data. *arXiv1710.10251v2*.

Athey, S. and G. W. Imbens (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *The Journal of Economic Perspectives 31(2)*, 3–32.

Bauguess, S. W. (2017, June). The Role of Big Data, Machine Learning, and AI In Assessing Risks: A Regulatory Perspective.

Ben-Michael, E., A. Feller, and J. Rothstein (0). The Augmented Synthetic Control Method. *arXiv:1811.04170*.

Bradford, A. (2020). *The Brussels Effect.*

Brandeis, L. D. (1914). *Other People's Money and How the Bankers Use It.* Frederick A. Stokes.

Bridy, A. (2015). Internet Payment Blockades. *Florida Law Review 67*, 1523.

Bridy, A. (2016). Three Notice Failures in Copyright Law. *Boston University Law Review*, 777.

Bright, P. (2011, March). How Operation b107 Decapitated the Rustock Botnet. *ARS TECHNICA*.

Cardenas, A., S. Radosavac, J. Grossklags, J. Chuang, and C. J. Hoofnagle (2010, September). An Economic Map of Cybercrime.

CFR (2016). 31 C.F.R. § 1020.200 et. seq.

Chachra, N., D. McCoy, S. Savage, and G. M. Voelker (2014). Empirically Characterizing Domain Abuse and the Revenue Impact of Blacklisting. *Proceedings of the Workshop on the Economics of Information Security*.

Christin, N. (2013). Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace. *22nd International Conference on World Wide Web*.

Cooke, E., J. Farnam, and D. McPherson (2005). The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets. *Proceedings of the 2005 Steps to Reducing Unwanted Traffic on the Internet Workshop*.

Cooter, R. and T. Ulen (2016). *Law and Economics, 6th Edition.* Berkeley Law Books.

Cowley, S. (2017, October). 2.5 Million More People Potentially Exposed in Equifax Breach. *The New York Times*.

Eichensehr, K. (2017). Public-Private Cybersecurity. *Texas Law Review* (95).

for Data Science, C. and P. P. at the University of Chicago (0). Temporal Cross-Validation.

Goel, S. and H. A. Shawky (2009, October). Estimating the Market Impact of Security Breach Announcements on Firm Values. *Information & Management*, 404–410.

Goel, S. and H. A. Shawky (2014, January). The Impact of Federal and State Notification Laws on Security Breach Announcements . *Communications of the Association for Information Systems 34*.

Goldman, Z. K. and D. McCoy (2016). Deterring Financially Motivated Cybercrime. *Journal of National Security Law & Policy 8*(595), 595–619.

Goldsmith, J. and T. Wu (2006). *Who Controls the Internet? Illusions of a Borderless World.*

Hiller, J. S. (2014). Civil Cyberconflict: Microsoft, Cybercrime, and Botnets. *Santa Clara High Tech Law Journal 31*, 163.

Hoofnagle, C. J. (2007). Identity Theft: Making the Known Unknowns Known. *Harvard Journal of Law & Technology 21*(1), 98–122.

Hoofnagle, C. J. (2008). Towards a Market for Bank Safety. *Loy. Consumer L. Rev. 21*.

Hoofnagle, C. J., I. Altaweel, and N. Good (2017). Online Pharmacies and Technology Crime. *Routledge Handbook of Technology, Crime, and Justice*.

Hoofnagle, C. J., J. M. Urban, and S. Li (2012). Mobile Payments: Consumer Benefits & New Privacy Concerns. *Berkeley Center for Law & Technology Research Paper*.

Houk, M. (2016). Counterfeit Goods. *Materials Analysis in Forensic Science*.

Jung, J., C. Concannon, R. Shroff, S. Goel, and D. G. Goldstein (2017, April). Simple Rules for Complex Decisions. *arXiv:1702.04690v3*.

Karami, M., Y. Park, and D. McCoy (2016). Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services. *Proceedings of the 25th International Conference on World Wide Web*.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015). Prediction Policy Problems. *American Economic Review*.

Kogan, S., D. Levin, B. R. Routledge, and J. S. Sagi (2009, June). Predicting Risk from Financial Reports with Regression. *Human Language Technologies: The 2009 Conference of the North American Chapter of the ACL*.

Kopel, K. (2013). Operation Seizing Our Sites: How the Federal Government is Taking Domain Names Without Prior Notice. *Berkeley Technology Law Journal 28*, 859.

Kravets, D. (2012, May). Feds Seized Hip-Hop Site for a Year, Waiting for Proof of Infringement. *WIRED*.

Krebs, B. (2014). *Spam Nation: The Inside Story of Organized Cybercrime - From Global Epidemic to Your Front Door*.

Lapowsky, I. (2018, April). Facebook Exposed 87 Million Users to Cambridge Analytica. *WIRED*.

Leuz, C. and P. Wysocki (2016, February). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestion for Future Research. *Journal of Accounting Research*.

Levchenko, K., A. Pitsilldis, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage (0). Click Trajectories: End-to-End Analysis of the Spam Value Chain.

Lichtman, D. and E. C. Posner (2006). Holding Internet Service Providers Accountable. *The Law and Economics of Cybersecurity*.

Liu, Y., A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, and M. Liu (2015). Cloudy With a Chance of Breach: Forecasting Cyber Security Incidents. *Proceedings of the 24th USENIX Security Symposium*.

Martinez, J. (0, August). US Government Dismisses Piracy Case Against Rojadirecta Site. *Hill*.

Mayer, J. (2016). Cybercrime Litigation. *University of Pennsylvania Law Review*.

McCoy, D., H. Dharmadasani, C. Kreibich, G. M. Voelker, and S. Savage (2012). The Role of Payments in Abuse-advertised Goods. *Proceedings of the 2012 ACM Conference on Computer and Communications Security*.

McCoy, D., A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko (2012). PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs. *Proceedings of the 21st USENIX Conference on Security Symposium*.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013, September). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301:3781v3*.

Mitts, J. (2014). Predictive Regulation. *SSRN Working Paper*.

Mitts, J. and E. Talley (2018). Informed Trading and Cybersecurity Breaches. *Harvard Business Review*.

Moore, T., R. Anderson, and R. Clayton (2009). The Economics of Online Crime. *The Journal of Economic Perspectives 23*(3), 3–20.

Mulligan, D. K. and F. B. Schneider (2011). Doctrine for Cybersecurity. *Daedalus*.

Musiani, F., D. L. Cogburn, L. DeNardis, and N. S. Levinson (2016).

Needles, S. A. (2009). The Data Game: Learning to Love the State-Based Approach to Data Breach Notification Law,. *North Carolina Law Review*.

Poulsen, K. (2017, July). Putin's Hackers Now Under Attack - From Microsoft. *Daily Beast*.

Regan, P. M. (2009). Federal Security Breach Notifications: Politics and Approaches.

Romanosky, S. and Z. K. Goldman (2016, November). What is Cyber Collateral Damage? And Why Does it Matter? *LawFare Blog*.

Romanosky, S., R. Telang, and A. Acquisti (2011). Do Data Breach Disclosure Laws Reduce Identity Theft? *Journal of Policy Analysis and Management 30*(2), 256–286.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* (75(371)), 591.

Silver-Greenberg, J. (2014, January). Justice Department Inquiry Takes Aim at Banks' Business With Payday Lenders. *The New York Times*.

USC (2012a). 12 U.S.C. § 635(i).

USC (2012b). 17 U.S.C. § 512 (2012) Limitations on liability relating to material online.

USC (2012c). 47 U.S.C § 230 Protection for private blocking and screening of offensive material.

Vigna, P. (2017, April). People Love Talking About Bitcoin More than Using It. *Wall St. Journal*.

Wang, D. Y., M. Der, M. Karami, L. Saul, and D. McCoy (2014). Search + Seizure: The Effectiveness of Interventions on SEO Campaigns. *Proceedings of the 2014 Conference on Internet Measurement*.

WIRED (2002). Gates Finally Discovers Security. *WIRED*.