THREE ESSAYS ON CYBERSECURITY-RELATED ISSUES

By HE LI

A dissertation submitted to the

Graduate School- Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Dr. Miklos A. Vasarhelyi

and approved by

____

Dr. Miklos A. Vasarhelyi

____

Dr. Won G. No

____

Dr. Alexander Kogan

____

Dr. J. Efrim Boritz

Newark, New Jersey

October 2017

ABSTRACT OF THE DISSERTATION

THREE ESSAYS ON CYBERSECURITY-RELATED ISSUES

By HE LI

Dissertation Director: Dr. Miklos A. Vasarhelyi

This dissertation consists of three essays that examine cybersecurity-related matters. In the first essay, I investigate whether external auditors respond to cyber incidents by charging higher audit fees and whether they price cybersecurity risk before the actual event happens when there is no explicit requirement from the regulators. Findings in the essay suggest that cyber incidents lead to increase in audit fees, and the increase is smaller for firms with prior cybersecurity risk disclosures. In addition, firms with repeated cyber incidents or cyber incidents that involve intellectual property experience larger increases in audit fees. However, auditor’s concern over cyber incidents is mitigated by monitoring from large and sophisticated external stakeholders.

The second essay examines the informativeness of cybersecurity risk disclosure and provides three main results. First, both the presence and length of cybersecurity risk disclosure are informative of future reported cyber incidents. Second, market participants are using information conveyed by the presence of cybersecurity risk disclosure, but not the information content which is measured by the adjusted length of the disclosure. Third, the presence of cybersecurity risk disclosure is no longer significantly associated with subsequently reported cyber incidents after the passage of cybersecurity disclosure guidance. However, the essay fails to find a significant association between firm-specific disclosure and cyber incidents.

In the third essay, issues regarding assurance on cybersecurity are discussed. In particular, I argue that data analytics should be an integral part of cybersecurity assurance, and introduce a process of using data analytics in testing cybersecurity controls. Illustrative examples of the process using synthetic data are provided to demonstrate that data analytics is a well-suited approach for providing assurance on cybersecurity. A set of critical challenges for applying data analytics in the assurance engagement are also discussed.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the help and support from my committee members, my colleagues, and my family. It is my great pleasure to convey my gratitude to them in my humble acknowledgment.

Above all, I would like to express the deepest appreciation to my advisor Dr. Miklos A. Vasarhelyi for his constant guidance, patience, support, and suggestions. He introduced me to the world of research and gave me the opportunity to work with talented people in the research center. I am very grateful to Dr. Won G. No for guiding me step-by-step in my dissertation journey. I learnt so much from his persistence and pursuit of excellence that I would benefit for the rest of my academic career. I deeply thank Dr. Alexander Kogan and Dr. J. Efrim Boritz for serving on the committee and providing valuable insight for my studies. Also, I would like to express my gratitude to Dr. Michael Alles, Dr. David Wang, and Dr. Kevin C. Moffitt for their important remarks.

It has always been a pleasure to exchange ideas with others. My special thanks go to my colleagues and friends at Rutgers: Zhengjie Sun, Feiqi Huang, Yunsen Wang, Yue Liu, Cheng Yin, Ting Sun, Qi Liu, Jun Dai, Pei Li, Kyunghee Yoon, Tiffany Chiu, Qiao Li, Zhaokai Yan, and Deniz Appelbaum.

Finally, I would like to thank my family members for their encouragement that allows me to pursue my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Cybersecurity is “the process of implementing and operating controls and other risk management activities to protect information and systems from security events that could compromise them and, when security events are not prevented, to detect, respond to, mitigate against, and recover from those events in a timely manner”[1]. Cybersecurity has attracted much attention in recent years. Both the general public and the business world are concerned about the growing cybercrimes that expose sensitive personal information, cause business disruptions, or steal trade secrets. PricewaterhouseCoopers (2016) reports that the average number of detected cyber incidents increased 38% and the theft of “hard” intellectual property increased 56% in 2015 compared with 2014. More than 20% of the breached firms experienced substantial loss of revenue, customer base, and business opportunities, and most of the breached firms spent millions of dollars improving defense technologies and expanding security procedures following the attacks (CISCO, 2017). Due to the potential impact on firm value and operation, firm executives are treating cybersecurity as one of the top priorities. About 88% of U.S. Chief Executive Officers (CEOs) are concerned that cyber threats could hinder the growth of their firms (Loop, 2016). Likewise, investors are clamoring for more information about cybersecurity risks and data breaches, and how firms are addressing those risks (Shumsky, 2016). To respond to the increasing cyber threats, the Securities and Exchange Commission (SEC) held a roundtable discussion to deliberate cybersecurity landscape and cybersecurity disclosure issues (SEC, 2014). The Standing Advisory Group of the Public Company Accounting Oversight Board (PCAOB) also discussed the potential implications of cybersecurity on

[1] The Cybersecurity Working Group of the AICPA Assurance Services Executive Committee.

financial reporting and auditing (PCAOB, 2014). Particularly, the SEC's Division of Corporation Finance issued a disclosure guidance regarding cybersecurity in 2011 to assist firms in assessing what, if any, disclosures should be provided related to cybersecurity risks and cyber incidents (SEC, 2011).

The purpose of this dissertation is to investigate several important yet unresolved issues regarding cybersecurity. Specifically, the second chapter examines the reaction of external auditors in the event of cyber incidents. Prior studies have examined various roles, including board members, top executives, and internal auditors, in addressing cybersecurity risks and cyber incidents (Higgs, Pinsker, Smith, & Young, 2014; Kwon, Ulmer, & Wang, 2013; Steinbart, Raschke, Gal, & Dilla, 2016; Steinbart, Raschke, Gal, & Dilla, 2013; Zafar, Ko, & Osei-Bryson, 2015). However, academic research remains silent on whether external auditors respond to cyber incidents experienced by their clients, and whether they consider cybersecurity risks prior to the materialization of the risk in the absence of mandatory regulatory requirement for auditors to address cybersecurity risks. This gap is surprising given the increased attention from regulators such as the Center for Audit Quality (CAQ) and the PCAOB. Using an audit fee change model, I find a significant positive relationship between increases in audit fees and cyber incidents. Furthermore, increases in audit fees following cyber incidents are smaller for those with prior cybersecurity risk disclosure, implying that auditors price material cybersecurity risk prior to the cyber-attacks and thus are responding less severely (are less surprised) when the actual event happens. In addition, firms with repeated cyber incidents or cyber incidents that involve intellectual property experience larger increases in audit fees. Finally, external monitoring, as measured by the percentage of institutional holdings and number of block holders, can mitigate auditor's concern over cyber incidents.

The third chapter of the dissertation studies the informativeness of cybersecurity risk disclosures in the risk factor section of annual report (i.e., item 1A in the 10-K). Since 2005, the SEC mandated firms to describe "the most significant factors that make the offering speculative or risky" in Item 1A of 10-K filed after December 1, 2005 with the object to "to provide investors with a clear and concise summary of the material risks to an investment in the issuer's securities" (SEC, 2005). Practitioners criticize that managers are likely to provide vague risk disclosure and simply list all uncertainties they face, providing little information for investors (Reuters, 2005). Results in the third chapter suggest that cybersecurity risk disclosure is largely informative. Specifically, both the presence and the content as measured by the adjusted length of cybersecurity risk disclosure are associated with subsequently reported cyber incidents. There is a substantial increase in the percentage of firms that disclose cybersecurity risks following the SEC's disclosure guidance. However, the presence of cybersecurity risk disclosure is no longer associated with future reported cyber incident in the post-guidance period. Contrary to Hilary, Segal, and Zhang (2017), I find evidence that the market reaction following the cyber incident is positively associated with firm's prior presence of cybersecurity risk disclosure, but not the length of the disclosure. To examine the SEC's concern that more firm-specific disclosure may compromise firm's cybersecurity effort by providing a roadmap to malicious parties, two measures are created based on the bag-of-words approach to capture firm-specific disclosure, but I fail to find a significant association between cyber incident and any of my two measures. Furthermore, the topic analyses show that business disruption and financial performance are the two major concerns regarding cybersecurity and remain relatively steady over time. Concerns over intellectual property and reputation, on the other hand, are relatively low but are increasing rapidly in recent years.

The fourth chapter discusses issues surrounding cybersecurity assurance. Growing cyber threats have prompted board members, analysts, investors, business partners, and regulators to demand information regarding how firms are managing cybersecurity risks. Realizing that there is no consistent and common language for describing cybersecurity risk management programs, the AICPA Assurance Services Executive Committee (ASEC) has developed a cybersecurity risk management reporting framework for firms to communicate information regarding cybersecurity risk management efforts and for practitioners [2] to examine and report on the management-prepared cybersecurity information[3]. Along with the reporting framework, the AICPA' ASEC Cybersecurity Working Group, in conjunction with the Auditing Standards Board (ASB), introduced an attestation guide named *Reporting on an Entity's Cybersecurity Risk Management Program and Controls* to assist practitioners to opine on the cybersecurity report. Although the attestation guide points out detailed requirements for practitioners at different stages in a cybersecurity assurance engagement, it contains limited guidance on how to systematically evaluate cybersecurity risks in the engagement, how to collect evidence pertaining to specific risks, and how to use the evidence in assessing the risks. The fourth chapter attempts to address the above issues from a data analytics perspective. Particularly, the chapter first discusses why data analytics should be an integral part of cybersecurity

[2] AICPA (2017c) refers a Certified Public Accountant (CPA) performing an attestation engagement as a practitioner. Accordingly, this dissertation uses "practitioner" rather than "auditor" throughout the paper to refer to a CPA in an engagement other than financial audit.

[3] See at https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/pages/cyber-security-resource-center.aspx.

assurance. A process of using data analytics in testing cybersecurity controls is then introduced, pointing out considerations that practitioners may need to have in the engagement. Illustrative examples of the process using synthetic data are also presented to demonstrate that data analytics is a well-suited approach for providing assurance on cybersecurity. Finally, I discuss a set of critical challenges for applying data analytics in the assurance engagement in the last section.

The remainder of this dissertation proceeds as follows: Chapter 2 examines the association between external audit and cyber incidents. Chapter 3 studies the informativeness of cybersecurity risk disclosure. Chapter 4 presents discussions on data analytics and cybersecurity assurance. The last chapter concludes this dissertation and discusses limitations.

CHAPTER 2: ARE EXTERNAL AUDITORS CONCERNED ABOUT CYBER INCIDENTS? EVIDENCE FROM AUDIT FEES

2.1 INTRODUCTION

Cybersecurity issues have attracted much attention in recent years, especially after several high-profile cybercrimes such as the data breach at Target Corporation[1] and the hacking attack at Sony Pictures Entertainment[2]. PricewaterhouseCoopers (2016) reports that the average number of detected cyber incidents increased 38% and the theft of "hard" intellectual property increased 56% in 2015 compared with 2014. To respond to the increasing cybersecurity threats, the Securities and Exchange Commission (SEC) held a roundtable discussion regarding cybersecurity and related issues, challenges it raises for market participants and public firms, and how to address those issues and challenges (SEC, 2014). Also, the Standing Advisory Group of the Public Company Accounting Oversight Board (PCAOB) assembled a panel discussion on cybersecurity issues and potential implications for financial reporting and auditing (PCAOB, 2014).

While there is still no formal disclosure requirement by the SEC or the PCAOB regarding cybersecurity, the issuance of Guidance on Disclosing Cybersecurity Risks by the SEC's Division of Corporation Finance demonstrates that regulators are concerned about the impact of cybersecurity on firms and investors (SEC, 2011). The speech by the SEC commissioner, Luis Aguilar, at the New York Stock Exchange reveals such concern:

[1] In later 2013, hackers gained access to millions of people's credit card data and personal information by exploring vulnerabilities in Target's Point of Sale (POS) systems. See at http://www.wsj.com/articles/SB10001424052702303754404579312232546392464.

[2] On November 24, 2014, a hacker group released confidential data from Sony Pictures Entertainment that include personal information about employees and their families, e-mails between employees, information about executive salaries at the company, copies of then-unreleased Sony films, and other information. See at https://en.wikipedia.org/wiki/Sony_Pictures_Entertainment_hack.

"… The impact of cyber-attacks may extend far beyond the direct costs associated with the immediate response to an attack. Beyond the unacceptable damage to consumers, these secondary effects include reputational harm that significantly affects a company's bottom line" (Aguilar, 2014).

Abundant literature demonstrates the negative impact of cyber incidents on breached firms' stock prices and various contingency factors that may mitigate or deepen the market reaction (Campbell, Gordon, Loeb, & Zhou, 2003; Cavusoglu, Mishra, & Raghunathan, 2004; Ettredge & Richardson, 2003; Gatzlaff & McCullough, 2010; Goel & Shawky, 2009; Gordon, Loeb, & Zhou, 2011; Hinz, Nofer, Schiereck, & Trillig, 2015; Yayla & Hu, 2011). Prior studies also show the role of board members, top executives, and internal auditors in addressing cybersecurity risks and cyber incidents (Higgs et al., 2014; Kwon et al., 2013; Steinbart et al., 2016; Steinbart et al., 2013; Zafar et al., 2015).

Academic research, however, remains silent on whether external auditors respond to cybersecurity incidents experienced by their clients, and whether they consider cybersecurity risks prior to the materialization of the risk. This gap is surprising given the increased attention from regulators on cybersecurity. In 2014, the Center for Audit Quality (CAQ) issued an alert regarding cybersecurity to summarize the responsibilities of independent external auditors with respect to cybersecurity matters (CAQ, 2014). For example, it suggests that the auditor should be responsible for evaluating the firm's accounting for cybersecurity-related losses, for assessing the impact on the firm's financial statements and disclosures, and for examining the firm's controls related to timely recording and disclosing the necessary information in the financial statements. Recent staff inspection reports also indicate that the inspections staff of PCAOB is examining how engagement teams evaluate the risks of material misstatement and related controls associated with cybersecurity and will continue to monitor auditors' practices regarding cybersecurity (PCAOB, 2015; PCAOB, 2016). Furthermore, the SEC has issued comment letters to encourage and request more disclosures on cyber incidents and has recently engaged in multiple active enforcement investigations involving data breach events concerning two aspects: disclosures and controls (Schubert, Cedarbaum, & Schloss, 2015). Some have argued that the SEC's cybersecurity disclosure guidance on cybersecurity will become a requirement and could be interpreted as an expansion of the scope of the integrated audit of internal control over financial reporting and the financial statements (Grant & Grant, 2014).

However, counter-arguments point out that despite regulators' concern about cybersecurity risks, there is no mandatory regulatory requirement for auditors to address cybersecurity risks. In the absence of such requirements, auditors would be averse to addressing cybersecurity risks beyond those affecting financial statements as doing so could needlessly expose them to liability and costs that would be difficult to recover. Also, the negative effect of cyber incidents on financial statements taken as a whole is sometimes quantitatively immaterial. For example, in the well-known Home Depot breach incident, the pretax net expense relating to the cyber incident was $119 million for the first three quarters of 2015, which is less than 1% of earnings before taxes.[3] Moreover, some believe that all firms operating in cyberspace will suffer a security event or breach at some point

[3] See http://www.auditanalytics.com/blog/when-is-a-cybersecurity-incident-material/. That said, it is important to recognize that cyber incidents can result in consequences such as reputational damage, loss of intellectual property, disruption of key business operations, fines and penalties assessed by governments litigation and remediation costs and exclusion from strategic markets that could be qualitatively material (AICPA, 2016).

in time[4], and that investors anticipate and price protect themselves against such risks, particularly if other firms that they monitor or pay attention to have experienced a cyber incident (Ettredge & Richardson, 2003). To sum up, it is an empirical question whether external auditors respond to cyber incidents in practice by noticeably extending their audit procedures and charging and successfully collecting higher fees for doing so. Because audit fees must be approved by the board of the client, external auditors must have a strong basis to justify the additional work performed pertaining to cyber incidents.

The main objective of this study is to investigate whether external auditors respond to cyber incidents by expanding their audit effort, resulting in higher audit fees, and whether external auditors are pricing material cybersecurity risks even before the actual adverse event happens. Using a change model specification, I find a significant positive relationship between increases in audit fees and cyber incidents. Furthermore, using firm's cybersecurity risk disclosure as the proxy for ex-ante material cybersecurity risk, it is shown that following cyber incidents, increases in audit fees are smaller for those with prior cybersecurity risk disclosure, implying that auditors price material cybersecurity risk prior to the cyber-attacks and thus are responding less severely (are less surprised) when the actual event happens. In addition, compared with firms that experience a cyber incident for the first time, firms with repeated cyber incidents are punished more severely by auditors as reflected in audit fees. Further, auditors increase audit fees most to respond to cyber incidents that involve intellectual property, the type of cyber incidents that threaten

[4] ASEC Cybersecurity Working Group Initiative; see at http://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/Pages/AICPACybersecurityInitiative.aspx

firm’s core value. Finally, external monitoring, as measured by the percentage of institutional holdings and number of block holders, can mitigate auditor’s concern over cyber incidents.

Overall, the findings of this study provide several contributions to the existing literature. First, I fill the gap in the prior literature by establishing the association between external audit activity and cyber incidents, suggesting that regulators’ concerns about cybersecurity issues are shared by external auditors. As regulators keep emphasizing that the impact of cyber incidents may go beyond the initial costs addressing the issues and can have further implications for financial reporting, my evidence that auditors are expanding their procedures following the incident provides some relief to the regulators and investors as auditors provide additional assurance for the quality of financial statements and internal controls.

Second, the finding that auditors are taking material cybersecurity risks into consideration before the actual cyber event happens indicates that they are proactively considering operational risks. Lawrence, Minutti-Meza, and Vyas (2016) point out that operational control risks can be indicative of financial control risks and urge stakeholders to consider operational control risks. While the question whether auditors price material cybersecurity risks to cover additional work or just price protect them against the risks is still not addressed, the fact that they are taking material cybersecurity risks into consideration is consistent with the emphasis on operational risks.

Third, the results suggest that auditors are not simply reacting to cyber incidents due to public pressure. Instead, they are most concerned about cyber incidents involving intellectual property, a type of incident that has the least exposure to the public compared with hacking of customer personal information and credit card. The evidence indicates that auditors are, at least in part, rational in evaluating cyber incidents, rather than just protecting themselves from public criticism.

Fourth, I extend research in the IT domain, particularly research on the consequences of cyber incidents. Prior research exclusively focuses on market reaction and firm performance after cyber incidents. This chapter empirically shows another consequence: increased audit fees. The finding should alert both practitioners and researchers that the impact of cyber incidents could be far more than anticipated and could concern various types of stakeholders.

Finally, I contribute to the audit fees literature by showing an additional factor that is valued by external auditors when setting audit fees. The magnitude of impact is larger than the impact of merger activities and more than half of the impact of material weakness in internal controls on audit fees, providing economic significance. The finding in this chapter suggests that future audit fees model may need to consider operational risk that is overlooked in prior audit literature.

From a practical point of view, this study provides evidence that may potentially alleviate regulator’s concerns about the aftermath of cyber incidents by suggesting that external auditors address such incidents even in the absence of regulatory requirements to do so. I argue that regulators carefully consider the status quo before introducing potential legislative rules for auditors on cybersecurity, as it appears in my study that auditors are reacting rationally based on the nature of the cyber incidents.

The remainder of this chapter proceeds as follows. The next section presents research background and introduces hypotheses. The third section addresses research design and sample selection procedure. The fourth section discusses results and describes additional tests. The last section concludes this paper.

2.2 BACKGROUND AND HYPOTHESIS DEVELOPMENT

Cybersecurity

Cybersecurity and information security are often used interchangeably.[5] The Cybersecurity Working Group of the AICPA Assurance Services Executive Committee defines cybersecurity as "the process of implementing and operating controls and other risk management activities to protect information and systems from security events that could compromise them and, when security events are not prevented, to detect, respond to, mitigate against, and recover from those events in a timely manner." The committee further defines cybersecurity compromise as "a loss of confidentiality, integrity, or availability of information, including any resultant impairment of (1) processing integrity or availability of systems or (2) the integrity or availability of system inputs or outputs, which have a negative effect on the achievement of the entity's business objectives and commitments (including cybersecurity commitments), as well as the laws and regulations related to cybersecurity risks and the cybersecurity program." The underlying premise is that "*all firms that operate in cyberspace will suffer a security event or breach at some point in time*." The assumption is supported by Ransbotham and Mitra (2009), who provide empirical evidence that all systems are potential victims of cyber-attacks. Firms not intrinsically attractive to attackers are not immune from attacks. For this study, cyber incidents are defined as "*cyber-attacks that are initiated by hackers to steal, tamper with,*

[5] Cybersecurity and information security are different in the sense that cybersecurity pertains to security risks related to cyberattacks while information security considers security of information and information systems regardless of the realm.

or destroy sensitive information in the cyber realm." Therefore, I exclude data breaches that are not related to cybersecurity, such as stolen laptop.

Although cybersecurity issues have been examined by multiple disciplines, there are two dominant streams of research. The first one is cybersecurity governance. Cybersecurity was traditionally viewed as purely a technical issue that should be handled by the IT department. Both practitioners and researchers have recently realized that cybersecurity should be considered from a managerial perspective and addressed at the highest level of the firm (ISACA, 2006; PricewaterhouseCoopers, 2016; Soomro, Shah, & Ahmed, 2016; Von Solms, 2005).[6] It has been shown that management has a critical role in encouraging cybersecurity policy compliance (Bulgurcu, Cavusoglu, & Benbasat, 2010; Hu, Dinev, Hart, & Cooke, 2012; Ifinedo, 2014). More recent literature focuses on specific roles. For instance, Kwon et al. (2013) find that putting IT executives in the top management team is negatively associated with the possibility of future cyber incidents, while Zafar et al. (2015) report that firms that have the CIO (or other top IT executive) in the top management team can recover damages or losses from cyber incidents quicker than the firms that do not. Because effective governance requires both monitoring and audit of performance, the internal audit function is also examined in relation to cybersecurity. Ideally, the feedback provided by internal audit can be used to improve the overall effectiveness of the firm's information security (Steinbart, Raschke, Gal, & Dilla, 2012). By conducting a series of semi-structured interviews with both internal auditors and information systems professionals, Steinbart et al. (2012) propose that internal auditors' IT

[6] A recent senate bill under review is suggesting that board members should have mandatory cybersecurity education. See at http://www.dandodiary.com/2016/01/articles/cyber-liability/senate-bill-would-require-disclosure-concerning-corporate-boards-cybersecurity-expertise/.

knowledge, communication skills, and attitude, as well as top management support, can influence the cooperation between internal audit and the information security function. Further studies by Steinbart et al. (2013) and Steinbart et al. (2016) substantiate the claims that a better relationship between the two functions is associated with fewer information security-related internal control weaknesses being reported to the board, more attacks stopped before they cause harm, and more attacks detected after they cause harm.

The second research stream concentrates on the consequences of cybersecurity breaches and cybersecurity-related events. Overall, there is much evidence that breached firms experience a negative market reaction (Campbell et al., 2003; Gatzlaff & McCullough, 2010; Goel & Shawky, 2009; Hinz et al., 2015), but there is no consensus on which types of the breaches (confidentiality, availability, and integrity) drive the decline in market value (Benaroch, Chernobai, & Goldstein, 2012; Goldstein, Chernobai, & Benaroch, 2011; Gordon et al., 2011). Furthermore, several studies report an array of contingency factors that influence the market response, including firm size, industries, and announcement texts (Acquisti, Friedman, & Telang, 2006; Das, Mukhopadhyay, & Anand, 2012; Wang, Ulmer, & Kannan, 2013; Yayla & Hu, 2011). In addition to the decline in market value, prior research finds that breaches caused by cyber-attacks are much more likely than breaches caused by lost or stolen hardware to be settled (Romanosky, Hoffman, & Acquisti, 2014), and that customers' overall satisfaction and revisit intentions are negatively affected by cybersecurity breaches (Berezina, Cobanoglu, Miller, & Kwansa, 2012). While cyber incidents are shown to be negative, previous literature also documents that information security investment (Chai, Kim, & Rao, 2011) and voluntary disclosure of information regarding cybersecurity (Gordon, Loeb, & Sohail, 2010; Wang, Kannan, &

Ulmer, 2013) can generate a positive market response. I extend this stream of literature to demonstrate that cyber incidents could also increase audit risks that are reflected in audit fees.

Cybersecurity and Audit Fees

I make two arguments about why external auditors should be concerned about cyber incidents: Internal Control over Financial Reporting (ICFR) and material misstatement.

Internal Control over Financial Reporting (ICFR)

ICFR is "a process designed by, or under the supervision of, the firm's principal executive and principal financial officers, or persons performing similar functions, and effected by the firm's board of directors, management, and other personnel, to provide reasonable assurance regarding the reliability of financial reporting and the preparation of financial statements for external purposes in accordance with generally accepted accounting principles" (PCAOB, 2004). ICFR also includes procedures and policies related to maintaining accounting records, documenting transactions, authorizing receipts and expenditures, and safeguarding assets (Hogan & Wilkins, 2008). Sarbanes-Oxley Act (SOX) section 404 requires management to assess and report on the effectiveness of their firms' ICFR. It also requires external auditors to attest and report on the assessments made by client management. Hence, external auditors are legally responsible for detecting deficiencies in firms' ICFR. Prior research documents that external auditors charge higher fees for clients with deficiencies in ICFR (Hoitash, Hoitash, & Bedard, 2008), and the fee premium persists several years after the deficiencies are fixed (Hoag & Hollingsworth, 2011; Munsif, Raghunandan, Rama, & Singhvi, 2011).

In the event of a cyber incident, external auditors are expected to consider its implications for ICFR. If the attack is directly on a firm’s accounting systems, the incident could involve, or could suggest the risk of, manipulation of the firm’s books and records, which could affect financial statements (PCAOB, 2014). Prior research posits that the negative market response following a cyber incident announcement is because such an event signals the presence of internal control material weaknesses (Benaroch et al., 2012). Likewise, the PCAOB’s staff inspection briefs indicate that inspection staffs are “reviewing how engagement teams evaluate the risks of material misstatement associated with cyber-security and the related controls in the integrated audit” (PCAOB, 2015) and cautioning external auditors to consider the implications for ICFR if cybersecurity incidents have occurred during the audit period (PCAOB, 2016). The SEC is also pursuing firms based on perceived shortcomings of their ICFR after cyber incidents to the extent that unauthorized persons are able to access, steal, or destroy material assets in their information systems (Association of Corporate Counsel, 2016).

Even if cyber-attacks have no direct impact on a firm’s accounting systems, external auditors may still need to exert additional efforts. Cyber-attacks on perimeter or internal network layers may indicate weaknesses in general IT controls, which could suggest risks in ICFR. Prior study observes a positive association between operational control weaknesses and material weakness in ICFR, suggesting that vulnerabilities in any of the systems and procedures could affect both operating and financial reporting activities (Lawrence et al., 2016). For instance, a report by Verizon (2016) demonstrates that older vulnerabilities are highly targeted and many breaches are permitted by known bugs or vulnerabilities. If a firm fails to remediate vulnerabilities in one particular area that eventually leads to a cyber incident, it is unlikely that the firm will be proactive in preventing vulnerabilities in other systems.[7] In the Target data breach case, a Senate report notes that the attackers who infiltrated Target's network with a vendor's credentials seemed to succeed in moving from less sensitive areas of Target's network to areas storing consumer data, suggesting that the firm failed to isolate its most sensitive network assets. As it appears that the attackers succeeded in moving through various key Target systems (United States Senate, 2013), legitimate concerns should be raised that attackers may be capable of exploring corporate networks in depth and attacking different layers of systems including Enterprise Resource Planning (ERP) systems and general ledger.

Given the central functionalities of a firm's accounting information systems and the wealth of data stored on those systems are likely to be of great interest to cybercriminals, external auditors should consider the potential risks that come from cybersecurity threats (Debreceny, 2014). Since external auditors respond to the higher levels of control risk by charging higher audit fees (Hoag & Hollingsworth, 2011; Hogan & Wilkins, 2008; Hoitash et al., 2008), I expect external auditors to charge higher fees after a cyber incident and expand their security-related ICFR audit procedures.

Material Misstatement

Cyber incidents may also be associated with the risks of material misstatement. The occurrence of cyber incidents could increase client business risk, which refers to "the risk that the client's economic condition will deteriorate in either the short term or long term" (Johnstone, 2000). Prior studies indicate that external auditors evaluate client business risk when determining whether to accept a new client (Khalil & Mazboudi, 2016), and are less

[7] According to Data Breach Litigation Report (2016), negligence is the most widely used legal theory against breached firms.

likely to accept a client's proposed accounting practice if client business risk is high (Chang & Hwang, 2003). A recent analysis reveals that following a cyber incident, firms, on average, experience more than 3.3% abnormal churn of existing customers, which is defined as a greater than expected loss of customers in the normal course of business (Ponemon Institute, 2016).[8] This is consistent with a behavioral study by Berezina et al. (2012) that shows participants' overall satisfaction, revisit intentions, and the likelihood of recommending a hotel to others were negatively affected by a cyber breach. The Ponemon study also indicates that indirect costs associated with cyber incidents (primarily lost business) are much larger than (almost twice) the direct costs such as costs to resolve the data, investments in technologies, or legal fees. Therefore, although the direct costs of cyber incident may not be material, the resulting indirect costs could be material enough to provide management incentives to bias the report.[9] As client's business risk is an important determinant of whether financial statements contain material misstatements (AICPA, 1997), external auditors may conduct more costly audit procedures to achieve an acceptable level of audit risk and may charge a fee premium if the additional effort is not sufficient to cover residual costs under heightened client business risk (Stanley, 2011).

In addition, cyber-attacks may have an indirect effect on financial statements by requiring the future recognition of asset impairments and loss contingencies, and may push a firm to reconsider projections. In auditing accounting estimates, external auditors normally should consider the firm's historical experience in making past estimates as well as their experience of other firms in the same industry. However, changes in facts,

[8] The report has controlled for outliers by considering only breaches that affect less than 100,000 records.

[9] The bias could be either downward or upward. For example, management could also use cybersecurity breach to explain bad firm performance and take a big bath.

circumstances, or a firm’s procedures may cause the firm and auditors to take into account different factors that were not considered in the past, but become significant to the accounting estimate (AU sec. 342). When planning and performing procedures to evaluate the reasonableness of the firm’s accounting estimates, the auditors should consider, with an attitude of professional skepticism, subjective and objective factors included in the estimate. If a cyber incident happens, the auditors may need to collect additional evidence regarding whether there would be a significant change in circumstances. For example, external auditors need to examine whether there is a substantial increase in returns that would affect the sales returns estimate, which could influence accounting numbers on financial statements materially. Another example is the impact on estimated goodwill impairment if expected future cash flows for a cash-generating unit are affected by a cyber incident. This is consistent with the SEC’s Disclosure Guidance, which recommends that subsequent to a security incident firms should reassess the assumptions that underlie the estimates made in preparing the financial statements and must explain any risk or uncertainty of a reasonably possible change in its estimates in the near-term that would be material to the financial statements (SEC, 2011). According to the guidance, cyber incidents may result in diminished future cash flows, thereby requiring consideration of impairment of certain assets including goodwill, customer-related intangible assets, trademarks, patents, capitalized software or other long-lived assets associated with hardware or software, and inventory.

In the event of a cyber incident, external auditors should also assess the risk of material misstatement that comes from the evaluation of the firm’s accounting for known cybersecurity-related losses that include contingent liabilities and claims (CAQ, 2014). An estimated loss from a loss contingency would be accrued by a charge to income if both of the following conditions are met: information available prior to issuance of the financial statements indicates that it is probable that an asset had been impaired or a liability had been incurred at the date of the financial statements, and the amount of loss can be reasonably estimated (FASB, 1975). In addition, the auditor should obtain evidential matter relevant to (1) the existence of a condition, situation, or set of circumstances indicating an uncertainty as to the possible loss to an entity arising from litigation, claims, and assessments, (2) the period in which the underlying cause for legal action occurred, (3) the degree of probability of an unfavorable outcome, and (4) the amount or range of potential loss (AU sec. 337). Specific to cybersecurity, approximately 5% of publicly reported data breaches led to class action litigation, and the conversion rate has remained relatively consistent over the years (Bryan Cave, 2016). If a firm had a material contingent liability for an actual cyber incident, in addition to performing audit procedures related to the reasonableness of the liability recorded, the auditor would also assess whether the disclosures in footnotes related to such liability are appropriate as they relate to the financial statements taken as a whole (CAQ, 2014). Because facts and impacts about cyber incidents may not be fully revealed until further investigation, auditors may need to exert additional effort to reduce the uncertainty of contingencies and claims.

It is arguable that in some cases, the impact of cyber incidents on financial statements may not be material quantitatively, and thus should not attract the auditor’s attention. However, as the SEC repeatedly forced several firms to disclose their cyber incidents even if the impact is immaterial to financial statement (e.g. Amazon was asked by the SEC to disclose the cyber-raid in its next quarterly filing in 2012 despite Amazon’s claims that the cyber-attack was not important), it is reasonable to expect that external auditors will increase professional skepticism with respect to firms' cyber incidents even if the impacts may not directly influence financial statements in a quantitatively material manner. Therefore, this study introduces the following hypothesis.

H1. Ceteris paribus, increases in audit fees are larger for firms that experienced cyber incidents than firms that did not experience cyber incidents.

The next hypothesis concentrates on the association between audit fees and ex-ante cybersecurity risk. While the above discussion argues that auditors will increase audit fees after the occurrence of cyber incidents as a responding strategy, it remains unexamined whether external auditors price material cybersecurity risks before the actual incident happens. Stanley (2011) find that external auditors price any expected cost arising from potential losses such as future litigation or reputational damage. As cybersecurity risk has implications for firm's future performance, customer relationship, and control environment, I would expect that external auditors incorporate material cybersecurity risk into audit fees even before the actual risk event happens.

It is not trivial to determine when ex-ante cybersecurity risk is becoming material as auditors are not required to audit and attest on firm's cybersecurity. To address this issue, firm's cybersecurity-related risk factor disclosure is used as the proxy for material cybersecurity risk. Because cybersecurity risk disclosure is negative information and is not mandatory, firms may have incentives to withhold the disclosure due to concerns over increased cost of capital or damaged future career (Kothari, Li, & Short, 2009; Kothari, Shu, & Wysocki, 2009). However, litigation costs could be high enough to motivate risk disclosures (Skinner, 1994). Managers could be sued or face legal liability if they fail to disclose a material risk (Campbell, Chen, Dhaliwal, Lu, & Steele., 2014). Consistent with the view, prior studies have shown that firms are not making boilerplate risk factor disclosures (Campbell et al., 2014; Filzen, 2015; Gaulin, 2017; Hope, Hu, & Lu, 2016; Kravet & Muslu, 2013). Therefore, I expect that firms are likely to make cybersecurity risk disclosure when cybersecurity risk is material. Since risk disclosure in 10-K (i.e., Item 1A - Risk Factors) is reviewed by external auditors, it is natural that the auditors should be aware of material cybersecurity risk. Considering that material cybersecurity risk may have an impact on firm's performance and controls and eventually could influence accounting numbers on financial statements materially, auditors may take material cybersecurity risk into account when they determine audit fees. If auditors incorporate material cybersecurity risk before a cyber incident happens, it is expected that external auditors respond to the cyber incident less severely (increase smaller audit fees) when there is prior disclosure of cybersecurity risk by the firm. On the other hand, if auditors do not price cybersecurity risk prior to a cyber incident, the reaction to the cyber incident should be unconditional on firm's prior cybersecurity risk disclosure. This leads to the following hypothesis.

H2. Ceteris paribus, increases in audit fees should be smaller for cybersecurity breached firms with prior cybersecurity risk disclosure than for cybersecurity breached firms without prior cybersecurity risk disclosure.

Note that while I assume that firms that have cybersecurity risk disclosures are facing material cybersecurity risk, the opposite may not be true. It is still possible that firms withhold disclosure regarding cybersecurity even if they have material cybersecurity risk. However, this is not a significant concern for my test as it will only bias against me finding any significant results if auditors are incorporating material cybersecurity risk that firms did not disclose.

2.3 RESEARCH DESIGN AND SAMPLE SELECTION

Estimation Model

To mitigate the concern of endogeneity that firms with higher audit fees may be more likely to be targeted by hackers, I use a change specification to examine the association between cyber incident and audit fees. I choose audit fee change model over two-stage model because Lennox, Francis, and Wang (2011) indicate that two-stage model is fragile and can generate almost any possible outcome by making minor changes in model specification. Propensity score matching is not selected because it can only control for endogeneity that arises from observable rather than unobservable factors (Lennox et al., 2011; Shipman, Swanquist, & Whited, 2017), which could be a significant problem in my research context given the fact that there is no well-specified model to evaluate the determinants of experiencing cyber incidents. As audit fee change model can eliminate endogeneity caused by unobservable factors under the assumption that these factors are time-invariant, it has been commonly used in recent audit fee literature (Desir, Casterella, & Kokina, 2013; Hardies, Breesch, & Branson, 2015; Khalil & Mazboudi, 2016; Stanley, 2011).

I estimate the change form of a traditional audit fees model that is adapted from prior studies (Doogar, Sivadasan, & Solomon, 2015; Elliott, Ghosh, & Peltier, 2013; Huang, Raghunandan, & Rama, 2009; Stanley, 2011).

$$\begin{aligned}\Delta logAUDIT_{it} = {} & \Delta Cyber\text{-}Incident_{it} + \Delta LNassets_{it} + \Delta InvRec_{it} + \Delta Segments_{it} + \Delta Foreign_{it} \\ & + \Delta Merger_{it} + \Delta Special_{it} + \Delta Loss_{it} + \Delta Growth_{it} + \Delta Btm_{it} + \Delta Big4_{it} \\ & + \Delta GCO_{it} + \Delta Initial_{it} + \Delta ROA_{it} + \Delta Leverage_{it} + \Delta Quick_{it} + \Delta ICW_{it}\end{aligned}$$

$$+ \mathit{Residual}_{it\text{-}1} + \mathit{Year\ Indicators} + \mathit{Industry\ Indicators} + \varepsilon_{it} \qquad (1)$$

where Δ represents one-year change in the level of each variable, and $\mathit{Residual}_{it\text{-}1}$ represents the prior-period unexpected audit fees measured as the residual from yearly estimations of the basic audit fees model (2) to control for the effect of mispricing and mean reversion over time (Francis & Wang, 2005; Mayhew, 2005; Stanley, 2011). Appendix A contains a detailed description of variable definitions.

$$\begin{aligned} \mathit{logAUDIT}_{it} = {} & \mathit{Cyber\text{-}Incident}_{it} + \mathit{LNassets}_{it} + \mathit{InvRec}_{it} + \mathit{Segments}_{it} + \mathit{Foreign}_{it} + \mathit{Merger}_{it} \\ & + \mathit{Special}_{it} + \mathit{Loss}_{it} + \mathit{Growth}_{it} + \mathit{Btm}_{it} + \mathit{Big4}_{it} + \mathit{GCO}_{it} + \mathit{Initial}_{it} \\ & + \mathit{ROA}_{it} + \mathit{Leverage}_{it} + \mathit{Quick}_{it} + \mathit{ICW}_{it} + \mathit{Busy}_{it} + \mathit{Year\ Indicators} \\ & + \mathit{Industry\ Indicators} + \varepsilon_{it} \qquad (2) \end{aligned}$$

The focus of this study is on the relationship between $\Delta \mathit{logAUDIT}_{it}$ and $\Delta \mathit{Cyber\text{-}Incident}_{it}$. A positive coefficient on $\Delta \mathit{Cyber\text{-}Incident}_{it}$ will support my hypothesis that external auditors increase audit fees in the fiscal year of a cyber incident. For control variables, I expect a positive coefficient on $\Delta \mathit{LNassets}_{it}$, as firm size is the primary driver of audit fees. $\Delta \mathit{InvRec}_{it}$, $\Delta \mathit{Segments}_{it}$, $\Delta \mathit{Foreign}_{it}$, $\Delta \mathit{Merger}_{it}$, and $\Delta \mathit{Special}_{it}$ are included to control for the complexity of the audit and anticipated positive coefficients. $\Delta \mathit{Big4}_{it}$ is included and expected to be positive as it accounts for fee premium. $\Delta \mathit{Loss}_{it}$, $\Delta \mathit{GCO}_{it}$, $\Delta \mathit{Leverage}_{it}$, and $\Delta \mathit{ICW}_{it}$ control for higher audit fees charged to riskier firms. Coefficients on $\Delta \mathit{Growth}_{it}$, $\Delta \mathit{Btm}_{it}$, $\Delta \mathit{ROA}_{it}$, and $\Delta \mathit{Quick}_{it}$ are anticipated to be negative because such firms pose fewer risks to the audit. Finally, $\Delta \mathit{Initial}_{it}$ is added to control for the lower fees due to lowballing in initial engagement.

To examine the second hypothesis, I create an indicator variable *Disclosure* that takes the value of 1 if a firm has prior-year cybersecurity risk disclosure in the risk factor disclosure section (i.e., Item 1A in 10-K), 0 otherwise. Cybersecurity risk disclosure is identified by searching keywords that are developed based on prior research (Gordon et al., 2010; Wang, Kannan, et al., 2013). Appendix B provides a list of keywords used in this study. A firm with risk factor disclosure that contains any of these keywords is considered to have cybersecurity risk disclosure. The interaction term Δ*Cyber-Incident*$_{it}$* *Disclosure* is added into equation (1). A negative coefficient would suggest that auditors increase fewer fees for the firms that have prior cybersecurity risk disclosures.

Sample Selection

I obtain my cyber incident data from the Audit Analytics cybersecurity database and Privacy Rights Clearinghouse (privacyrights.org). Audit Analytics cybersecurity database collects cybersecurity breaches for U.S. public firms while Privacy Rights Clearinghouse publishes data breaches that involve individual's identity. I start with 738 data breaches, of which 303 are related to cyber incidents (cyber-attacks)[10]. I first remove cyber incidents for firms in the financial industry (SIC 6000-6999) as they have a different audit fee structure. If a firm experienced more than one cyber-attack in one year (e.g. Hyatt Hotels Corp. was hacked twice in 2015), I keep only one incident per year to prevent over-sampling. Finally, observations that do not have the necessary financial or audit data are excluded. These procedures result in a final sample of 140 cybersecurity breached firm observations. Any firm-year that is not in my initial sample of cyber incidents is considered to be a non-cybersecurity breached observation (*Cyber-Incident*=0). My final sample consists of 140 cybersecurity breached observations and 29,627 non-cybersecurity

[10] Data breach could happen due to reasons other than cyber-attacks. For example, stolen laptop or improperly disposed documents could result in breach of sensitive information. These types of data breaches are not considered as they are not related to cybersecurity. In addition, column 2 of Table 5 also indicates that external auditors are not concerned about such type of data breaches.

breached firm observations. Table 1 summarizes the sample selection procedure.

Table 1. Sample Selection Criteria

Number of firm-years with cyber incidents		140
Original Number of cyber incidents	303	
Minus: observations that are in financial industries	(24)	
Minus: observations that have more than one cyber incident in a year (keep each firm-year only once)	(-76)	
Minus: observations that have missing data for the analysis	(-68)	
Number of firm-years without cyber incidents (i.e., control groups)		29,627
Total number of observations		29,767

There are two potential limitations that might affect my data set. The first one is that a firm experienced a cyber incident but never discovered the attack. The second scenario is that a firm recognized that it was hacked and notified its external auditor, but the incident was not publicly announced, thus not recorded in my sample. Under both situations, I may incorrectly classify a cybersecurity breached firm as a non-breached firm. However, the validity of my results should not be affected by these possibilities because they will only act as a bias against me, thus weaken my findings.

Table 2. Descriptive Statistics

Variable	Variables in the Original Form									Variables in the Change Form		
	Total Sample			*Firms with Cyber Incidents*			*Firms without Cyber Incidents*					
	Mean	*Std*	*Median*	*Mean*	*Std*	*Median*	*Mean*	*Std*	*Median*	*Mean*	*Std*	*Median*
logAUDIT	13.6863	1.3001	13.7280	15.5013	1.1434	15.5454	13.6785	1.2953	13.7231	0.0206	0.2588	0.0131
Lnassets	6.1726	2.2492	6.1912	9.2160	1.7234	9.4533	6.1595	2.2423	6.1823	0.0518	0.2685	0.0375
InvRec	0.2371	0.1788	0.2039	0.1746	0.1398	0.1271	0.2374	0.1789	0.2044	0.0002	0.0540	0.0002
Segments	1.9346	1.2642	1.0000	2.9766	1.8675	3.0000	1.9301	1.2591	1.0000	0.0100	0.3701	0.0000
Foreign	0.3517	0.4775	0.0000	0.4531	0.4998	0.0000	0.3513	0.4774	0.0000	0.0113	0.2145	0.0000
Merger	0.1966	0.3975	0.0000	0.3750	0.4860	0.0000	0.1959	0.3969	0.0000	0.0310	0.3812	0.0000
Special	0.6772	0.4676	1.0000	0.8359	0.3718	1.0000	0.6765	0.4678	1.0000	0.0204	0.5220	0.0000
Loss	0.3543	0.4783	0.0000	0.1719	0.3788	0.0000	0.3551	0.4785	0.0000	0.0131	0.4286	0.0000
Growth	0.1434	0.5980	0.0609	0.0563	0.1463	0.0442	0.1438	0.5992	0.0610	-0.0573	0.6206	-0.0242
Btm	0.5487	0.9372	0.4593	0.4519	0.4129	0.3698	0.5491	0.9388	0.4597	0.0128	0.5769	0.0015
Big4	0.7099	0.4538	1.0000	0.9531	0.2122	1.0000	0.7088	0.4543	1.0000	-0.0080	0.1321	0.0000
GCO	0.0630	0.2429	0.0000	0.0234	0.1519	0.0000	0.0631	0.2432	0.0000	0.0092	0.1870	0.0000
Initial	0.0587	0.2350	0.0000	0.0000	0.0000	0.0000	0.0589	0.2355	0.0000	-0.0002	0.3169	0.0000
ROA	-0.0136	0.2954	0.0603	0.0941	0.03	0.0831	-0.0141	0.2959	0.0603	-0.0074	0.1330	-0.0006
Leverage	0.5402	0.3783	0.4904	0.5995	0.2372	0.6033	0.5400	0.3788	0.4901	0.0173	0.1433	0.0025
Quick	2.2318	2.4193	1.4701	1.4781	1.0762	1.1526	2.2350	2.4230	1.4718	-0.0608	1.2509	-0.0098
ICW	0.0913	0.2881	0.0000	0.0313	0.1747	0.0000	0.0916	0.2885	0.0000	-0.0097	0.2990	0.0000
Busy	0.7401	0.4386	1.0000	0.5703	0.4970	1.0000	0.7408	0.4382	1.0000			
Disclosure	0.3835	0.4863	0.0000	0.8614	0.3473	1.0000	0.3812	0.4857	0.0000			

Note: All variables are winsorized at 1 and 99 percent. All Variables are defined in Appendix A.

Table 3. Correlations among Variables Included in the Audit Fees Model

Panel A: Variables in the Original Form

	logAUDIT	*Cyber-Incident*	*Lnassets*	*InvRec*	*Segments*	*Foreign*	*Merger*	*Special*	*Loss*	*Growth*	*Btm*	*Big4*	*Initial*	*GCO*	*ROA*	*Leverage*	*Quick*	*ICW*	*Busy*	*Disclosure*
logAUDIT	1.000																			
Cyber-Incient	**0.094**	1.000																		
Lnassets	**0.872**	**0.094**	1.000																	
InvRec	**-0.061**	**-0.028**	**-0.152**	1.000																
Segments	**0.413**	**0.048**	**0.398**	**0.034**	1.000															
Foreign	**0.251**	**0.013**	**0.133**	**0.103**	**0.092**	1.000														
Merger	**0.227**	**0.032**	**0.203**	**-0.032**	**0.098**	**0.107**	1.000													
Special	**0.316**	**0.019**	**0.236**	**-0.023**	**0.142**	**0.140**	**0.340**	1.000												
Loss	**-0.306**	**-0.031**	**-0.426**	**-0.082**	**-0.210**	**-0.050**	**-0.076**	**0.032**	1.000											
Growth	**-0.091**	-0.009	**-0.072**	**-0.083**	**-0.064**	**-0.032**	**0.027**	**-0.056**	**0.028**	1.000										
Btm	**-0.035**	-0.008	**0.035**	**0.076**	**0.041**	0.005	**-0.013**	-0.011	**-0.027**	**-0.053**	1.000									
Big4	**0.640**	**0.036**	**0.588**	**-0.125**	**0.182**	**0.115**	**0.113**	**0.174**	**-0.227**	**-0.048**	**-0.045**	1.000								
Initial	**-0.111**	-0.011	**-0.102**	**0.022**	**-0.039**	-0.005	**-0.025**	-0.003	**0.064**	**0.020**	0.005	**-0.144**	1.000							
GCO	**-0.273**	**-0.017**	**-0.345**	-0.005	**-0.113**	**-0.063**	**-0.089**	**-0.018**	**0.284**	**0.020**	**-0.212**	**-0.225**	**0.046**	1.000						
ROA	**0.370**	**0.027**	**0.502**	**0.118**	**0.203**	**0.092**	**0.111**	**0.060**	**-0.537**	**-0.065**	**0.153**	**0.254**	**-0.051**	**-0.501**	1.000					
Leverage	**0.021**	0.008	**-0.032**	0.001	**0.019**	**-0.068**	**-0.025**	**0.093**	**0.187**	0.007	**-0.471**	**-0.030**	0.001	**0.395**	**-0.351**	1.000				
Quick	**-0.204**	**-0.019**	**-0.208**	**-0.210**	**-0.172**	-0.004	**-0.060**	**-0.126**	**0.084**	**0.073**	**0.062**	**-0.055**	0.008	**-0.117**	**-0.060**	**-0.396**	1.000			
ICW	**-0.148**	-0.011	**-0.206**	**0.041**	**-0.054**	**-0.017**	**-0.034**	0.001	**0.149**	**0.028**	-0.010	**-0.180**	**0.076**	**0.219**	**-0.162**	**0.127**	**-0.033**	1.000		
Busy	**0.029**	**-0.022**	**0.028**	**-0.187**	**0.014**	**-0.012**	0.005	0.006	**0.061**	**0.062**	**-0.028**	**0.040**	0.004	**0.029**	**-0.069**	**0.072**	**0.031**	-0.002	1.000	
Disclosure	**0.198**	**0.068**	**0.191**	**-0.061**	**0.039**	**0.028**	**0.161**	**0.090**	**-0.077**	**-0.031**	**-0.054**	**0.134**	**-0.016**	**-0.074**	**0.099**	**0.013**	**-0.079**	**-0.047**	**-0.015**	1.000

Note: This table presents correlations for all variables in the original form. Significant correlations are represented in bold (two-sided and threshold: .05). All Variables are defined in Appendix A.

Table 3. Correlations among Variables Included in Audit Fees Model (continued)

Panel B: Variables in the Change Form

	Δ*logAUDIT*	Δ*Cyber-Incident*	Δ*Lnassets*	Δ*InvRec*	Δ*Segments*	Δ*Foreign*	Δ*Merger*	Δ*Special*	Δ*Loss*	Δ*Growth*	Δ*Btm*	Δ*Big4*	Δ*Initial*	Δ*GCO*	Δ*ROA*	Δ*Leverage*	Δ*Quick*	Δ*ICW*
Δ*logAUDIT*	1.0000																	
Δ*Cyber-Incident*	0.0105	1.0000																
Δ*Lnassets*	**0.2771**	0.0031	1.0000															
Δ*InvRec*	**-0.0210**	-0.0034	**-0.2430**	1.0000														
Δ*Segments*	**0.0840**	**-0.0121**	**0.1084**	**0.0308**	1.0000													
Δ*Foreign*	0.0031	-0.0068	0.0081	0.0092	0.0026	1.0000												
Δ*Merger*	**0.0899**	0.0005	**0.1305**	**-0.0334**	**0.0461**	0.0095	1.0000											
Δ*Special*	**0.0593**	-0.0053	0.0114	-0.0063	**0.0321**	0.0054	**0.2139**	1.0000										
Δ*Loss*	**0.0470**	0.0019	**-0.1097**	**0.0122**	0.0068	-0.0031	**0.0214**	**0.0774**	1.0000									
Δ*Growth*	**0.0302**	0.0049	**0.1610**	**0.0839**	**0.0375**	0.0011	**0.0300**	-0.0034	**-0.1151**	1.0000								
Δ*Btm*	**0.0417**	-0.0008	**0.1220**	**-0.0540**	**0.0209**	**0.0207**	0.0085	**0.0312**	**0.0510**	-0.0076	1.0000							
Δ*Big4*	**0.2113**	0.0029	**0.0483**	**-0.0119**	-0.0061	-0.0062	0.0069	**0.0118**	**0.0126**	0.0015	**0.0132**	1.0000						
Δ*Initial*	**-0.1200**	0.0003	0.0055	-0.0050	0.0018	-0.0050	-0.0100	0.0032	-0.0054	**0.0163**	0.0015	**-0.1066**	1.0000					
Δ*GCO*	-0.0078	-0.0074	**-0.1537**	**0.0419**	**-0.0160**	-0.0075	-0.0052	0.0086	**0.0508**	**-0.0341**	**-0.0625**	**0.0123**	**0.0147**	1.0000				
Δ*ROA*	-0.0009	0.0034	**0.3722**	**-0.0260**	0.0002	0.0016	**-0.0141**	**-0.0448**	**-0.2618**	**0.2602**	**0.0284**	0.0036	0.0075	**-0.1761**	1.0000			
Δ*Leverage*	**0.0403**	-0.0043	**-0.2286**	**0.1784**	**0.0240**	-0.0103	**0.0254**	**0.0336**	**0.1678**	**-0.0267**	**-0.2533**	-0.0066	-0.0044	**0.1865**	**-0.3162**	1.0000		
Δ*Quick*	**-0.0495**	0.0027	**0.1628**	**-0.2200**	**-0.0450**	**-0.0121**	**-0.0591**	**-0.0280**	**-0.0795**	**-0.0334**	**0.0328**	0.0038	0.0053	**-0.1141**	**0.2136**	**-0.3368**	1.0000	
Δ*ICW*	**0.1212**	0.0071	**0.0190**	0.0044	0.0015	0.0107	0.0079	**0.0171**	**0.0282**	-0.0053	**-0.0119**	**0.0306**	0.0070	**0.0174**	**-0.0145**	**0.0393**	**-0.0227**	1.0000

Note: This table presents correlations for all variables in the change form. Significant correlations are represented in bold (two-sided and threshold: 0.05). All Variables are defined in Appendix A.

Table 2 reports the descriptive statistics for the variables used in the analysis. Firms with reported cyber incidents tend to be larger than their counterparts (9.2160 vs 6.1595, $p < 0.001$). In addition, about 86% of cybersecurity breached firms have prior cybersecurity risk disclosure, while only about 38% of non-breached firms have such disclosures. Table 3, panel A presents univariate correlations among the variables in equation (2) while Panel B reports univariate correlations among the change variables. The dependent variable, *logAUDIT*, is significantly correlated with all independent variables. The variable of interest, *Cyber-Incident*, is significantly correlated with the dependent variable and several independent variables, with the largest correlation being 0.094. In the correlation matrix of change variables, Δ*logAUDIT* is not significantly correlated with Δ*Cyber-Incident*, Δ*Foreign*, Δ*GCO*, and Δ*ROA*. Therefore, I turn to multiple regression to control for other determinants of Δ*logAUDIT*.

2.4 RESULTS

Main Findings

Table 4 shows the results of the multiple regression in equation (2). The traditional audit fee model is highly significant and captures about 84.65% of the variation in *logAUDIT* using my independent variables. The coefficient on *Cyber-Incident* is 0.216 ($p < 0.0001$), providing some initial support for my hypothesis. Except for *GCO*, *Leverage*, and *Busy*, all the control variables are significant in the predicted direction. Specifically, *LNassets*, *InvRec*, *Segments, Foreign*, *Merger*, *Special*, *Loss*, *Big4*, and *ICW* are positively associated with *logAUDIT*, while *Growth*, *Btm*, *Initial*, *ROA* and *Quick* are negatively correlated with *logAUDIT*.

Table 4. Regression of Cyber Incidents on Audit Fees using Equation (2)

Independent Variables	*Estimates*	*t-statistics*
Cyber-Incident	0.216	5.18***
Lnassets	0.495	95.90***
InvRec	0.463	10.04***
Segments	0.060	10.10***
Foreign	0.118	8.25***
Merger	0.049	4.33***
Special	0.150	15.25***
Loss	0.121	11.35***
Growth	-0.031	-6.65***
Btm	-0.066	-10.38***
Big4	0.395	21.28***
Initial	-0.089	-5.66***
GCO	-0.039	-1.69*
ROA	-0.286	-12.96***
Leverage	0.011	0.58
Quick	-0.015	-5.89***
ICW	0.163	10.18***
Busy	0.003	0.21
Intercept	10.229	113.44***
Industry Effects	Included	
Year Effects	Included	
Adjusted R square	84.65%	
# Observations	36,565	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Column 1 of Table 5 reports the results of the audit fee change model in equation (1). As expected, the explaining power of the change model is much smaller than that of the traditional audit fee model (adjusted *R* square = 24.98%), but is similar to those reported in prior studies (Desir et al., 2013; Hardies et al., 2015; Khalil & Mazboudi, 2016). My variable of interest, Δ*Cyber-Incident*, is positively associated with Δ*logAUDIT,* supporting my first hypothesis. The result is also economically significant. The increase in audit fees after cyber incident (0.045) is about twice the increase after firms suffer loss (0.024), and about 60% of the increase after firms report material weakness in internal controls (0.074). As for control variables, all except Δ*Foreign*, Δ*Growth*, Δ*Btm*, and Δ*GCO* are significant in the predicted direction.

While my focus is on cyber incidents that are initiated by malicious third parties and happen in the cyber realm (i.e. hacking), the regression results for data breaches that do not involve hacking is also reported in Column 2 of Table 5 as a comparison. The coefficient of Δ*Non_Cyber-Incident* (a binary variable that equals 1 if the firm suffers a data breach that does not involve cyber-attack, 0 otherwise) is not statistically significant, suggesting that external auditors are not concerned about data breaches that are less severe, such as stolen laptop or unintentional disclosure of sensitive information online. Overall, results in Table 5 support my hypothesis that external auditors are responding to cyber incident by charging higher audit fees.

Table 5. Regression of Cyber incidents on Audit Fees increases using Equation (1)

Independent Variables	*Cyber-Incident*		*Non_Cyber-Incident*	
	Estimates	*t-statistics*	*Estimates*	*t-statistics*
ΔCyber-Incident	0.045	2.86***		
ΔNon_Cyber-Incident			0.019	1.37
ΔLnassets	0.277	37.22***	0.276	37.06***
ΔInvRec	0.133	4.28***	0.131	4.23***
ΔSegments	0.025	5.85***	0.025	5.86***
ΔForeign	0.009	1.28	0.009	1.27
ΔMerger	0.029	7.70***	0.029	7.78***
ΔSpecial	0.025	9.37***	0.025	9.37***
ΔLoss	0.024	6.62***	0.024	6.59***
ΔGrowth	0.000	0.01	0.000	0.03
ΔBtm	-0.004	-1.09	-0.004	-1.11
ΔBig4	0.335	21.27***	0.334	21.26***
ΔInitial	-0.076	-12.06***	-0.076	-12.05***
ΔGCO	0.013	1.34	0.013	1.34
ΔROA	-0.127	-8.81***	-0.126	-8.77***
ΔLeverage	0.059	4.29***	0.059	4.25***
ΔQuick	-0.010	-6.67***	-0.010	-6.65***
ΔICW	0.074	11.32***	0.074	11.34***
Residual	-0.152	-41.31***	-0.151	-41.33***
Intercept	0.034	1.88*	0.035	1.90*
Industry Effects	Included		Included	
Year Effects	Included		Included	
Adjusted R square	24.98%		24.95%	
# Observations	29,767		29,725	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Regression results for testing whether external auditors price material cybersecurity risk prior to the cyber incident are presented in Table 6. Consistent with my hypothesis, there is a statistically significant and negative coefficient on Δ*Cyber-Incident* Disclosure*, indicating that increase in audit fees is smaller for those cybersecurity breached firms that have prior cybersecurity risk disclosures. On average, firms without prior cybersecurity risk disclosure are punished three times larger than those with prior cybersecurity risk disclosure (0.12 vs. 0.12-0.09). The results provide evidence that auditors indeed price cybersecurity risks even before the actual adverse event happens[11].

Sensitivity Analyses

Multiple Breaches for a Single Firm

Several firms experienced cyber incidents in multiple years, which could introduce over-sampling bias in my test. Although standard errors are clustered by firm to correct time series dependence in my model, tests were reperformed by keeping only the first cyber incident for each firm if it undergoes several cyber incidents to further address the concern. The results are still significant with the predicted directions when using this reduced sample (untabulated).

[11] An alternative explanation is that firms making cybersecurity risk disclosures are simply experiencing less severe cyber-attacks, which result in smaller increase in audit fees. However, this is not likely given that firms will disclose negative information only when they deem the risk is material. In fact, this will only bias against finding a negative interaction.

Table 6. Regression of Cyber incidents and Prior Cybersecurity Risk Disclosure on Audit Fees Increases using Equation (1)

Independent Variables	*Estimates*	*t-statistics*
Δ*Cyber-Incident*	0.120	3.77***
Disclosure	0.011	3.16***
Δ*Cyber-Incident* * *Disclosure*	-0.090	-2.38**
Δ*Lnassets*	0.273	32.57***
Δ*InvRec*	0.145	4.21***
Δ*Segments*	0.029	5.81***
Δ*Foreign*	0.008	1.03
Δ*Merger*	0.030	7.51***
Δ*Special*	0.026	8.82***
Δ*Loss*	0.026	6.33***
Δ*Growth*	-0.003	-0.73
Δ*Btm*	-0.004	-1.05
Δ*Big4*	0.343	19.40***
Δ*Initial*	-0.085	-11.17***
Δ*GCO*	0.004	0.35
Δ*ROA*	-0.125	-7.90***
Δ*Leverage*	0.061	4.08***
Δ*Quick*	-0.010	-6.03***
Δ*ICW*	0.081	10.87***
Residual	-0.157	-36.13***
Intercept	0.017	0.78
Industry Effects	Included	
Year Effects	Included	
Adjusted R square	27.62%	
# Observations	20,883	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Propensity Score Matching

Although propensity score matching is not the appropriate choice to address endogeneity arising from unobservable factors (Lennox et al., 2011; Shipman et al., 2017), which is a significant concern in the current context, the results were nevertheless examined using a traditional audit fee model in equation (2) using a propensity score matched sample. I generated propensity scores using a logistic regression that models the likelihood that a firm will experience cyber incidents[12]. The following logit model was used based on Wang, Kannan, et al. (2013), Higgs et al. (2014), and Sheneman (2017)):

$$\begin{aligned} Prob\ (Breach = 1) = {} & LNassets_{it} + Segments_{it} + ROA_{it} + Growth_{it} + Loss_{it} \\ & + Leverage_{it} + ICW_{it} + Year\ Indicators \\ & + Industry\ Indicators + \varepsilon_{it} \qquad (3) \end{aligned}$$

A detailed description of variable definitions can be found in Appendix A. After obtaining propensity scores, I matched each cybersecurity breached firm observations with non-breached firm observations that have propensity scores within 10% of the treatment firm. Table 7 summarizes the regression results using the propensity-matched sample. Column 1 indicates that audit fees are higher for firms experiencing cyber incidents ($p <$ 0.05), while Column 2 suggests that firms with prior cybersecurity risk disclosures have smaller fee increases ($p < 0.05$). Overall, findings using propensity score matching are similar to those reported in the main model.

[12] I reiterate that there is no well-specified model for explaining the probability of experiencing cyber incident.

Table 7. Regression Results of Equation (2) using Propensity Score Matched Sample

Independent variables	Cyber Incident		Prior Risk Disclosure	
	Estimates	*t-statistics*	*Estimates*	*t-statistics*
Cyber-Incident	0.131	2.01**	0.386	2.74***
Disclosure			0.197	2.09**
*Cyber-Incident*Disclosure*			-0.321	-2.12**
Lnassets	0.554	18.41***	0.543	15.86***
InvRec	1.442	3.25***	1.443	3.63***
Segments	0.019	0.78	0.038	1.60
Foreign	0.125	1.77*	0.187	2.43**
Merger	-0.003	-0.05	0.042	0.57
Special	0.062	0.7	-0.025	-0.26
Loss	0.141	1.42	0.033	0.32
Growth	-0.110	-0.97	-0.137	-1.08
Btm	-0.116	-1.62	-0.071	-0.90
Big4	0.356	2.81***	0.415	3.25***
Initial	-0.222	-1.16	-0.074	-0.50
GCO	-0.692	-3.46***	0.000	.
ROA	-1.012	-2.41**	-1.414	-3.01***
Leverage	0.044	0.23	0.086	0.42
Quick	-0.056	-1.86*	-0.038	-1.30
ICW	0.585	1.98**	0.367	1.25
Busy	-0.077	-0.93	-0.040	-0.46
Intercept	10.180	25.33***	9.330	16.86***
Industry Effects	Included		Included	
Year Effects	Included		Included	
Adjusted R square	79.59%		82.09%	
# Observations	545		412	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Additional Tests

Repeated Cyber Incidents

Since several firms experience multiple cyber incidents, I examine whether auditors are responding differently for firms having past cyber incidents. Firms experiencing more than one cyber incident can be hardly explained as coincidence because experiencing multiple cyber incidents could be indicative of severe weaknesses in firm's internal controls over operations and management's lack of commitment to maintain a sound internal control environment and remediate past vulnerabilities that result in the past cyber incidents. Thus, auditors are expected to perceive such firms as riskier and increase more audit fees.

An indicator variable *Past_Breach* was created to capture firm's past cyber incidents and interact this variable with Δ*Cyber-Incident*. The regression results are presented in Table 8. The coefficient on the interaction, Δ*Cyber-Incident* Past_Breach*, is positive and significant, suggesting that auditors increase larger audit fees for cybersecurity breached firms that have past cyber incidents. On average, the increase in audit fees for breached firms with past cyber incidents is more than twice of those that experience cyber incident for the first time (0.040+0.054 vs 0.040), demonstrating that auditors are especially concerned about the systematic problems underscored by repeated cyber incidents.

Table 8. Regression of Cyber incidents and Past Breach on Audit Fees Increases using Equation (1)

Independent Variables	*Estimates*	*t-statistics*
Δ*Cyber-Incident*	0.040	2.31**
Past_Breach	0.007	0.63
Δ*Cyber-Incident *Past_Breach*	0.054	2.12**
Δ*Lnassets*	0.276	37.05***
Δ*InvRec*	0.131	4.24***
Δ*Segments*	0.025	5.88***
Δ*Foreign*	0.009	1.28
Δ*Merger*	0.029	7.79***
Δ*Special*	0.025	9.37***
Δ*Loss*	0.024	6.61***
Δ*Growth*	0.000	0.02
Δ*Btm*	-0.004	-1.11
Δ*Big4*	0.334	21.26***
Δ*Initial*	-0.076	-12.05***
Δ*GCO*	0.013	1.35
Δ*ROA*	-0.126	-8.77***
Δ*Leverage*	0.058	4.25***
Δ*Quick*	-0.010	-6.66***
Δ*ICW*	0.074	11.34***
Residual	-0.152	-41.37***
Intercept	0.034	1.88*
Industry Effects	Included	
Year Effects	Included	
Adjusted R square	24.96%	
# Observations	29,853	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Type of Information Hacked

While cybersecurity breaches are generally more severe than other types of data breaches (e.g. stolen laptop) because it is initiated by malicious third parties, the type of information hacked could determine the severity and implications of the incident. In this section, I specifically consider intellectual property because intellectual property is the most important assets that firms should protect, and the damage of intellectual property theft could be material. Reuters (2015) reported that after Chinese hackers have stolen intellectual property from an Australian firm, the firm was forced to slash price of its products in half to compete with the counterfeiters. As intellectual property is the core of firm's value, theft of intellectual property could result in the forfeiture of competitive advantage, reduced market share, and loss of profitability (Gelinne, Fancher, & Mossburg, 2016). Compared with theft of customer personal information and credit card information, cybercrime towards intellectual property has stronger and more direct implications for firm's financial positions, including but not limited to future cash flows, valuation of intangible assets, and going concerns, all of which require auditors exert additional efforts to reduce the risk of material misstatement. In addition, since intellectual property is one of the most important assets for firms and has the strongest protection, breach of it could indicate material weakness in firm's internal controls over operations, which could be indicative of material weakness in internal controls over financial reporting (Lawrence et al., 2016). I create a variable *IP* that equals 1 if the cyber incident involves intellectual property, 0 otherwise. Δ*Cyber-Incident**IP* is added into equation 1 to capture the differential effect of different types of information hacked. Results are presented in Table 9. Consistent with my expectation, the interaction is statistically significant and negative

($p < 0.05$), suggesting that external auditors have differential responses to different types of cyber incidents.

Mitigating Channel

In this section, I explore whether auditor's reaction to cyber incident will be mitigated by external monitoring. Particularly, I focus on institutional ownership and block holders (i.e., shareholders who hold at least 5% of the shares outstanding). There is rich literature on the effect of block holders and institutional ownership on corporate governance. The overall finding is that larger block holders and institutional owners can improve corporate governance, mitigate agency problem, and reduce the risk of material misstatement and fraud (Edmans, 2014; Sharma, 2004). Because large and sophisticated shareholders provide active monitoring of corporate affairs and firm's accounting practices (Mitra, Hossain, & Deis, 2007), they may help mitigate auditor's concern to cyber incident as these firms post less risk to auditors. For example, these firms are less likely to have a weak control environment as they are actively monitored by large and sophisticated shareholders.

Two variables are used to capture external monitoring: the percentage of institutional holdings (*INST*) and the number of block holders (*NUM*). The results of interacting these two variables with Δ*Cyber-Incident* are summarized in Table 10. Both interactions are negatively associated with the increase in audit fees, providing evidence that external monitoring could mitigate auditor's concern over cyber incident.

Table 9. Regression of Cyber incidents and Intellectual Property on Audit Fees Increases using Equation (1)

Independent Variables	*Estimates*	*t-statistics*
Δ*Cyber-Incident*	0.024	1.65*
IP	-0.057	-5.47***
Δ*Cyber-Incident *IP*	0.092	2.21**
Δ*Lnassets*	0.276	36.98***
Δ*InvRec*	0.130	4.19***
Δ*Segments*	0.024	5.83***
Δ*Foreign*	0.008	1.19
Δ*Merger*	0.029	7.78***
Δ*Special*	0.025	9.39***
Δ*Loss*	0.024	6.59***
Δ*Growth*	0.000	0.10
Δ*Btm*	-0.003	-1.05
Δ*Big4*	0.334	21.26***
Δ*Initial*	-0.076	-12.00***
Δ*GCO*	0.014	1.37
Δ*ROA*	-0.127	-8.79***
Δ*Leverage*	0.059	4.30***
Δ*Quick*	-0.010	-6.61***
Δ*ICW*	0.075	11.46***
Residual	-0.152	-41.31***
Intercept	0.034	1.88*
Industry Effects	Included	
Year Effects	Included	
Adjusted R square	25.05%	
# Observations	29,682	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

Table 10. Regression of Cyber incidents and External Monitoring on Audit Fees Increases using Equation (1)

Independent variables	*(1)*		*(2)*	
	Estimates	*t-statistics*	*Estimates*	*t-statistics*
Δ*Cyber-Incident*	0.072	4.11***	0.075	3.22***
NUM	0.001	1.12		
Δ*Cyber-Incident *NUM*	-0.019	-2.46**		
INST			0.005	1.20
Δ*Cyber-Incident *INST*			-0.068	-1.76*
Δ*Lnassets*	0.277	37.18***	0.276	36.87***
Δ*InvRec*	0.133	4.29***	0.133	4.29***
Δ*Segments*	0.025	5.91***	0.025	5.92***
Δ*Foreign*	0.009	1.31	0.009	1.31
Δ*Merger*	0.029	7.70***	0.029	7.69***
Δ*Special*	0.025	9.37***	0.025	9.37***
Δ*Loss*	0.024	6.61***	0.024	6.62***
Δ*Growth*	0.000	-0.02	0.000	-0.02
Δ*Btm*	-0.004	-1.11	-0.004	-1.08
Δ*Big4*	0.335	21.27***	0.335	21.25***
Δ*Initial*	-0.076	-12.05***	-0.076	-12.06***
Δ*GCO*	0.014	1.36	0.014	1.37
Δ*ROA*	-0.127	-8.80***	-0.126	-8.78***
Δ*Leverage*	0.059	4.32***	0.060	4.35***
Δ*Quick*	-0.010	-6.66***	-0.010	-6.65***
Δ*ICW*	0.074	11.31***	0.074	11.31***
Residual	-0.152	-41.35***	-0.152	-41.34***
Intercept	0.034	1.89*	0.035	1.91*
Industry Effects	Included		Included	
Year Effects	Included		Included	
Adjusted R square	24.99%		24.99%	
# Observations	29,761		29,761	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels (one-tailed), respectively. Test statistics are based on coefficient standard errors that are heteroscedasticity-consistent and are clustered at firm level. Estimated coefficients for year and industry dummy variables are not reported for brevity. All Variables are defined in Appendix A.

2.5 CONCLUDING REMARKS

This chapter demonstrates a potential relationship between the external audit and cyber incidents. Specifically, using data on cyber incidents for the period 2005 to 2015, I empirically examine the relationship between the increase in audit fees and cyber incidents. Consistent with my expectation, a significant positive association is observable between audit fee increase and cyber incidents using an audit fee change model. In addition, increases in audit fees are smaller for firms with prior cybersecurity risk disclosure following cyber incidents, implying that auditors have priced material cybersecurity risk prior to the cyber-attacks. Evidence in this chapter also demonstrates that firms with repeated cyber incidents are charged higher audit fees than firms that are only breached for the first time. Furthermore, auditors differentiate the type of information hacked. Increases in Audit fees are higher for firms with cyber incidents that involve intellectual property than for firms not involving intellectual property hacking. Finally, I document that auditor's concern over cyber incidents is mitigated by external monitoring, as measured by the percentage of institutional holdings and number of block holders. Collectively, results in this chapter should be valuable to regulators and academics who are interested in understanding auditor's opinion over cyber incidents. The findings that auditors both price cybersecurity risk ex-ante and respond to cyber incidents ex-post disagree with the concern that auditors are not taking cybersecurity seriously.

There are several limitations that must be considered when interpreting the findings. Although auditors should respond to cyber incidents because they may indicate deficiencies in ICFR and risks of material misstatement, there could be other reasons why external auditors would increase audit fees following a cyber incident. In-depth case studies or interviews with external auditors should be conducted to build a more comprehensive understating of how external auditors respond to cybersecurity risks and cyber incident. In addition, the results of the study do not address how external auditors are evaluating cybersecurity risks prior to cyber incidents. A thorough investigation is necessary to advance our understanding of cybersecurity risk anticipation. For example, analogous to “contagion” effects in stock price reactions reported by Ettredge and Richardson (2003), do auditors of firms that are similar to firms that have experienced cyber incidents increase their audit procedures and audit fees to identify potentially unidentified cyber incidents among those clients and to address potential consequences?

CHAPTER 3: CYBERSECURITY RISK DISCLOSURE AND CYBERSECURITY DISCLOSURE GUIDANCE

3.1 INTRODUCTION

Cybersecurity has attracted much attention in the past ten years. Both the general public and the business world are concerned about the growing cybercrimes that expose sensitive personal information, cause business disruptions, or steal trade secrets, especially after a series of high-profile data breaches such as the ones at Target, Home Depot, and Yahoo. According to a recent Annual Cybersecurity Report, more than 20% of the breached firms experienced substantial loss of revenue, customer base, and business opportunities, and most of the breached firms spent millions of dollars improving defense technologies and expanding security procedures following the attacks (CISCO, 2017). Due to the potential impact on firm value and operation, cybersecurity is becoming one of the top priorities for firm executives. About 88% of U.S. Chief Executive Officers (CEOs) are concerned that cyber threats could hinder the growth of their firms (Loop, 2016). Likewise, investors are clamoring for more information about cybersecurity risks and data breaches, and how firms are addressing those risks (Shumsky, 2016).

To respond to the increasing cyber threats, the Securities and Exchange Commission (SEC) held a roundtable discussion to deliberate cybersecurity landscape and cybersecurity disclosure issues (SEC, 2014). The Standing Advisory Group of the Public Company Accounting Oversight Board (PCAOB) also discussed the potential implications of cybersecurity on financial reporting and auditing (PCAOB, 2014). Particularly, the SEC's Division of Corporation Finance issued a disclosure guidance regarding cybersecurity in 2011 to assist firms in assessing what, if any, disclosures should be provided related to cybersecurity risks and cyber incidents (SEC, 2011). Although the guidance is not technically a ruling, the SEC has issued comment letters to several firms pointing out the inadequacies of their cybersecurity risk disclosures by referring to the guidance. Therefore, some have argued that the guidance is becoming a de facto ruling (Grant & Grant, 2014).

In this chapter, I investigate the informativeness of cybersecurity risk disclosures in the risk factor section of annual report (thereafter cybersecurity risk disclosure). Informativeness of cybersecurity risk disclosures is defined in this study as "*the ability to help stakeholders assess the probability of future adverse events (i.e. cyber incidents)*." Understanding the information conveyed by cybersecurity risk disclosures is important as it can help investors to assess firm's cybersecurity risk, and shed light on any potential subsequent legislative rules regarding cybersecurity disclosures. Cybersecurity disclosures, particularly cybersecurity risk disclosure, have been criticized to be uninformative and boilerplate by both practitioners and academics. They argue that firms use boilerplate language every year (Bennett, 2015), a common criticism for risk factor disclosures in Item 1A. To examine the effectiveness of public firm disclosures, during the roundtable discussion organized by the SEC in 2014, a panel was formed to discuss disclosures concerning cybersecurity risks and cyber incidents, "*focusing on what public firms are currently disclosing about their cybersecurity threats and breaches, both potential and those that have already occurred, and how they determine the appropriate disclosure, the timing of that disclosure, and what information about cybersecurity investors need to know to make informed voting and investment decisions*".[1] Most panelists

[1] For more details, visit https://www.sec.gov/spotlight/cybersecurity-roundtable/cybersecurity-roundtable-transcript.txt.

raised concerns that many cybersecurity disclosures are boilerplate and admitted the difficulty striking the balance between providing meaningful disclosure and not adversely affecting the firm’s reputation and performance. For example, Keith Higgins, the director of the Division of Corporation Finance of the SEC, indicated in the panel discussion: “*If you take boilerplate on the one hand and on the far side you take a look at the specific road map of the company’s vulnerabilities and what the consequences of those vulnerabilities could be, where do you find the balance? How do you -- is there somewhere in the middle that will be helpful to investors while at the same time not harmful to companies?*” The issue is further complicated by the lack of clarity in the SEC’s cybersecurity disclosure guidance. As the guidance acknowledged, there is no explicit requirement for disclosure of cybersecurity risks or cyber incidents so far. The guidance only pointed out several areas where cybersecurity disclosures may be necessary. Accordingly, firms have great discretion in deciding whether, what, and how much to disclose.

Two aspects of cybersecurity risk disclosure are considered: presence and length. Specifically, I examine whether the presence of cybersecurity risk disclosure in a firm’s annual report signals higher cybersecurity risk as measured by subsequent cyber incidents, and whether the content of the disclosure, measured by adjusted length, is associated with increased likelihood of subsequent cyber incidents. The findings suggest that both the presence and the length of cybersecurity risk disclosure are associated with subsequent cyber incidents, indicating that cybersecurity risk disclosure is informative. There is a substantial increase in the percentage of firms that disclose cybersecurity risks following the SEC’s disclosure guidance, and that the presence of cybersecurity risk disclosure is no longer associated with subsequent cyber incidents in the post-guidance period, suggesting that the SEC's guidance led to more cybersecurity risk disclosures by firms regardless of their degree of cybersecurity risk. To examine the SEC's concern that more firm-specific disclosure may compromise firms' cybersecurity efforts by providing a roadmap to malicious parties, two measures based on the bag-of-words approach are created to capture firm-specific disclosure. I fail to find a significant association between cyber incidents and any of the two measures, demonstrating that cybersecurity risk disclosures in firm's annual report are far from the level of detail that could eventually hurt the firm.

An important question not addressed in the above findings is whether the market participants are utilizing information in cybersecurity risk disclosures. Contrary to Hilary et al. (2017), I find evidence that abnormal return calculated over the three days around the disclosure of a cyber incident is positively associated with firm's prior presence of cybersecurity risk disclosure. However, the content in the disclosure is not incorporated by the market participants as the adjusted length of disclosures describing cybersecurity risk is not associated with the market reaction. The consequences of cybersecurity risks and cyber incidents that firms are most concerned about are further examined. The topic analyses show that business disruption and financial performance are the two major concerns and remain relatively steady over time. Concerns over intellectual property and reputation, on the other hand, are relatively low but are increasing rapidly in recent years.

The findings of this study make several contributions to the existing literature. First, the study contributes to the cybersecurity disclosure literature. Early studies on cybersecurity focus on the market reaction following cyber incidents and have examined a set of contingency factors such as type of breaches (Gordon et al., 2011; Yayla & Hu, 2011), firm characteristics (Ettredge & Richardson, 2003), and distribution channels

(Benaroch et al., 2012) that could deepen or mitigate the market reaction. I extend this literature by showing that the investors are less surprised when there is prior disclosure of cybersecurity risks. Specific to cyber-related disclosure, Gordon et al. (2010) find that on average, voluntary disclosure relating to information security increases stock prices by more than 6 percent, and the voluntary disclosure concerning proactive security measures have the greatest impact on the firm's stock price, followed by the disclosure of vulnerabilities. This study complements Gordon et al. (2010) by exclusively focusing on cybersecurity risks (vulnerabilities) and providing evidence that cybersecurity risk disclosure is informative of future cyber incidents, and that the market reaction following cyber incidents is contingent on the presence of cybersecurity risk disclosure. Wang, Kannan, et al. (2013) examined the ex-post odds of cyber incidents and market reaction following voluntary disclosures, revealing that firms that disclose information security risk factors in their annual reports with actionable information are less likely to be associated with future cyber incidents. Firms that did not provide any actionable plans will be punished more severely when an actual incident happens than firms that disclosed actionable information. The paper complements Wang, Kannan, et al. (2013) in at least three key ways. My sample includes 326 cyber incidents, which is much larger than 62 cyber incidents in their study. More importantly, the sample covers both the pre-guidance period and the post-guidance period, which enables me to examine the changes in disclosure informativeness. Different from Wang, Kannan, et al. (2013), the identification of individual cybersecurity risk factors is automated by benefiting from text mining techniques, especially taking advantage of the contextual clues in HyperText Markup Language (HTML) tags. My approach enables analyses on a much larger scale to demonstrate that firms facing greater cybersecurity risks devote a greater portion of their disclosures towards describing cybersecurity risks. Another significant difference is that I empirically show the presence or absence of cybersecurity risk disclosure is valuable information, which is not explicitly examined in prior studies.

Second, this research also contributes to the risk disclosure literature. While findings in the study are largely in line with recent accounting literature showing that risk factor disclosure is not boilerplate, I use the actual adverse event (i.e., cyber incident) rather than market-based measures of firm risks (Campbell et al., 2014) or investors' risk perceptions (Kravet & Muslu, 2013) to capture the risks that a firm faces. As the objective of providing risk factor disclosure is to discuss "the most significant factors that make the firm risky" (SEC, 2005), my risk measure that focuses on actual risk event is more consistent with the SEC's intention than measures based on the assumption of market efficiency, and provides more direct evidence that risk disclosures are informative of future operational failures. Furthermore, different from prior studies that examined the variation of qualitative disclosures that are already included in risk factor disclosure section, my unique setting allows me to show that the presence or absence of risk disclosure could be informative of the risk. The study also indirectly demonstrates that market participants use information released in cybersecurity risk disclosures. This finding is in contrast with those reported in Hilary et al. (2017), but is consistent with prior studies that show investors incorporate information conveyed by risk factor disclosures into firm stock price (Campbell et al., 2014; Hope et al., 2016). I attribute the inconsistency with Hilary et al. (2017) to the difference in the sample characteristics (more types of cyber incidents such as hacking of intellectual property are included), sample size, and the way of identifying cybersecurity risk disclosures. The software and computer industry are excluded in this study.

Third, this paper makes contributions to the textual analysis literature. When examining disclosures related to cybersecurity, prior studies use manual collection (Wang, Kannan, et al., 2013), take several number of words around the keywords (Gordon et al., 2010), or simply count the number of predetermined keywords (Hilary et al., 2017). I develop methods that first identify individual risk factors from item 1A and then identify security-related risk factors. This helps me to more accurately examine the content of cybersecurity risk disclosure, and is also consistent with recent research effort that calls for analysis at individual risk factor level (Bao & Datta, 2014; Gaulin, 2017). In addition, the topic analysis using word-term patterns help to obtain a thorough understanding with respect to the consequences of cyber incidents that firms are most concerned about, which is not examined in prior studies.

Fourth, the results could also help policymakers to determine the benefits and consequences of cybersecurity risk disclosures and disclosure guidance. The findings support the decision to emphasize cybersecurity risk disclosures, as both the presence and the content of cybersecurity risk disclosures are informative of subsequent cyber incidents. However, my findings also reveal that the SEC's disclosure guidance leads to an unintentional consequence that more firms make cybersecurity risk disclosures even though they do not face higher cybersecurity risks. As the SEC warned firms to "avoid generic risk factor disclosure that could apply to any company", the outcome is counter to the SEC' intention. Such outcome is caused by the ambiguity in the guidance and comment letters sent by the SEC to force firms to disclose cybersecurity risks (Ferraro, 2013).

Therefore, it may be necessary for the SEC to revise the guidance to encourage firms who are exclusively facing higher cybersecurity risks to make such disclosures. The SEC may not want to elevate the guidance to the commission level, a suggestion made by Senator Jay Rockefeller in 2013, as that may push more firms to issue cybersecurity risk disclosures without having high cybersecurity risks. Additionally, while Ferraro (2013) criticizes that the SEC did little to resolve the concern about revealing too much information publicly could provide potential hackers with a roadmap for successful attacks, I find no evidence supporting such claim.

The remainder of this chapter is organized as follows. The next section provides research background and hypothesis development. This is followed by the details of sample selection procedures and research methodology. Next, empirical results and additional analyses are presented. The last section concludes this paper.

3.2 BACKGROUND AND HYPOTHESES DEVELOPMENT

Risk Factor Disclosure

On June 29, 2005, the SEC mandated firms to describe “the most significant factors that make the offering speculative or risky” in Item 1A of 10-K filed after December 1, 2005 with the objective being “to provide investors with a clear and concise summary of the material risks to an investment in the issuer’s securities” (SEC, 2005). Since firms are only required to provide qualitative descriptions and do not need to quantify the likelihood or impact of the disclosed risks, they have a great degree of discretion in what to disclose and how to disclose. Practitioners criticize that managers are likely to provide vague risk disclosure and simply list all uncertainties they face, providing little information for investors (Reuters, 2005). Similarly, Robbins and Rothenberg (2005) argue that risk factor disclosures are the cheapest form of insurance as "*firms that cannot point to such a risk factor when faced with a lawsuit will wish they could turn back the clock and insert such language*", implying that firms have incentives to make uninformative risk factor disclosures for legal protection. Realizing the problem, the SEC has issued comment letters to require more risk information from firms (Johnson, 2010), and has warned firms to "*avoid risk factor disclosure that could apply to any issuer or any offering*" (SEC, 2010).

The concern that risk factor disclosures may be boilerplate is alleviated by recent studies. Campbell et al. (2014) show that firms disclose more risk factors when facing greater risks, and devote a greater portion of the disclosures towards describing risks that are more significant. They also find that the unexpected portion of risk factor disclosures is associated with systematic risk, idiosyncratic risk, information asymmetry, and abnormal returns following the disclosure, indicating that the information conveyed by risk factor disclosures is perceived by market participants. Similarly, Kravet and Muslu (2013) reveal that increases in the number of risk-related sentences are positively associated with stock volatility, trading volume around and after the filings, and dispersed forecast revisions around the filings. However, the effect is largely driven by industry-level risk disclosures rather than firm-level disclosures. Hope et al. (2016) demonstrate that the level of specificity in risk factor disclosures is positively associated with the market reaction to 10-K filings and can help analysts assess firms' fundamental risk. Two contemporary papers examine the effect of comment letters. Brown, Tian, and Tucker (2015) identify that firms significantly modify their risk factor disclosures after receiving comment letters. More importantly, spillover effect exists in that firms not receiving comment letters still revise their risk factor disclosures if industry leader, close rival, or industry peers receive comment letters regarding risk factor disclosures, suggesting a deterrence benefit of the SEC's review process. Beatty, Cheng, and Zhang (2015) find that financial constraints risk factor disclosures are associated with firms' expected level of financial constraints, ex-ante litigation risk, and realized financial constraints outcomes. However, the association is significantly reduced after firms increase disclosures to respond to comment letters, demonstrating the concerns that firms may make disclosures that they otherwise deem immaterial simply to fulfill regulatory requirement. Several recent studies focus on risk factor updates. Filzen (2015) indicates that firms with risk factor updates in their quarterly reports have lower abnormal returns around the filing dates, lower future unexpected earnings, and larger likelihood of experiencing future negative earnings shock. A subsequent study by Filzen, McBrayer, and Shannon (2016) documents that quarterly risk factor updates are negatively associated with future returns and that the association is stronger for firms using more direct words related to firm fundamentals. Finally, Gaulin (2017) emphasizes the importance of using individual risk factors by showing that managers add new risk factors and remove stale risk factors on a timely basis, and that such activities predict future economic changes even after controlling for ex-ante risk and firm performance. In addition, firms respond to the SEC comment letters by improving the level of specificity while they respond to securities litigation by expanding the number of risks they identified without increasing the definitiveness of the disclosures, supporting the litigation shield hypothesis.

Cybersecurity Disclosure Guidance

In 2011, the SEC's Division of Corporation Finance issued disclosure guidance related to cybersecurity, pointing out sections that may be relevant for cybersecurity-related disclosure. Regarding risk factor disclosure, the guidance states that "*in determining whether risk factor disclosure is required, we expect registrants to evaluate their cybersecurity risks and take into account all available relevant information, including prior cyber incidents and the severity and frequency of those incidents…. Registrants should consider the probability of cyber incidents occurring and the quantitative and qualitative magnitude of those risks, including the potential costs and other consequences resulting from misappropriation of assets or sensitive information, corruption of data or operational disruption*" (SEC, 2011). Although the guidance explicitly specifies that it is not a ruling, the SEC has used comment letters to prompt cybersecurity risk disclosures. For example, in the comment letter addressing Freeport-McMoRan Copper & Gold Inc.'s annual report of 2011, the SEC states that: "*We note that none of your risk factors, or other sections of your Form 10-K, specifically address any risks you may face from cyber attacks, such as attempts by third parties to gain access to your systems to compromise sensitive business information, to interrupt your systems or otherwise try to cause harm to your business and operations. In future filings, beginning with your next Form 10-Q, please provide risk factor disclosure describing the cybersecurity risks that you face or tell us why you believe such disclosure is unnecessary*." Since comment letters are often considered as de facto rulings, it is argued that the disclosure guidance is becoming disclosure requirement (Grant & Grant, 2014).

Research on the disclosure guidance is recently emerging. Ferraro (2013) argues that the disclosure guidance both procedurally overreaches and substantively underachieves. The author criticizes that the SEC is using the non-legislative guidance as a legislative rule. More importantly, the paper points out that the guidance is vague, similar across industries that will bring little information to the market. Consistent with this view, Hilary et al. (2017) fail to find a significant association between the market reaction following cyber incidents and firms' prior cyber disclosures.

Hypotheses Development

All the hypotheses in this study focus on the informativeness of cybersecurity risk disclosure (i.e., predictability for future cyber incidents) and on whether the investors incorporate risk information conveyed by cybersecurity risk disclosures (i.e., whether the stock price will change based on cybersecurity risk disclosure). The maintained assumptions underlying the hypotheses are that managers are at least partially knowledgeable about the cyber threats firms face and the security measures they have taken, and that the market has certain degree of efficiency.

The first hypothesis centers on the presence of cybersecurity risk disclosures. The disclosure literature suggests that managers have incentives to disclose favorable information and withhold negative information (Beyer, Cohen, Lys, & Walther, 2010; Verrecchia, 2001). The bias against providing bad news result from concerns over increasing cost of capital, damaging future career opportunities, and revealing proprietary information to competitors (Ke, Huddart, & Petroni, 2003; Kothari, Li, et al., 2009; Kothari, Shu, et al., 2009). Hence, if cybersecurity risk disclosure includes unfavorable information, managers are less willing to disclose such information.

Although managers have incentives to withhold negative information due to business and career concerns, they may face legal penalties for not disclosing such information. Litigation costs could be high enough to motivate disclosures of bad news (Skinner, 1994). Consistent with this view, recent studies document that risk factor disclosures are generally informative (Campbell et al., 2014; Hope et al., 2016; Kravet & Muslu, 2013). Particularly, with respect to cybersecurity risk disclosure, lawsuits may be filed if a material cyber incident happens, but the firm fails to alert the investors about the risk in advance. For example, Heartland Payment Systems was sued for "*misrepresenting or failing to disclose that the company's safety and security measures designed to protect consumers' financial records and data from security breaches were inadequate and ineffective*".[2]

Taken together, firms tend to provide cybersecurity risk disclosure when they deem the risk as a material matter. That is, firms that provide cybersecurity risk disclosures face higher cybersecurity risk, and thus are more likely to experience cyber incidents. Accordingly, the following hypothesis is introduced.

H1: The presence of cybersecurity risk disclosure is positively associated with the likelihood of subsequently reported cyber incident.

The second hypothesis examines the content of cybersecurity risk disclosure. While the presence of cybersecurity risk disclosure signals elevated cybersecurity risk that prompts the firm to disclose, the variation of the disclosure content could also be informative. Consider the following two cybersecurity risk disclosures:

> 1) Security breaches or intrusion into our information systems, and the breakdown, interruption in or inadequate upgrading or maintenance of our information processing software, hardware or networks may impact our business. Security breaches or intrusion into the systems or data of the third parties with whom we conduct business may also harm our business.[3]
>
> 2) Experienced computer programmers and hackers may be able to penetrate our security controls and misappropriate or compromise our confidential information or that of third parties, create system disruptions or cause shutdowns. Computer programmers and hackers also may be able to develop and deploy viruses, worms and other malicious software programs that attack our websites, products or otherwise exploit any security vulnerabilities of our websites and products. The costs to

[2] For more details, visit http://securities.stanford.edu/filings-case.html?id=104260.

[3] Excerpted from the 10-K of GRACO INC for the fiscal year 2013 (https://www.sec.gov/Archives/edgar/data/42888/000119312514056452/d675621d10k.htm).

> us to eliminate or alleviate cyber or other security problems, bugs, viruses, worms, malicious software programs and security vulnerabilities could be significant, and our efforts to address these problems may not be successful and could result in interruptions, delays, cessation of service and loss of existing or potential customers that may impede our sales, manufacturing, distribution or other critical functions. We manage and store various proprietary information and sensitive or confidential data relating to our business and third party business. Breaches of our security measures or the accidental loss, inadvertent disclosure or unapproved dissemination of proprietary information or sensitive or confidential data about us or our partners or customers, including the potential loss or disclosure of such information or data as a result of fraud, trickery or other forms of deception, could expose us, our partners and customers or the individuals affected to a risk of loss or misuse of this information, result in litigation and potential liability for us, damage our brand and reputation or otherwise harm our business. In addition, the cost and operational consequences of implementing further data protection measures could be significant. Delayed sales, significant costs or lost customers resulting from these system security risks, data protection breaches, cyber-attacks and other related cybersecurity issues could adversely affect our financial results, stock price and reputation.[4]

It may be inaccurate to treat these two cybersecurity risk disclosures the same as they differ significantly in the amount of information provided. Practitioners, regulators, and academics have expressed concerns that cybersecurity risk disclosures may be boilerplate (Bennett, 2015; Hilary et al., 2017). If the concern is true, the content of cybersecurity risk disclosure is not expected to be associated with the likelihood of reported future cyber incidents. On the other hand, Campbell et al. (2014) show that the level of risk determines the amount of disclosure firms devote to address that risk. Similarly, Filzen (2015) argues that the more discussions of potential negative outcomes, the greater the likelihood of the negative event. If cybersecurity risk disclosure is informative, it is expected that firms facing higher cybersecurity risks are more likely to devote a greater portion of the disclosures to describe their cybersecurity risks. Therefore, it is an interesting empirical question whether the content of cybersecurity risk disclosure, as measured by adjusted length to capture the relative importance of the risk in firm's risk portfolio, is informative. This leads to the following hypothesis.

H2: The length of cybersecurity risk disclosure is positively associated

[4] Excerpted from the 10-K of DIODES INC for the fiscal year 2013 (https://www.sec.gov/Archives/edgar/data/29002/000119312514073365/d633786d10k.htm).

with the likelihood of subsequently reported cyber incident.

The next hypothesis concentrates on the market perception of cybersecurity risk disclosure. Prior studies indicate that changes in risk factor disclosures are associated with abnormal returns surrounding the release date, information asymmetry, analyst forecast dispersion, and risk perceptions (Campbell et al., 2014; Filzen, 2015; Hope et al., 2016; Kravet & Muslu, 2013). However, such studies examine risk factor disclosures at an aggregate level, rather than individual risk factor level. It is ex-ante not clear whether the market incorporates information conveyed by the disclosure that describes cybersecurity risk. Since directly examining the market reaction to cybersecurity risk disclosure is not feasible due to confounding effects such as information contained in the concurrently released 10-K filings, I indirectly test whether the market reaction following cyber incident is conditional on firms' disclosure practices. If investors incorporate information from cybersecurity risk disclosure, they should respond less severely for firms with prior cybersecurity risk disclosure.

H3a: The market reaction following cyber incident is less severe for firms with prior cybersecurity risk disclosure.

H3b: The market reaction following cyber incident is less severe for firms with lengthy cybersecurity risk disclosure.

The last hypothesis investigates the effect of the SEC's cybersecurity disclosure guidance. Firms are increasingly disclosing their cybersecurity risks following the guidance. The percentage of firms providing cybersecurity risk disclosures jumps from 27.29% in 2010 to 42.12% in 2011 (see Figure 1). However, less is known about whether the increase in regulatory pressure will result in uninformative disclosures. Since risk factor disclosure in Item 1A is qualitative and does not require assessment of probability, firms may disclose all possible risk factors to fulfill regulatory requirements (Campbell et al., 2014). Consistent with this view, Beatty et al. (2015) document that disclosures become less reflective of future financial constraints following the SEC comment letters. To the extent that the SEC's cybersecurity disclosure guidance could be viewed as a regulatory shock, the following hypothesis is examined in this study.

H4: The association between the presence of cybersecurity risk disclosure and subsequent cyber incident is different before and after the introduction of the SEC's cybersecurity disclosure guidance.

3.3 EMPIRICAL DESIGN AND SAMPLE SELECTION

Empirical Design

The first hypothesis predicts that the presence of cybersecurity risk disclosure is associated with subsequent cyber incidents. Variable *Disclosure* is constructed that equals to one if there is any cybersecurity risk disclosure in that fiscal year, zero otherwise. To examine the second hypothesis, I create the variable *length* that measures the total word count of cybersecurity risk disclosure, normalized by the average word count of individual risk factors for that firm-year. The normalization is important as it controls for a firm's tendency to provide longer disclosure. A logit model is estimated with *Breach* as the dependent variable that takes the value of one if the firm experiences cyber incident in year t+1, zero otherwise.

$$P(Breach_{it+1} = 1) = Cyber\_dis_{it} + Past\_breach_{it} + Size_{it} + LN\_segments_{it} + Age_{it} + Loss_{it} + LN\_analyst_{it} + Foreign_{it} + Merger_{it} + Growth_{it} + ICW_{it} \qquad (1)$$

Appendix C provides a detailed definition of each variable. *Cyber_dis* is the variable of interest, be it either *Disclosure* or *Length*. A positive coefficient on this variable would support the hypotheses. A set of control variables based on prior literature are also included (Hilary et al., 2017; Sheneman, 2017; Wang, Kannan, et al., 2013). Specifically, I control for consumer and finance industry as these are the two sectors that witness most cyber incidents. Positive coefficients are expected for firm size, age, growth, and number of analysts following, as these variables control for the visibility of the firm. Further, firm's financial conditions are controlled using *Loss*. As financially constrained firms are less likely to invest sufficiently into their financial reporting control systems (Doyle, Ge, & McVay, 2007), it is expected that firms with losses are also less likely to make sufficient investment in their internal controls over operations. $Foreign_{it}$, $Merger_{it}$, and $LN\_segments_{it}$ are included to control for the complexity of a firm's business. Positive coefficients are expected on these variables as more complex and dispersed operations are likely to result in ineffective and inconsistent controls (Sheneman, 2017). *ICW* is included to control for a firm's internal control environment. Since internal controls over financial reporting and internal controls over operations are correlated (Lawrence et al., 2016), Firms with material weaknesses in internal controls over financial reporting are more likely to experience cyber incidents. Finally, an indicator variable *Past_breach* is included to capture whether the firm had cyber incidents in any previous year.

For testing the market reaction, abnormal returns over the three days around the cyber incident announcement date are calculated and adjusted using the Fama-French three-factor model. Similar to Hilary et al. (2017), observations that are confounded with earnings announcements and 8-K filings are removed. In addition, the time period (pre or post guidance), market cap, book-to-market ratio, leverage, loss, and severity of the cyber incident are also controlled. I expect negative coefficients on loss and severity while a positive coefficient on market cap.

$$CAR_{it} = Cyber\_dis_{it} + Guidance_{it} + Market\_cap_{it} + Severity_{it} + Leverage_{it} + Btm_{it} + Loss_{it} + \varepsilon_{it} \qquad (2)$$

Variable definitions are provided in Appendix C.

Sample Selection

The cyber incident data comes from Privacy Rights Clearinghouse (privacyrights.org) and Audit Analytics cybersecurity database. Privacy Rights Clearinghouse publishes data breaches that involve individual's identity while the Audit Analytics cybersecurity database collects hacking incidents. To identify cybersecurity risk disclosures, item 1A of the 10-K is first extracted using an approach similar to Campbell et al. (2014)[5]. After obtaining the whole item 1A section, individual risk factors are identified using information provided in HTML tags. The SEC requires that each risk factor should be preceded by a subcaption that summarizes that risk[6]. Similar to Gaulin (2017), I identify each subcaption that is highlighted (bold, italic, or underlined) and is located at the beginning of a paragraph or isolated on a separate line. The content between two highlighted subcaptions is considered to be a unique risk factor. A detailed description of the procedure can be found in Appendix D.

keyword search is then used to identify risk factors related to cybersecurity. These keywords are identified from prior research (Gordon et al., 2010; Wang, Kannan, et al.,

[5] All the 10-K filings filed between January 2005 and December 2015 are downloaded.

[6] Item 503(c) of Regulation S-K.

2013) and have been refined to prevent misidentification.[7] Risk factors that contain any of these keywords are considered cybersecurity risk disclosure. To ensure the quality of the identification, I randomly selected 200 documents for manual inspection. All of them are accurately identified. Appendix E provides a list of these keywords.

[7] Several keywords that could generate false positives are excluded. For example, while Trojan typically refers to malicious program that is used to hack into a computer, it can also refer to a condom brand. In addition, some new keywords are added, such as ransomware and key logger.

Table 11. Sample Selection

Number of firm-years with cyber incidents		326
Original number of cyber incidents	758	
Minus: observations that have more than one cyber incidents in a year (keep each firm-year only once)	(-78)	
Minus: observations that are in the computer and software industry (SIC 3570-3579, 7370-7379)	(-93)	
Minus: observations for which item 1A cannot be extracted	(-185)	
Minus: observations that have missing values on any one of the variables used in the study	(-76)	
Number of firm-years without cyber incidents		29,205
Total number of firm-years		29,531

Table 11 summarizes the sample selection procedure. The paper starts with 790 cyber incidents that can be mapped to Compustat[8]. For firms that experienced more than one cyber incidents in the same year, only one incident is kept in the sample. Observations in the software and computer industry (SIC between 3570-3579 and 7370-7379) are further deleted because their cybersecurity risk disclosures cannot be accurately determined[9]. Lastly, observations that I cannot extract item 1A and observations that have missing values on any of the independent variables are deleted. In total, the sample contains have 29,205 non-breached observations and 326 breached observations.

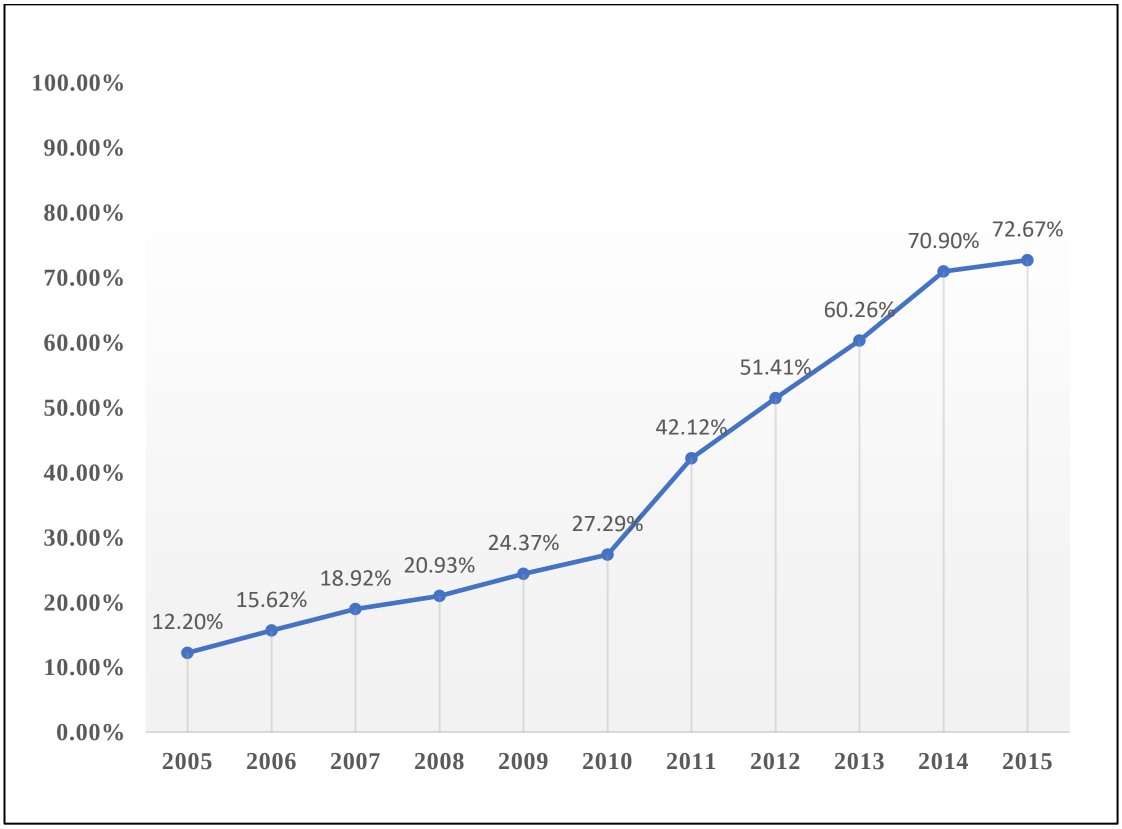
Figure 1 shows the percentage of firms providing cybersecurity risk disclosures in

[8] Many incidents reported in the database occur in non-profit or private firms, thus cannot be linked to Compustat.

[9] Business of firms in these industries could include providing security solutions to customers. The keyword search method will misidentify these security solutions as risk factors related to their business. For instance, disclosure regarding how the sales of intrusion detection products would influence stock price is incorrectly identified as risk factor, which has nothing to do with cybersecurity risk.

the sample. While the overall trend is upward, there is an unusual jump following the SEC's cybersecurity disclosure guidance in 2011. In addition, the annual increase in the percentage of firms providing cybersecurity risk disclosures is much larger following the disclosure guidance.

Figure 1. Percentage of Cybersecurity Risk Disclosures Across Years



Panel A of Table 12 presents the descriptive statistics of the variables used in this study. The mean for *Breach* is 0.011, suggesting that only about 1% of the firms in my sample experience cyber incidents. The percentage of firms making cybersecurity risk disclosures are 69.9% (with cyber incidents) and 36% (without cyber incidents), providing initial support for my argument that firms with high cybersecurity risks are more likely to provide cybersecurity risk disclosures. Similarly, cybersecurity risk disclosures of firms with cyber incidents are much longer than those of firms without cyber incidents (2.086 vs

1.545). Panel B of Table 12 describes the variables for testing the market reaction. The mean and median *CAR* is -0.2% and -0.3%, respectively, indicating that investors view cyber incidents as adverse events.

Table 13 reports the univariate correlations among variables examined in this study. The variables of interest, *Disclosure* and *Length*, are both positively correlated with the dependent variable *Breach*.

Table 12. Descriptive Statistics

Panel A: Descriptive Statistics for Variables in Equation (1)

Variable	*Total sample*			*Firms without cyber incidents*			*Firms with cyber incidents*		
	(n=29,531)			*(n=29,205)*			*(n=326)*		
	Mean	*Std*	*Median*	*Mean*	*Std*	*Median*	*Mean*	*Std*	*Median*
Breach	0.011	0.104	0.000	0.000	0.000	0.000	1.000	0.000	1.000
Past_breach	0.029	0.168	0.000	0.026	0.158	0.000	0.328	0.470	0.000
Disclosure	0.364	0.481	0.000	0.360	0.480	0.000	0.699	0.459	1.000
Length	1.556	1.208	1.154	1.545	1.196	1.148	2.086	1.599	1.471
Size	6.439	2.307	6.586	6.408	2.293	6.559	9.238	1.774	9.161
LN_Segments	1.453	0.488	1.386	1.452	0.488	1.386	1.579	0.482	1.609
Age	21.676	14.668	17.000	21.595	14.641	17.000	28.921	15.251	26.000
Loss	0.413	0.492	0.000	0.415	0.493	0.000	0.187	0.391	0.000
LN_Analyst	1.356	1.191	1.386	1.348	1.186	1.386	2.039	1.432	2.565
Foreign	0.253	0.435	0.000	0.253	0.435	0.000	0.230	0.422	0.000
Merger	0.165	0.371	0.000	0.164	0.370	0.000	0.261	0.440	0.000
Growth	0.183	0.697	0.063	0.184	0.701	0.063	0.089	0.271	0.056
ICW	0.092	0.289	0.000	0.092	0.289	0.000	0.055	0.229	0.000

Note: This table reports descriptive statistics for the variables used in Equation (1). All variables are winsorized at 1 and 99 percent. All variables are defined in Appendix C.

Table 12. Descriptive Statistics (continued)

Panel B: Descriptive Statistics for Variables in Equation (2)

Variable	*Mean*	*Std*	*Median*
CAR	-0.002	0.026	-0.003
Disclosure	0.686	0.465	1.000
Length	2.160	1.745	1.602
Guidance	0.360	0.481	0.000
Size	9.216	1.964	9.248
Severity	0.398	0.490	0.000
Leverage	0.685	0.277	0.640
Btm	0.505	0.499	0.422
Loss	0.165	0.371	0.000

Note: This table reports descriptive statistics for the variables used in Equation (2). All variables are winsorized at 1 and 99 percent. All variables are defined in Appendix C.

Table 13. Correlations among Variables in Equation (1)

	Breach	*Past_breach*	*Disclosure*	*Length*	*Size*	*LN_Segments*	*Age*	*Loss*	*LN_Analyst*	*Foreign*	*Merger*	*Growth*	*ICW*
Breach	1.000												
Past_breach	**0.189**	1.000											
Disclosure	**0.074**	**0.151**	1.000										
Length	**0.065**	**0.131**	.	1.000									
Size	**0.128**	**0.206**	**0.247**	0.008	1.000								
LN_Segments	**0.027**	**0.041**	**0.057**	**-0.017**	**0.319**	1.000							
Age	**0.052**	**0.096**	**0.110**	**-0.054**	**0.329**	**0.269**	1.000						
Loss	**-0.048**	**-0.075**	**-0.126**	**-0.038**	**-0.404**	**-0.154**	**-0.228**	1.000					
LN_Analyst	**0.061**	**0.093**	**0.132**	**0.028**	**0.345**	**0.107**	**-0.116**	**-0.160**	1.000				
Foreign	-0.005	**-0.013**	**0.012**	**-0.030**	**0.106**	**0.349**	**0.063**	**-0.020**	**0.077**	1.000			
Merger	**0.027**	**0.051**	**0.149**	**0.044**	**0.158**	**0.160**	**0.046**	**-0.062**	**0.114**	**0.108**	1.000		
Growth	**-0.014**	**-0.031**	**-0.047**	-0.010	**-0.115**	**-0.121**	**-0.150**	**0.084**	0.002	**-0.038**	**0.013**	1.000	
ICW	**-0.013**	**-0.038**	**-0.064**	0.000	**-0.240**	**-0.072**	**-0.119**	**0.174**	**-0.150**	**-0.012**	**-0.038**	**0.076**	1.000

Note: This table presents correlations for all variables used in Equation (1) (two-sided). All variables are defined in Appendix C.

3.4 RESULTS

Main Findings

Panel A of Table 14 shows the results for testing H1. Consistent with my expectation, the coefficient of *Disclosure* is positive and significant (0.742, $p < 0.01$). The result suggests that firms with prior cybersecurity risk disclosures are more likely to experience subsequent cyber incidents. As for control variables, larger firms, firms with more analysts following, firms undergoing merger, firms with material weaknesses in internal controls, firms operating in consumer section, and firms with history of cyber incidents are more likely to have future cyber incident. Panel B of Table 14 presents the test results for H2. The coefficient of *Length* is 0.199 and is statistically significant, revealing that firms providing lengthy cybersecurity risk disclosure are more likely to experience subsequent cyber incidents. In untabulated test, I also explore alternative measures of *Length*. Specifically, *Length* is replaced with the log number of words in cybersecurity risk disclosure as well as the number of words in cybersecurity risk disclosure normalized by the total number of words in item 1A. Similar results are obtained using both measures. Overall, results reported in Table 14 suggest that both the presence and content of cybersecurity risk disclosures as measured by adjusted length are informative of future cyber incidents, providing support for the SEC’s intention to encourage cybersecurity risk disclosures.

Table 14. Logit Regression of Cybersecurity Risk Disclosure on Cyber Incidents

Independent variables	*Panel A*		*Panel B*	
	Estimates	*z-statistics*	*Estimates*	*z-statistics*
Disclosure	0.742	3.85***		
Length			0.199	4.13***
Past_breach	1.414	7.45***	1.337	6.90***
Size	0.611	11.49***	0.525	9.19***
LN_Segments	0.053	0.34	0.185	0.90
Age	-0.003	-0.65	-0.004	-0.75
Loss	-0.108	-0.68	-0.006	-0.03
LN_Analyst	0.104	2.01**	0.072	1.16
Foreign	-0.033	-0.20	0.061	0.35
Merger	0.247	1.62*	0.097	0.58
Growth	-0.125	-0.70	-0.063	-0.38
ICW	0.500	1.75**	0.080	0.19
Finance	-0.133	-0.65	-0.116	-0.50
Consumer	1.298	6.75***	1.205	5.33***
Intercept	-10.291	-22.12***	-9.168	-15.54
Year Effects	Included		Included	
Pseudo R Square	0.253		0.218	
# Observations	29,531		10,480	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels based on two-tailed *p*-values (one-tailed when predicted), respectively. Test statistics are based on robust standard errors clustered by firm. All variables are defined in Appendix C.

Table 15 shows the results of the multiple regression in equation (2). All the coefficients are multiplied by 100 for readability. The significant and positive coefficient on *Disclosure* in Panel A of Table 15 supports H3a, indicating that the market reaction following cyber incidents is less severe for firms with prior cybersecurity risk disclosures. In addition, the market responds more negatively for more severe cyber incidents and firms with loss. However, H3b is not supported. The coefficient on *Length* is not significant, suggesting that the market seems not using information conveyed by the content of cybersecurity risk disclosures. Taken together, results in Table 15 demonstrate that investors only care about the presence or absence of cybersecurity risk disclosure, but not the information content of the disclosure.

Table 15. Regression of Cybersecurity Risk Disclosures on Cumulative Abnormal Return

Independent variables	*Panel A*		*Panel B*	
	Estimates	*t-statistics*	*Estimates*	*t-statistics*
Disclosure	0.766	2.53***		
Length			-0.113	-1.04
Guidance	-0.034	-0.1	0.071	0.18
Market_cap	-0.025	-0.29	-0.034	-0.34
Severity	-0.443	-1.33*	-0.264	-0.58
Leverage	-0.172	-0.32	-0.691	-0.94
Btm	0.239	0.54	0.032	0.05
Loss	-0.609	-1.49*	-1.009	-2.18**
Intercept	2.469	1.25	3.092	1.91
Industry Effects	Included		Included	
R Square	0.198		0.224	
# Observations	389		267	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels based on two-tailed *p*-values (one-tailed when predicted), respectively. Test statistics are based on robust standard errors clustered by firm. Coefficients have been multiplied by 100 for readability. All variables are defined in Appendix C.

To examine H4, the sample is partitioned into a pre-guidance period group and a post-guidance period group and reexamine equation (1). Results are presented in Table 16. Panel A of Table 16 reveals that *Disclosure* is only significant in the pre-guidance period group, but not significant in the post-guidance period group. Panel B of Table 16 also shows that the coefficients of *Length* for both periods are positive and significant. Both the effect and significance of *Length* increase in the post-guidance period. These findings support the argument that the introduction of the SEC's cybersecurity disclosure guidance leads to disclosures by firms that do not have material cybersecurity risks. Furthermore, it is noticeable that both the magnitude and statistical significance are increased from the pre-guidance period to the post-guidance period for *Length*, suggesting that the content of cybersecurity risk disclosures is becoming more informative of cybersecurity risks in the post-guidance period.

Table 16. Logit Regression of Cybersecurity Risk Disclosure on Cyber Incidents by Period

Independent variables	*Panel A*				*Panel B*			
	Pre-Guidance		*Post-Guidance*		*Pre-Guidance*		*Post-Guidance*	
	Estimates	*z-statistics*	*Estimates*	*z-statistics*	*Estimates*	*z-statistics*	*Estimates*	*z-statistics*
Disclosure	0.891	4.63***	0.304	0.88				
Length					0.158	1.93**	0.225	3.77***
Past_breach	1.348	5.30***	1.539	6.29***	1.220	4.32***	1.453	5.83***
Size	0.671	9.99***	0.514	7.22***	0.579	7.04***	0.456	6.31***
LN_Segments	0.040	0.22	0.073	0.29	0.154	0.53	0.253	0.97
Age	-0.004	-0.62	-0.002	-0.33	-0.006	-0.91	-0.002	-0.29
Loss	-0.132	-0.62	-0.064	-0.24	0.043	0.16	-0.093	-0.31
LN_Analyst	0.108	1.73**	0.092	1.28	0.094	1.10	0.053	0.72
Foreign	0.082	0.41	-0.189	-0.80	0.235	0.91	-0.057	-0.23
Merger	0.225	1.00	0.268	1.31*	0.065	0.24	0.132	0.61
Growth	-0.066	-0.30	-0.332	-1.10	0.052	0.31	-0.268	-0.68
ICW	0.647	1.94**	0.092	0.17	-0.186	-0.30	0.282	0.49
Finance	-0.399	-1.56*	0.250	0.93	-0.396	-1.19	0.158	0.56
Consumer	1.106	4.74***	1.576	5.78***	0.923	3.03***	1.512	5.30***
Intercept	-10.668	-18.00***	-9.380	-13.28***	-9.209	-11.16***	-9.164	-11.44***
Year Effects	Included		Included		Included		Included	
Pseudo R Square	0.252		0.247		0.204		0.236	
# Observations	19546		9441		4561		5919	

Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels based on two-tailed p-values (one-tailed when predicted), respectively. Test statistics are based on robust standard errors clustered by firm. All variables are defined in Appendix C.

Additional Tests

Firm-specific Disclosure

In this section, I try to address the concern that more firm-specific cybersecurity risk disclosures could lead to more attacks. The SEC stated that "we *are mindful of potential concerns that detailed disclosures could compromise cybersecurity efforts -- for example, by providing a 'roadmap' for those who seek to infiltrate a registrant's network security -- and we emphasize that disclosures of that nature are not required under the federal securities laws*" (SEC, 2011). Ferraro (2013) criticizes the SEC's failure to address this issue and argues that any disclosure that is meaningful for investors is likely to contain information to hackers seeking future attacks. To test if the claim is valid, I use equation (1) but substitute *Cyber_dis* with two measures: *Score* and *Informativeness*. Both measures are based on the bag-of-words approach that represents documents as vectors with each dimension representing a unique word. The first measure, *Score*, is adapted from Brown and Tucker (2011), which is calculated as one minus the cosine similarity between a firm's disclosure and industry-year's average disclosure, adjusted by document length using Tayler expansion[10]. The variable captures how a firm's disclosure deviates from the industry average practice. The second measure, *Informativeness*, is calculated as the percentage of unique words that are not used by any other firms in the same industry for the same fiscal year. The variable represents how a firm's disclosure includes firm-specific information in terms of word usage. The regression results are presented in Table 17. Neither of these two measures is statistically significant. While the results do not invalidate

[10] Brown and Tucker (2011) analytically prove that the similarity score between two documents is a function of document length. Accordingly, they propose to use Tayler expansion to adjust the similarity score.

the concern, they seem to suggest that the level of firm-specific information in item 1A of 10-K is not informative enough to jeopardize a firm's cyber endeavor.

Table 17. Logit Regression of Firm-specific Disclosure on Cyber Incidents

Independent variables	*Panel A*		*Panel B*	
	Estimates	*z-statistics*	*Estimates*	*z-statistics*
Score	0.336	0.43		
Informativeness			1.215	1.24
Past_breach	1.400	7.17***	1.415	6.98***
Size	0.533	8.87***	0.525	8.16***
LN_Segments	0.164	0.77	0.138	0.64
Age	-0.004	-0.72	-0.002	-0.32
Loss	0.033	0.18	-0.009	-0.04
LN_Analyst	0.076	1.19	0.069	1.03
Foreign	0.122	0.64	0.148	0.77
Merger	0.099	0.57	0.105	0.58
Growth	-0.057	-0.35	-0.059	-0.33
ICW	0.090	0.21	0.141	0.32
Finance	-0.040	-0.16	-0.083	-0.32
Consumer	1.345	5.65***	1.494	6.01***
Intercept	-8.952	-14.84***	-9.249	-13.64***
Year Effects	Included		Included	
Pseudo R Square	0.202		0.216	
# Observations	10207		9295	

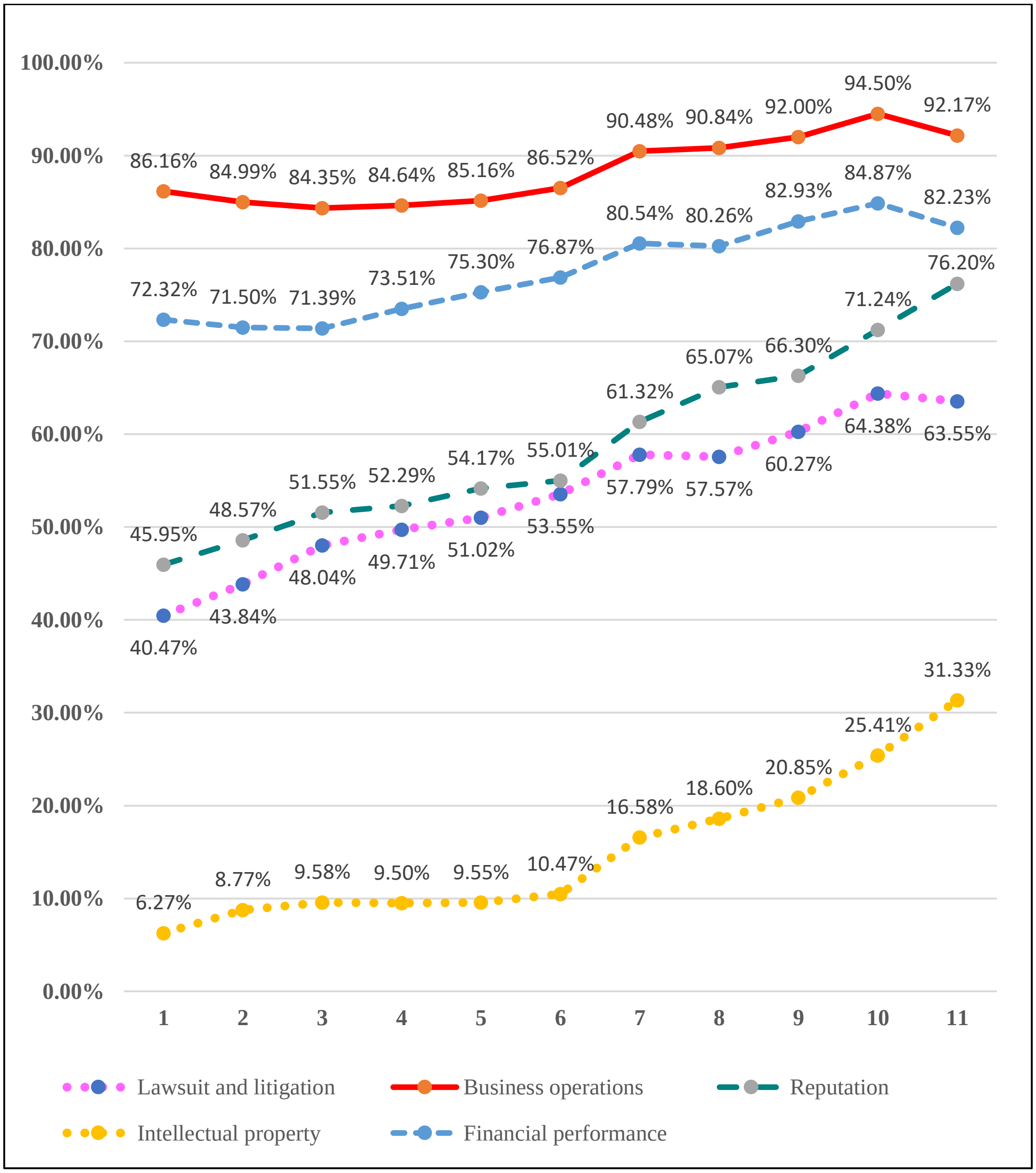
Note: *, **, *** represent significance at the 0.10, 0.05, and 0.01 levels based on two-tailed *p*-values (one-tailed when predicted), respectively. Test statistics are based on robust standard errors clustered by firm. All variables are defined in Appendix C.

Topic Analysis

To further understand cybersecurity risk disclosure, a topic analysis is conducted to investigate firm's concerns about cybersecurity. Specifically, all two-word phrases that occur in at least 2% but no more than 98% of all cyber disclosures are extracted, which gives me 1,042 phrases in total [11]. I manually read these phrases, choose 211 phrases that are meaningful, and classify these phrases into five topics of consequences: business operations, financial performance, reputation, lawsuit and litigation, and intellectual property. Appendix E lists the phrases used for classification. Figure 2 shows the percentage of firms that mention each type of risk across years. The figure offers two important findings. First, the disruption of business operations is the biggest concern regarding cybersecurity. More than 85% of the firms disclose the potential impact of cyber incidents on business operations, and the rate remains relatively stable over time. Impact on financial performance is the second biggest concern, with more than 70% of the firms mentioning this topic. Second, while intellectual property is the least mentioned topic, I observe a significant jump in the recent years. Similarly, concerns over reputation are steadily increasing over years, which is consistent with the public perception that cybersecurity is attracting greater attention in recent years.

[11] All the words are stemmed, and stop words are removed. Phrases that consist of two words are used to increase the interpretability of the outcome. In addition, I use 2% as the threshold to get rid of specific phrases such as firm names as well as 98% threshold to filter out uninformative phrases that are used by all disclosures. The results do not change when these parameters are varied.

Figure 2. Percentage of Firms Disclosing Different Topics Across Years



3.5 CONCLUDING REMARKS

In this chapter, I examine whether cybersecurity risk disclosure is informative for future cyber incidents. Results are summarized in Table 18. Specifically, I focus on two measures: the presence of cybersecurity risk disclosure and the length of cybersecurity risk disclosure. Consistent with my expectation, both the presence and length of cybersecurity risk disclosure are positively associated with subsequent cyber incidents, suggesting that cybersecurity risk disclosure is not boilerplate. In addition, I test whether the market participants are using information in cybersecurity risk disclosure. The results demonstrate that investors are only using information conveyed by the presence of, but not the content of cybersecurity risk disclosure. Furthermore, there is a differential effect before and after the SEC's cybersecurity disclosure guidance. The presence of cybersecurity risk disclosure is no longer associated with subsequent cyber incidents, revealing that the SEC's emphasis on cybersecurity risk disclosures results in more disclosures by firms not having material cybersecurity risks. I fail to find a significant association between firm-specific disclosure and cyber incidents, providing some relief for the regulator's concern that more firm-specific disclosure may provide information for hackers. Finally, the topic analysis indicates that firms are more concerned about business operations and financial performance when encountering cybersecurity issues. Moreover, there is a growing concern regarding reputation damage and loss of intellectual property due to cyber incidents. Collectively, results in this chapter should be valuable to practitioners, regulators, and academics who are interested in the informativeness of cybersecurity risk disclosures. I stand with the SEC to emphasize the importance of cybersecurity risk disclosure, but raise the question for the unintended consequence result from cybersecurity disclosure guidance.

Table 18. Summary of Findings

Hypothesis	Supported?
The presence of cybersecurity risk disclosure is positively associated with the likelihood of subsequent cyber incident.	YES
The length of cybersecurity risk disclosure is positively associated with the likelihood of subsequently reported cyber incident.	YES
The market reaction following cyber incident is less severe for firms with prior cybersecurity risk disclosure.	YES
The market reaction following cyber incident is less severe for firms with lengthy cybersecurity risk disclosure.	NO
The association between the presence of cybersecurity risk disclosure and subsequent cyber incident is different before and after the introduction of the SEC's cybersecurity disclosure guidance.	YES

There are several limitations in this study. I maintain the assumption that managers have knowledge of the cybersecurity risks firms face, which may not necessarily hold. If firms are not aware of the level of cyber threats, they are less likely to provide meaningful disclosures. In addition, cyber incidents are used as the proxy for cybersecurity risks, which may not be the most accurate measure as theoretically any system can be breached. Future study may benefit by using information at a more disaggregated level, such as data from intrusion detection system (IDS). Further, this paper did not answer the question why investors are not utilizing information conveyed in the length of cybersecurity risk disclosure. There could be at least two explanations. Market participants may be unaware of the informativeness of the content thus not pricing the information into stock price. Alternatively, investors may recognize such information, but believe that firms providing lengthy cybersecurity risk disclosures are more likely to invest heavily to address cybersecurity risks, reducing the probability of future cyber incidents. Future research is needed to explore this issue.

CHAPTER 4: CYBERSECURITY ASSURANCE AND DATA ANALYTICS

4.1 INTRODUCTION

Cybersecurity has attracted a great deal of attention in recent years. The number of cyber incidents detected is continuously increasing, and top managements show concern that cyber threats could hinder the growth of their firms (Loop, 2016). Regulators also have displayed concerns about cybersecurity issues and potential implications for financial reporting and auditing. In this chapter, I discuss the application of data analytics in cybersecurity assurance that has not received much attention in the literature but is of great interest to the audit profession.

Cybersecurity can be defined as "the processes and controls implemented by a firm to manage cybersecurity risks" (AICPA, 2017c). Due to the ever-increasing reliance on information systems and the Internet, information assets become one of firm's most valuable resources (Gordon et al., 2010). Recent high-profile cybercrimes, such as the Yahoo data breach and the Wannacry Ransomware attack, have re-emphasized the importance of cybersecurity, which has become fundamental to organizations' IT mission (Debreceny, 2013). In 2015, the number of reported security incidents increased 38 percent from the previous year, with a substantial surge of incidents involving intellectual property and business plans (PWC, 2016). These numbers tend to be underestimated since firms may hesitate to report such information due to concerns over negative public image (D'Arcy, Hovav, & Galletta, 2009). The growing threats against cybersecurity have prompted the Securities and Exchange Commission (SEC) to hold a roundtable discussion to deliberate about the cybersecurity landscape and related issues (SEC, 2014). In the same year, the Standing Advisory Group of the Public Company Accounting Oversight Board

(PCAOB) also assembled a panel discussion on cybersecurity's potential implications for financial reporting and auditing (PCAOB, 2014).

In addition to the regulatory emphasis, there is growing interest in cybersecurity reporting. The cybersecurity disclosure guidance issued by the SEC in 2011 is a snapshot that reporting on cybersecurity matters is top on regulator's agenda (SEC, 2011). Recently, a new bill called "Cybersecurity Disclosure Act of 2017" was introduced in the U.S. Senate, requiring firms to disclose the level of cybersecurity expertise of their board members and to describe what other cybersecurity steps have been taken by the firms (Reed, 2017). Realizing that there is no consistent and common language for describing cybersecurity risk management programs, the AICPA Assurance Services Executive Committee (ASEC) has developed a cybersecurity risk management reporting framework for firms to communicate information regarding cybersecurity risk management efforts and for practitioners to examine and report on the management-prepared cybersecurity information. Along with the reporting framework, the AICPA' ASEC Cybersecurity Working Group, in conjunction with the Auditing Standards Board (ASB), introduced an attestation guide titled *Reporting on an Entity's Cybersecurity Risk Management Program and Controls* to assist practitioners to opine on the cybersecurity risk management report. Although the attestation guide points out detailed requirements for practitioners at different stages in a cybersecurity assurance engagement, it contains limited guidance on how to systematically evaluate cybersecurity risks in the engagement, how to collect evidence pertaining to specific risks, and how to use the evidence in assessing the risks. This chapter makes an attempt to address the above issues from a data analytics perspective. Although data analytics has been used in the audit of financial statements, its applicability in cybersecurity assurance has not yet been established. Given that these two engagements differ in nature and objectives, it is essential to discuss and evaluate the usefulness of data analytics in the new engagement. Therefore, this essay first analyzes the potential benefits of using data analytics in cybersecurity assurance to demonstrate that analytics should be an integral part of the engagement. A seven-step process of using data analytics in testing cybersecurity controls is then introduced, pointing out considerations that practitioners may need to have in the engagement. An illustrative example of the process using synthetic insider threat data is presented to show the usefulness of data analytics. Finally, critical issues related to the use of data analytics in cybersecurity assurance are outlined.

This essay is organized as follows. The next section provides a brief overview of related literature and the AICPA's reporting framework. The third section argues for the use of data analytics in cybersecurity assurance while the fourth section discusses critical issues. The last section concludes this essay.

4.2 BACKGROUND AND RELATED LITERATURE

Overview of the AICPA's Cybersecurity Reporting Framework

The AICPA recently announced an entity-level cybersecurity reporting framework for firms to meet the needs of a broad range of stakeholders by communicating useful information about their cybersecurity risk management efforts (Tysiac, 2017). The framework is voluntary and consists of three components: a management description, a management's assertion, and a practitioner's opinion. The first component is a narrative description of a firm's cybersecurity risk management program that provides a basis for understanding the way in which the firm manages its cybersecurity risks and the controls the firm implements in response to those risks. The second component is an assertion made by management about whether the management description was presented in accordance with the selected description criteria and whether the controls implemented as part of the program were effective to achieve the firm's cybersecurity objectives against the selected control criteria. The last component is a practitioner's opinion on the fair presentation of management's description and the suitability of the design and, if applicable, the effectiveness of controls implemented in the program (AICPA, 2017b). In addition, two sets of related criteria have been published in conjunction with the reporting framework: description criteria and control criteria. Description criteria are designed for management to explain the firm's cybersecurity risk management program in a consistent manner, while control criteria are intended for management to develop cybersecurity controls and for practitioners to opine on the effectiveness of the controls implemented in the program (Tysiac, 2017).

The reporting framework is flexible enough to allow management to select description criteria and control criteria as long as they are considered suitable under *Concepts Common to All Attestation Engagements* (AT-C 105), and exhibit all four characteristics: relevance, objectivity, measurability, and completeness. However, the reporting framework encourages the use of the AICPA proposed criteria to promote consistency and comparability of information provided by different firms.[1] There are two distinct subject matters in the cybersecurity risk management examination: the description of a firm's cybersecurity risk management program and the design and operating effectiveness of the controls for achieving the firm's cybersecurity objectives.

[1] The AICPA's ASEC integrates information from various sources such as Control Objectives for Information and Related Technologies (COBIT) 5, National Institute of Standards and Technology (NIST) Cybersecurity Framework, and Committee of Sponsoring Organizations of the Treadway Commission (COSO)'s 2013 Internal Control - Integrated Framework in the criteria development process.

Accordingly, the management makes two main assertions: description assertion (i.e., the description was presented in accordance with the description criteria) and control assertion (i.e., the controls implemented as part of the firm's cybersecurity risk management program were suitably designed, and if applicable to the engagement, effective to achieve its cybersecurity objectives based on a specified set of suitable control criteria) (AICPA, 2017c). While the practitioner needs to evaluate both assertions, the discussion presented in this chapter exclusively focuses on the control assertion rather than the description assertion for two reasons. First, how to audit qualitative disclosure has already been extensively discussed in the domain of Corporate Social Responsibility (e.g., Cohen & Simnett, 2014; O'Dwyer, 2011; Wallage, 2000). Second, the fair representation of the management description is partly contingent on the evaluation of the control assertion because firms discuss cybersecurity control processes in the description.

Literature Review

Numerous prior studies have examined cybersecurity issues. The first stream of literature that is related to this essay is about cybersecurity disclosure. Early studies examined the market reaction following the firm's voluntary disclosure of cybersecurity matters. For example, Gordon et al. (2010) demonstrated that voluntary disclosure concerning proactive security measures has the greatest positive impact on a firm's stock price, followed by voluntary disclosure of cybersecurity vulnerabilities. A subsequent study by Wang, Kannan, et al. (2013) documented that firms disclosing cybersecurity risk factors in their annual reports with actionable information are less likely to be associated with subsequent security incidents. By contrast, firms that only disclose cybersecurity risk but reveal no actionable plans are more likely to have security incidents in the future, and are punished more severely when the actual security incidents happen. After the SEC issued the cybersecurity disclosure guidance in 2011 which highlights sections that may be relevant for cybersecurity disclosure, researchers started to focus on the informativeness of cybersecurity disclosure. Hilary et al. (2017) failed to find a significant association between firm's prior cybersecurity risk disclosure and the market reaction following the security incidents, claiming that cybersecurity risk disclosures are not informative for investors. However, a contemporary paper by Li, No, Wang, and Vasarhelyi (2017) showed the opposite result, suggesting that both the presence and the length of cybersecurity risk disclosure are predictive of future security incidents. The previous research addressed disclosures that were not assured by independent practitioners. The AICPA framework may lead to the issuance of assured cybersecurity reports and the impact of such reports would be of future research interest. This chapter contributes to the development of this stream of research by focusing on the assurance process of the cybersecurity report.

Data analytics is the process of generating insights from financial, operational, and other forms of electronic data internal or external to a firm (Schneider, Dai, Janvrin, Ajayi, & Raschke, 2015; Wang & Cuthbertson, 2015). The exponential increase of structured and unstructured data motivates firms to use available information to extract knowledge and generate business values. By effectively leveraging data analytics in business, firms can achieve about five percent productivity gains (Brynjolfsson, Hammerbacher, & Stevens, 2011; Warren Jr, Moffitt, & Byrnes, 2015). Chen, Chiang, and Storey (2012) identified five areas that are most likely to benefit from using data analytics: e-commerce and market intelligence, e-government and politics, science and technology, health and well-being, and security and public safety. For using data analytics in auditing, prior studies examined the usefulness of data analytics in assessing fraud risk (Lin, Hwang, & Becker, 2003), weakness in internal controls with respect to segregation of duties (Jans, Alles, & Vasarhelyi, 2014), and anomalies in a business process (Kogan, Alles, Vasarhelyi, & Wu, 2014). In addition, some studies, by emphasizing the potential of data analytics, introduced a list of research opportunities pertaining to data analytics for the accounting and auditing profession (e.g., Schneider et al., 2015; Wang & Cuthbertson, 2015) while several other studies discussed how analytics can change accounting and auditing domain in the context of big data (e.g., Brown-Liburd, Issa, & Lombardi, 2015; Cao, Chychyla, & Stewart, 2015; Vasarhelyi, Kogan, & Tuttle, 2015; Warren Jr et al., 2015; Yoon, Hoogduin, & Zhang, 2015; Zhang, Yang, & Appelbaum, 2015). A comprehensive discussion of data analytics relating to management accounting and auditing can be found at Appelbaum, Kogan, Vasarhelyi, and Yan (2017) and Appelbaum, Kogan, and Vasarhelyi (2017), respectively. This chapter is different from this stream of literature in that it introduces data analytics into cybersecurity, an emerging assurance field – cybersecurity - that is attracting attention in the audit profession, but has not been examined previously.

4.3 DATA ANALYTICS AND CYBERSECURITY ASSURANCE

Necessity

Data analytics refers to using various analytical techniques and explanatory and predictive models to analyze structured and unstructured data and to provide valuable information for users to make informed decisions (Schneider et al., 2015). The availability of inexpensive computing power and ever-increasing storage capacity allow data analytics to be applied to many fields. Given the nature of the cybersecurity control activities, data analytics should be an integral part of the evidence collection process in a cybersecurity assurance engagement for the following reasons.

First, similar to the traditional audit where the decision whether to use sampling depends on “the cost and time required to examine all the data and the adverse consequences of possible erroneous decision based on the conclusions resulting from examining only a sample of the data” (PCAOB, 2011), practitioners in a cybersecurity assurance engagement need to consider the cost and consequence of using sampling. The cost of data collection and analysis is quite low because of the digitization of information and the availability of analytic techniques (Krahel & Titera, 2015). For regulatory compliance and forensic analysis, many firms have already collected security-related data, such as system logs, active directory, and network events (Cardenas, Manadhata, & Rajan, 2013). In addition, some components of the system of internal controls use data stored in electronic form for configuration, operation, and monitoring of controls, resulting in such data being readily available for use. By contrast, use of sampling in cybersecurity assessment may not provide practitioners with sufficient, appropriate evidence to reduce risk to an acceptable level. For certain cybersecurity controls, even a one-time failure of the control could be an indicator of an ineffective system of internal control related to cybersecurity which permits material security incidents to occur. Because data analytics enables the analysis of the whole population accumulated over the period under examination (Wang & Cuthbertson, 2015), practitioners should rely on data analytics to analyze 100 percent of the data when it is possible [2].

Second, using data analytics can improve efficiency. In a cybersecurity assurance

[2] When a security incident has not yet occurred, practitioners may not identify any suspicious item even if they are using data analytics. However, a downside of examining 100 percent of the population is that practitioners may obtain too much confidence in firms’ controls, potentially influencing their professional skepticism in subsequent procedures.

engagement, practitioners may need to examine thousands of records and documents, such as service-level agreements and contracts with business partners, for understanding risks arising from interaction with third parties and identifying required controls. In addition, such documents and records are often the original sources of evidence that must be evaluated. However, the sheer amount of the documents together with a large number of pages in each document makes manual inspection highly inefficient. Given that such documents are highly standardized, this problem can be resolved using a data analytic technique called text mining. Practitioners can presumably use keyword search to identify deficiencies or exceptions in contracts, or use similarity analysis to identify the deviation of a document from other similar documents to highlight potential issues (Yan, Moffitt, & Vasarhelyi, 2017). Likewise, using motion detection techniques such as the ones described in Konrad (2000) can save a tremendous amount of efforts when examining video feeds to identify unauthorized access to restricted areas.

Third, data analytics can offer an independent view of a firm's control system that is otherwise not available from evidence collected through inquiry, inspection, observation, and walkthroughs. Because many controls in cybersecurity context are implemented using computer programs that automatically create logs with a tremendous amount of relevant information, analyzing such data that are generated independent of the employees who are performing the procedures can provide more reliable evidence about whether the controls were implemented and properly operated. For example, company insiders who are entitled to certain level of access may not always comply with the entity's cybersecurity policies and controls and may abuse their privilege (Vance, Lowry, & Eggett, 2013). To the extent that the insiders are the subject of an inquiry and have incentives to conceal their noncompliant behaviors, practitioners may consider evidence gathered through inquiry to be insufficient. Instead, sufficient and relevant evidence may be collected by analyzing insiders' activities recorded in system logs.

Fourth, many technology-based automated controls use data analytics to prevent and detect security events. For example, to deal with potential data breaches, firms employ data loss prevention software, which uses both supervised and unsupervised learning to identify potential incidents with network data (Shabtai, Elovici, & Rokach, 2012).[3] While security software and systems developed commercially tend to function in a consistent manner, practitioners may consider evaluating the design and functioning of the controls by analyzing the same data using the same or different data analytics and perform cross-validation.

Process of Using Data Analytics in Cybersecurity Assurance

This subsection discusses the process of using data analytics in testing cybersecurity controls, which is adapted from the forthcoming AICPA Audit Data Analytics Guide[4]. The process consists of seven stages that are visualized in Figure 3:

1:) determine assertion to be examined,

2:) identify controls implemented by the management that support the assertion,

[3] Machine learning algorithms can be broadly categorized into supervised and unsupervised learning. Supervised learning is the task of inferring a function from labeled training data which consist of a set of training examples. For instance, if all historical data about system access are correctly labeled as valid or invalid, based on the historical data, supervised machine learning algorithms can automatically learn the difference between the two types of data and detect invalid access in the future. Unsupervised learning is the task of inferring a function to describe hidden structure from unlabeled data. For example, when there is no prior knowledge about data, unsupervised machine learning may help to identify hidden groups or patterns based on the data characteristics.

[4] For more information, see at https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/pages/auditdataanalyticsguide.aspx.

3:) determine if data analytics can be used to evaluate the control,

4:) develop procedure that involves the use of data analytics,

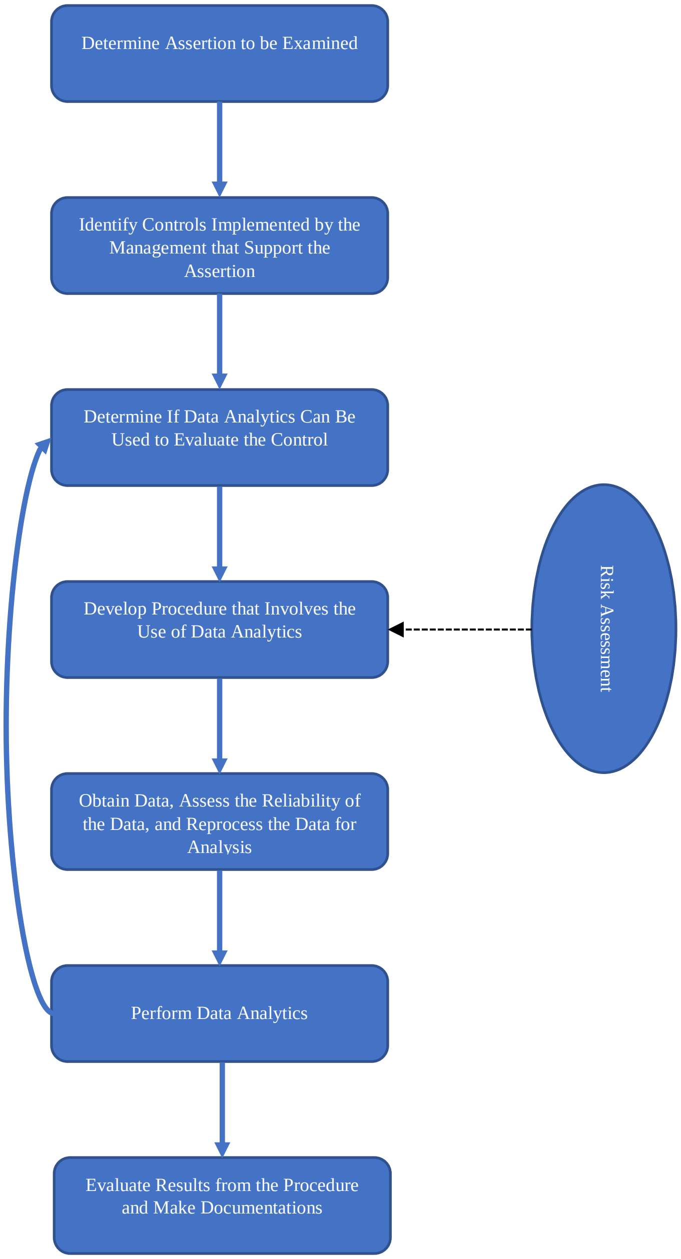
5:) obtain data, assess the reliability of the data, and reprocess the data for analysis,

6:) perform data analytics,

7:) evaluate results from the procedure and make documentations.

Detailed considerations that practitioners should have at each stage are discussed below.

Figure 3. Process of Using Data Analytics


Determine Assertion to be Examined
Identify Controls Implemented by the Management that Support the Assertion
Determine If Data Analytics Can Be Used to Evaluate the Control
Develop Procedure that Involves the Use of Data Analytics
Obtain Data, Assess the Reliability of the Data, and Reprocess the Data for Analysis
Perform Data Analytics
Evaluate Results from the Procedure and Make Documentations
Risk Assessment

Stage 1: Determine Assertion to be Examined

In the first stage, the practitioner selects the assertions to be tested. If the engagement is to examine the operating effectiveness of cybersecurity controls, the main assertion is that "the controls implemented as part of the firm's cybersecurity risk management program were effective to achieve its cybersecurity objectives based on a specified set of suitable control criteria" (AICPA, 2017c). Although a firm can tailor its cybersecurity objectives to reflect its business objective, four main objectives are shared across all firms: availability, confidentiality, information integrity, and processing integrity. Accordingly, the main assertion can be further divided into four sub-assertions. Table 19 summarized these sub-assertions with illustrative controls for each sub-assertion.[5] In the engagement, the practitioner, at the very least, should obtain sufficient evidence as to whether each relevant sub-assertion can be supported. Other assertions that only apply to some firms can also be determined and examined using the same process.

[5] Illustrative controls were identified from Trust Services Criteria, available at https://www.aicpastore.com/InternalControls/trust-services-principles-and-criteria/PRDOVR~PC-TSPC13/PC-TSPC13.jsp.

Table 19. Assertions and Illustrative Controls

Main Assertion	*Sub-Assertion*	*Illustrative Control*
The controls within an entity's cybersecurity risk management program was effective to achieve the entity's cybersecurity objectives based on the control criteria.	Controls were effective to enable timely, reliable, and continuous access to and use of information and systems (availability).	Future processing demand is forecasted and compared to existing capacity on a daily basis.
	Controls were effective to protect information from unauthorized access and disclosure (confidentiality).	Access to data is restricted to authorized applications through access control software.
	Controls were effective to guard against improper information modification or destruction of information (information integrity).	Weekly full-system and daily incremental backups of information are performed using an automated system.
	Controls were effective to guard against improper use, modification, or destruction of systems (processing integrity).	Systems backups are transported and stored offsite by a third-party storage provider.

Stage 2: Identify Controls Implemented by the Management that Support the Assertion

After determining assertion to be examined, the practitioner identifies controls implemented by the firm pertaining to each assertion for subsequent examination. Different controls can be implemented to achieve the same cybersecurity objective. For example, smaller and less complex firms may address the risk that threatens the achievement of firm's cybersecurity objectives using fewer and less sophisticated controls than larger and multinational firms (AICPA, 2017c).

Stage 3: Determine If Data Analytics Can Be Used to Evaluate the Control

In the next stage, the practitioner should develop procedures to assess the suitability and operating effectiveness of controls identified in the previous stage. These procedures include, but not be limited to, inquiry of employees, walk-throughs, inspection of files, and reperformance of controls. The nature, timing, and extent of procedures to be performed should be contingent on the risk assessment procedures in the planning stage. If the risk associated with the assertion is high, the practitioner will need to respond by revising the timing of the procedure, modifying how the controls are tested, increasing the number of procedures to be performed, or expanding the sample size. It is notable that risk assessment in cybersecurity assurance is conceptually different from that in financial statement assurance. In the audit of financial statements, the auditors rely on the audit risk model[6] as a conceptual framework to analyze risks and determine the extent of testing[7]. However, the same model cannot be directly applied to cybersecurity assurance because the two assurances differ in objectives. The objective of auditors in a financial statement audit is to ensure the fair presentation of financial statements, whereas the objective of the practitioner in a cybersecurity examination engagement is to ensure the suitability and effectiveness of controls related to management's description of its cybersecurity risk management program. In a cybersecurity assurance engagement, the fact that there has been a security incident does not necessarily mean that cybersecurity controls are not effective because any firm could experience security incidents at some point of time even if the level of controls is high (AICPA, 2017c; Mitra & Ransbotham, 2015). By contrast, there could be a material weakness in a firm's cybersecurity controls even if there is no security incident[8]. Thus, it is necessary for the practitioner to use a conceptual model that is relevant to such

[6] For more information, see AU Section 312: Audit Risk and Materiality in Conducting an Audit.

[7] The ASB also introduced an attestation risk model that generalized the audit risk model for use when the subject matter of the engagement is something other than financial statements. See AT-C 105.

[8] Determining materiality is a significant challenge in cybersecurity assurance because there is no dollar amount. The issue is discussed in detail in the fourth section.

circumstances. For example, the internal control risk model proposed by Akresh (2010) can be used as the starting point for risk assessment in cybersecurity assurance[9].

Some procedures can only be performed in non-analytical ways (e.g. observation of the application of the control). On the other hand, data analytics can be highly effective in performing some procedures to obtain evidence. When determining if data analytics should be used in the procedures, the practitioner needs to consider what type of data can be analyzed. Generally, three types of data can be highly effective in cybersecurity assurance. First, evidence about the potential occurrence of security events can serve as indicators that controls may not be effective to achieve a firm's cybersecurity objectives. Although cybersecurity assurance is the audit of a firm's control process, the outcome from the control process can reflect deficiencies in the controls, pointing out potential areas that require further examination. Such data could come from internal sources such as monitoring tools over user activities or system performance, or external sources such as social media. For example, by analyzing data from system downtime monitoring tools, the practitioner can gain an understanding of the assertion pertaining to controls over the availability of information and systems. As an alternative example, social media data can be extracted, parsed, and analyzed to infer if a firm fails to deliver timely service or has customers who experienced identity theft.

Second, System data, especially metadata, can reveal the operations of controls in practice. Information systems are capable of automatically generating valuable metadata. Metadata is data that describes the underlying data. One unique advantage of metadata in

[9] Akresh (2010) argues that the audit risk model is not suitable for audits of internal control. To address the issue, the author developed an internal control risk model that focuses on three elements: inherent risk, control design and implementation risk, and control operating effectiveness risk.

assurance service is its independence. Because metadata is generated independent of the people performing the control procedures, it provides more reliable evidence for evaluating cybersecurity controls. Jans et al. (2014) analyzed event logs that contain both data entered by the auditee and metadata recorded by an ERP system. Using process mining, they identified numerous instances that violate established policies, including payments without approval and violation of segregation of duties, that are not detected by internal auditors using conventional audit procedures. Because many controls in cybersecurity context are implemented using computer programs, performing analytics on metadata is preferable to other types of procedures for testing the operating effectiveness of such controls.

Third, many automated controls use heuristics and machine learning techniques to identify suspicious network traffic patterns (Cardenas et al., 2013). Although it is unlikely that the practitioner will examine data at such granularity level, network data, such as packet data, can be analyzed using both supervised and unsupervised methods to identify anomalies that should be compared against alerts generated by control tools implemented by the firm. Any mismatch could be indicative of a potential control problem. It is possible that a firm under examination does not record or store all the data due to cost or privacy concerns. Thus, the practitioner should determine what data is recorded at an early stage of the engagement. If data is not available for certain procedures, the practitioner should resort to other procedures to collect evidence and consider the implications of the absence of such data.

Stage 4: Develop Procedure that Involves the Use of Data Analytics

The next stage is to define the specific objective of the data analytics procedure and select the appropriate techniques. Data analytics can be broadly categorized into three types: descriptive, predictive, and prescriptive (Appelbaum, Kogan, & Vasarhelyi, 2017). Descriptive analytics describes what has happened using techniques such as descriptive statistics, visualization, and text mining. Predictive analytics, on the other hand, demonstrates what could happen (IBM, 2013) by transforming historical data into knowledge to predict future events using a predictive or probability model. Prescriptive analytics shows what should happen using an optimization approach (Appelbaum, Kogan, Vasarhelyi, et al., 2017). Descriptive analytics and predictive analytics are more relevant techniques in cybersecurity assurance because descriptive analytics demonstrates potential control deviations while predictive analytics reveals potential threats. For descriptive analytics, descriptive statistics, visualization, and clustering would presumably be most effective because these techniques enable the practitioner to narrow down to the areas that are most likely to have control deficiencies. For predictive analytics, supervised methods, such as decision tree and time-series regression, are generally more applicable because rich historical information is accumulated over time for analysis, allowing the practitioner to build benchmarks for identifying high-risk area and examine if firms have addressed the identified issues. A summary of data analytics techniques can be found in Appelbaum, Kogan, Vasarhelyi, et al. (2017).

Stage 5: Obtain Data, Assess the Reliability of the Data, and Reprocess the Data for Analysis

At the fifth stage, data are obtained and validated. Two characteristics of data must be evaluated: accuracy and completeness (AICPA, 2017a). Inaccurate data may result in unreliable evidence while incomplete data may fail to produce sufficient evidence. The practitioner can use simple statistics to evaluate the reliability of data. For example, if gaps in system log data are identified (i.e., there is no log information at all for certain days), the data is unlikely to be complete. In the case that data were considered inaccurate or incomplete, the practitioner needs to consider the implications for firm’s cybersecurity controls as some controls implemented by the firm may also rely on the same data. Separate procedures that examine the scripts used to extract the data may be performed.

Significant efforts may be required to prepare data for analytics since analytics in cybersecurity assurance usually relies on data that are in various formats, capturing activities of multiple entities at different points of information systems. Data transformation is necessary to convert data in such a way that data analytics procedure can be efficiently performed to achieve the objective. In particular, the degree of aggregation should be carefully selected. Since such data usually captures information at the most disaggregated level (e.g. in terms of time interval, activities are recorded every millisecond), the practitioner will need to make the choice of the aggregation level. On the one hand, anomalies or control deviations may go undetected if data is aggregated over a longer period. On the other hand, too many false positives may be generated that consume the practitioner’s limited resource if data is analyzed at the most disaggregated level. As there is still extensive debate in the profession as to what extent data should be aggregated (Kogan et al., 2014), the practitioner need to exercise professional judgment regarding this matter.

Stage 6: Perform Data Analytics

The process of performing data analytics is iterative in nature and can be characterized by hypothesis generation and testing. When the initial results from performing data analytics indicate control deviations, the practitioner should continue to explore the data by generating possible explanations for the deviations. These explanations are further tested by reperforming analytics, usually on a subgroup of the data that is generated by grouping or filtering. If a hypothesis derived from the initial results is supported, the practitioner will focus on the cases that cannot be explained by the hypothesis. If the initial hypothesis is not supported, the practitioner may need to generate an alternative hypothesis or take all deviations as potentially problematic cases. The iterative process continues until the practitioner has determined that no additional hypothesis can be developed or tested in the procedure.

Stage 7: Evaluate Results from the Procedure and Make Documentations

In the final stage, evidence is evaluated to determine whether the objective of the procedure is achieved. The iterative application of data analytics in the previous stage either confirms or disconfirms the practitioner's expectation. The practitioner should decide if additional procedures are needed depending on the assessment of the evidence. In such case, another round of examination starting from stage 3 should be conducted by developing new procedures. If the objective of the data analytics procedure is achieved, the practitioner should draw conclusions from the procedure, make proper documentation, and respond to the assessed control deviations. It is important to note that while the presence of control deviations may indicate that there are control deficiencies or material weaknesses, the absence of deviations alone may not be enough to confirm that the control is operating effectively. Evidence obtained from other procedures pertaining to the same assertion should be evaluated together with the evidence derived from the data analytics procedure.

Example Application of the Process

To demonstrate the applicability of data analytics in cybersecurity assurance, I provide illustrative examples in this subsection following the proposed process. Synthetic insider threat test datasets obtained from the Computer Emergency Response Team (CERT) Division of the Carnegie Mellon University were used for the demonstration.[10] Realizing the issue that the paucity of security-related data may impede research on cybersecurity, the CERT Division, in partnership with ExactData, LLC (www.exactdata.com), generates a synthetic collection of logs from sensors that monitor the performance of all the computer workstations in an organization with 1,000 employees over a 500-day period (Glasser & Lindauer, 2013). Using synthetic data that are publicly available is preferable for my illustration as it overcomes the difficulty of obtaining real-world data and the confidentiality and privacy issues that may be associated with such data[11]. The dataset contains rich information regarding users' activity, including computers logon and logoff activity, external device connect and disconnect activity, files sent to external devices, emails exchanged during the day, and websites browsed by the user. In addition, the data contains previously identified suspicious activities and an employee list including each employee's system account, role, department, supervisor, and other descriptive information.

Stage 1: determine assertion to be examined.

The auditor decides to examine the assertion that "controls were effective to protect

[10] For more information, see at https://www.cert.org/insider-threat/tools/index.cfm.

[11] The downside of using synthetic data is that the data is realistic in only a limited number of dimensions that are specifically controlled by the developer. The lack of realism may result in failure to identify interesting and meaningful patterns when conducting analysis. However, it is not a significant concern in the current context because this study is only interested in demonstrating the usefulness of data analytics, rather than applying data analytics to extract insight.

information from unauthorized access".

Stage 2: identify controls implemented by the management that support the assertion.

The auditor identifies a set of controls that have been implemented by the management to support the assertion. Particularly, the management develops and implements controls by referring to the Trust Services Criteria for Security, Availability, Processing Integrity, Confidentiality, and Privacy of the AICPA.

Stage 3: determine if data analytics can be used to evaluate the control.

The auditor decides to evaluate the suitability of design and operating effectiveness of two specific controls: the firm constantly reviews user credentials and removes user access when an individual no longer requires such access, and the firm actively monitors system components and the operation of those components for anomalies that are indicative of malicious acts (CC7.2, AICPA, 2017c). By examining the inventory of data that is available in the firm, the auditor determines that data analytics can be used in procedures to test these controls. The auditor concludes that log data for users' logon and logoff activity and the employee file can be extracted and integrated to examine the operating effectiveness of the first control, while several sources of data that captures account users' activities can be used to examine the second control.

Stage 4: develop procedures that involve the use of data analytics.

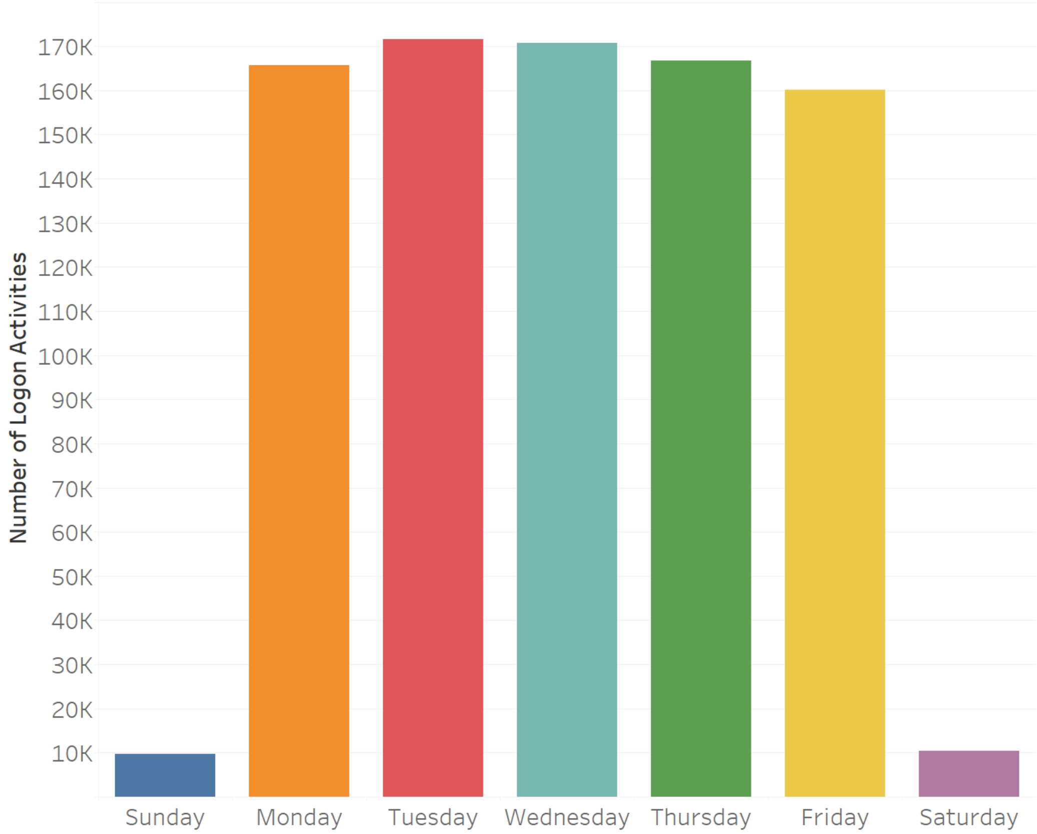
The specific objective of the procedure testing the first control is to examine if user access credential is immediately removed after employment termination. Descriptive statistics and visualization are selected to identify existing control deviations. The specific objective of the procedure testing the second control is to identify accounts that have suspicious behaviors so that the auditor can examine if these accounts were properly identified and addressed by the firm. Because the client firm stores information on previously confirmed suspicious activities, the auditor decides to use a decision tree, a supervised machine learning method, to predict if an account was used anomalously. Decision tree learning is a non-parametric supervised learning method that is widely used for classification purposes such as document classification (Harish, Guru, & Manjunath, 2010). By using historical labeled data (e.g. previously identified suspicious activities), the algorithm builds a decision tree that consists of internal nodes and leaves. Each internal node performs a test function that generates discrete outcomes. An instance is classified by recursively testing against the internal nodes in the tree until a leaf is reached, where the label of the leaf is the predicted category of the instance. Upon completion, the tree divides the full population into mutually exclusive subgroups (Kirkos, Spathis, & Manolopoulos, 2007). The biggest advantage of the decision tree is interpretability because the tree structure can be converted to a set of rules (Alpaydin, 2014), enabling the auditor to justify his conclusion.

Stage 5: obtain data, assess the reliability of the data, and reprocess the data for analysis.

After obtaining data, the auditor performs procedures to evaluate the reliability of the data using descriptive statistics. Specifically, the auditor scans the data to examine (1) if there is any gap in the data, and (2) if the distribution of the data is reasonable. For example, visualization shows that most logon activities happen on weekdays (Figure 4), which is consistent with the auditor’s expectation. After performing other analyses to validate the data, the auditor determines that the data to be analyzed is complete and accurate.

For the first procedure, the date part of the timestamp is extracted as the objective of the analysis is to determine if there is any logon activity after the day an employee is terminated. For the second procedure, the auditor identifies eight variables that are likely to be associated with suspicious behaviors. Specifically, the auditor considers whether there is abnormal number of emails exchanged, whether there is an abnormal number of attachments in the emails, whether there is an abnormal number of files transferred to external device, whether there is an abnormal number of hours logged onto a user's workstation, whether there are any after-hour logons, whether there is an abnormal number of times that an external device was connected to a user's workstation, whether there is any after-hour device plugin, and whether there is an abnormal number of times logging on to other's workstation. A full description of how these variables are calculated is presented in Appendix F. Because the data is highly imbalanced (1,362 threat activities versus 299,932 non-threat activities), both oversampling and undersampling are used to create a dataset with an equal number of observations in each class (13,620 observations in each group).

Figure 4. Number of Logon Activities by Day of the Week



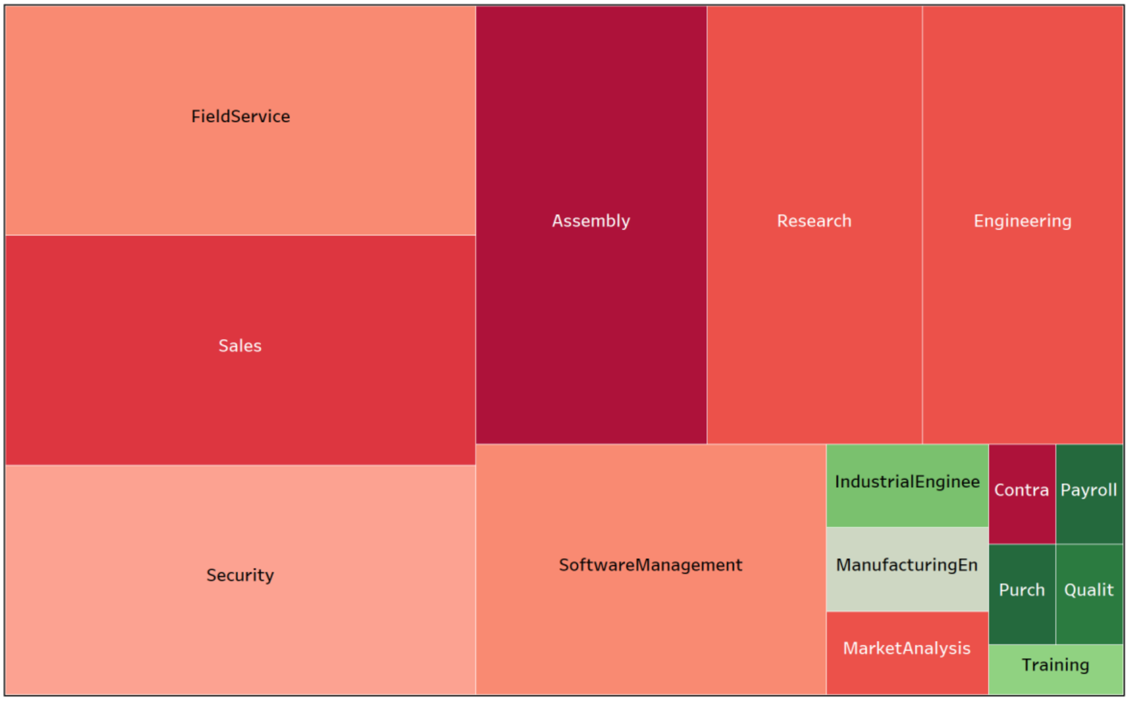
Stage 6: perform data analytics.

For the first procedure, the auditor first calculates the descriptive statistics for the gap between termination of employment and last logon activity. Table 20 shows the median, minimum, and maximum gap is 15 days, 2 days, and 32 days, respectively. Since a value larger than zero indicates that the credential is still being used after the employee is terminated, the descriptive statistics clearly demonstrate that there are control deviations. Because a significant number of exceptions is identified, data visualization is further utilized to explain the results of the analysis (Dilla, Janvrin, & Raschke, 2010) and group exceptions by category. The auditor decides to use a tree map to illustrate violations by department, with the size representing the number of employees that violate the controls and the color representing the largest gap between termination of employment and last logon activity in that department. By displaying two important elements in the tree map, visualization enables the auditor to locate the department that has the most serious control weakness (i.e. assembly department). In addition, Figure 5 indicates that the control deviations occur pervasively in the entire firm rather than just in one or several departments.

Table 20. Descriptive Statistics of the Gap between Termination of Employment and Last Logon Activity in Days

N	*MEAN*	*MIN*	*MEDIAN*	*MAX*	*STD*
154	15.66	2	15	32	8.48

Figure 5. Tree Map of Control Violations



For the second procedure, the auditor constructs a decision tree using C4.5 algorithm[12] embedded in Weka with 10-fold cross validation[13]. The model fit is reported in Table 21.

Table 21. Fit Statistics for the Decision Tree

Class	*Precision*	*Recall*	*ROC*
Threat Activity	87.1%	52.8%	72.4%
Non-threat Activity	66.1%	92.2%	72.4%

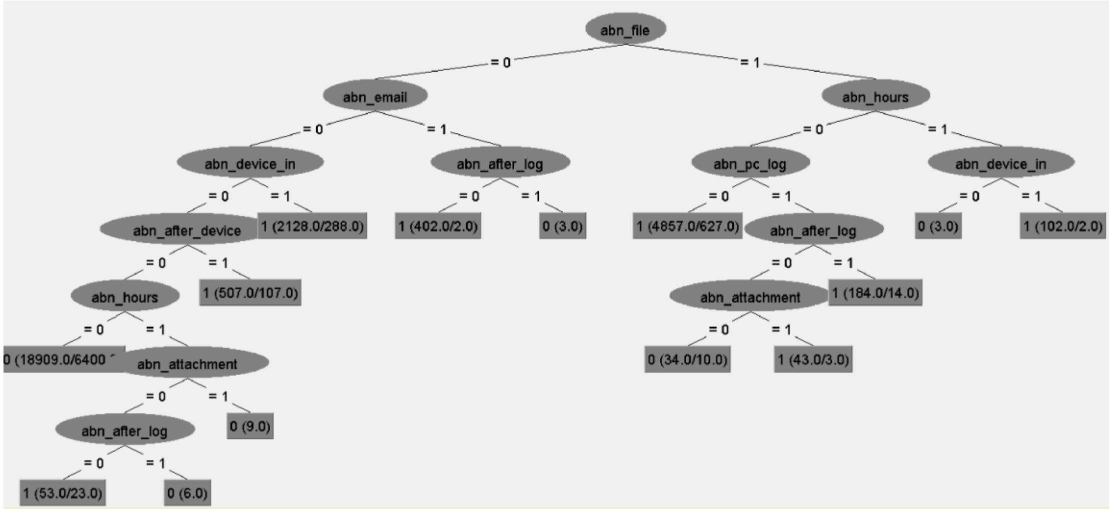
Specifically, the precision and recall [14] for the model are 0.871 and 0.528, respectively, suggesting that the predictive model can identify abnormal account activity with a high level of accuracy, but only around half of such activities can be identified by the model. Depending on the tradeoff between the cost of missing a true positive and the cost of investigating each identified case, the auditor can use a cost matrix to balance the achieved recall and precision. The decision tree is presented in Figure 6. The tree can be parsed to IF-THEN rules for explaining the logic underlying the identification. For example, one particular rule in the decision tree is that, if the account transferred an abnormal number of files to an external device, logged on for an abnormal number of hours, and inserted an external device an abnormal number of times, the activity is identified as an abnormal activity. The auditor concludes that the model has a reasonable fit and the IF-THEN rules generated by the decision tree is not counter-intuitive.

[12] For more information, see at https://en.wikipedia.org/wiki/C4.5_algorithm.

[13] Cross validation is to run the same algorithm on a specific number of resampled versions of the same dataset to fine-tune the model (Alpaydin, 2014). 10-fold cross validation partitions the sample into 10 equal sized subgroups. For each of the 10 interactions, 1 group is used as the validation data while the other 9 groups are used as the training data to build the model.

[14] Precision is the number of retrieved and relevant observations divided by the total number of retrieved observations, while recall is the number of retrieved relevant observations divided by the total number of relevant observations (Alpaydin, 2014).

Figure 6. Decision Tree



Stage 7: evaluate results from the procedure and make documentations.

After evaluating evidence obtained in the first procedure, the auditor concludes that controls were not operating effectively to constantly review user credentials. Both the procedure and the evidence were documented. The results of the analysis also lead the auditor to reassess the risk at the assertion level and redesign procedures because the evidence indicates increased risk of pervasive control weaknesses.

The decision tree model developed in the second procedure is used to identify suspicious account activities. The auditor compares the identified abnormal activities with firm's documentation and discovers that these instances are also identified by the firm and are properly addressed.

4.4 CRITICAL ISSUES

In this section, several issues pertaining to the use of data analytics in cybersecurity assurance are discussed. My intention is not to provide solutions but to point out areas that require further consideration. These practical and conceptual issues must be clarified for assurance process to be complete and coherent.

Are CPAs Qualified for Cybersecurity Assurance Engagement?

Since cybersecurity is about protecting information and systems in the cyber realm, a thorough understanding of information systems and network infrastructure is a prerequisite for a cybersecurity assurance engagement. In addition, the data-driven nature of the evidence collection process requires practitioners to possess a sufficient level of analytical skill. The prevalent view of the AICPA is that the Certified Public Accountants (CPAs) should be the ones to take the lead in a cybersecurity assurance engagement as the audit profession has established its reputation and credibility in assuring financial statement. Since the publication of Statement on Auditing Standards No. 70, *Service Organizations* in 1992, many CPAs began specializing in information technology risk and controls, with most moderate to large size public accounting firms developed sizable practices related to this specialty. As these practices grew, the firms added security and technology specialists who were not CPAs to their practices. Early in 1996, the AICPA identified information systems reliability as one of the six potential assurance areas that could generate revenue for public accounting firms (AICPA, 1996). In 2011, the AICPA published the Guide, *Reporting on Controls at a Services Organization Relevant to Security, Availability, Processing Integrity, Confidentiality and Privacy*, which significantly increased the number of examination reports issued by CPAs on information security. Compared with other professionals, the audit profession has advantages in several aspects for providing cybersecurity assurance. First and foremost, the CPAs use the concepts and terminology from traditional audit to provide rigor and consistency for new assurance area (Free, Salterio, & Shearer, 2009; O'Dwyer, 2011; Power, 1997). In addition, because Sarbanes-Oxley Act (SOX) Section 404 requires auditors to attest to and report on the management assessment of the effectiveness of Internal Controls over Financial

Reporting (ICFR), the audit profession has accumulated knowledge and expertise at risk assessment and evaluation for a subset of cybersecurity controls. The CPAs are also active in providing information security services or advisory engagements, with four of the top ten information security consultants being public accounting firms (AICPA, 2017b).

However, insufficient numbers of qualified personnel as well as gaps in knowledge and skills may preclude the audit profession from taking the leading role in emerging cybersecurity assurance. The CPA practitioners specializing in information technology controls are constantly challenged to obtain and expand their knowledge and skills in information technology, data analytics, and statistical modeling, which are imperative in assuring cybersecurity. Expanding the number of CPAs to support such examinations by training those that have been focused on financial auditing will be challenging. A similar concern was raised by No and Vasarhelyi (2017), who argue that even seeking aid from IT professionals and data scientists would be difficult in the absence of a certain level of education on statistics and information technology. The CPA firms usually involve IT specialists examining their clients' IT general controls as part of the ICFR evaluation; nonetheless, such arrangement is unlikely to be efficient and effective for a technology-centric and data-centric engagement like cybersecurity assurance, and as a result, such engagement will need to be led by the CPAs IT specialists. In addition, CPAs are not accustomed to collecting and analyzing nonfinancial information from various sources such as social network (Brown-Liburd et al., 2015). A change in the knowledge sets of CPAs should precede an assurance engagement on cybersecurity.

How to Address the Flood of Exceptions?

Using data analytics in a cybersecurity examination is likely to bring about technical challenges. A downside of examining the whole population and data from various sources is that a large number of anomalies (or exceptions) are likely to be generated and will require further examination by the practitioners (Cao et al., 2015; Debreceny, Gray, & Rahman, 2003). How to deal with these anomalies remains a challenge. It is inefficient for practitioners to manually examine all the anomalies as many of them may be false positives. For example, when examining system logs, all unusual activities will be identified for further inspection. Instead of investigating each anomaly, a more feasible and effective way is to develop a methodology that can identify true anomalies (called exceptional exceptions) by developing a system of prioritization (Issa, 2013). In the context of cybersecurity, it is ex-ante not clear how to optimally correlate data from different sources to prioritize exceptions and direct practitioners' attention to the true exceptions.

What is Materiality?

Practitioners must consider qualitative and quantitative factors when developing the overall engagement strategy (AICPA, 2017a). In traditional financial reporting, the concept of materiality indicates that some matters, either individually or in the aggregate, are important for the fair presentation of financial statements in conformity with Generally Accepted Accounting Principles (GAAP) while other matters are not important (SAS No. 107). The concept of materiality is fundamentally different in cybersecurity assurance. For the description assertion, materiality relates to the consideration whether description misstatement such as omissions of information in the presentation, individually or in the aggregate, could reasonably be expected to influence decision makings (AICPA, 2017c). Determining materiality for qualitative disclosure is itself a challenge because there is no common unit of measurement (Cohen & Simnett, 2014; Wallage, 2000). When there is no dollar amount, determination of materiality is significantly more judgmental and draws on the knowledge and experience of the practitioner[15]. For the control assertion, the concept of materiality becomes complicated because it is intrinsically difficult to differentiate material and immaterial matters. For example, breach of access to a financially immaterial unit may be inconsequential itself, but could enable malicious parties to exploit other systems that contain sensitive information such as business plans, which could be considered as material. The interconnected nature of information systems implies that vulnerabilities in any part of the system could be exploited to penetrate other segments in the same system, which makes the determination of materiality more challenging. In addition, the consequences of cybersecurity issues could range from damaged reputation to loss of intellectual property, which cannot be directly measured in monetary terms as in a traditional audit (No & Vasarhelyi, 2017). While data analytics can facilitate examining 100 percent of the data evidencing the operating effectiveness of cybersecurity controls, it provides little value for determining materiality. Absent numeric values, the concept of materiality should be revisited in the cybersecurity context to take the technical nature into consideration. For example, is it possible to use time, such as the number of hours system is down, as the unit for determining materiality?

Can Privacy be Preserved?

Privacy refers to an individual's right to disclose personal information at his or her own discretion (Shapiro & Baker, 2001). While data analytics, especially big data analytics, is always associated with privacy concerns (Cao et al., 2015), privacy issues can

[15] The description criteria released by the AICPA provide some suggestions for materiality consideration when preparing and evaluating management description. For more information, see at http://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/DownloadableDocuments/Cybersecurity/Description-Criteria.pdf.

be more salient in cybersecurity assurance. Yoon et al. (2015) argue that mining emails is a serious privacy issue that must be addressed before adopting big data in auditing. Analyzing user activities using system logs, on the other hand, would potentially reveal much more personal information than email mining. At an extreme case, keystrokes and web browsing history can be accessed and analyzed by the practitioner. Almost certainly, analytics on data at this granularity level would reveal sensitive information that is likely to violate one's privacy[16]. It may be impractical to inform the employees in advance as the usefulness of data may be impaired if one, due to the notice, starts to be careful about his or her activities while using the system. Anonymization of the data may be a solution to address the privacy issue; however, it introduces new issues such as the difficulty of integration with other data and the potential manipulation of data by the firm.

Is Data Always Easily Accessible?

Although data analytics can be beneficial in cybersecurity assurance, data acquisition could be difficult and may hinder the application of data analytics. Even in traditional financial audit, auditors do not have ready access to their clients' accounting and transaction data, which prompted the AICPA ASEC Emerging Assurance Technologies Task Force to recently issue the audit data standards[17] (Zhang, Pawlicki, McQuilken, & Titera, 2012). Acquiring data may be more challenging in cybersecurity assurance engagement because such data are more sensitive. It is not difficult to extract information from social media because it is publicly available; in contrast, firms may be reluctant to deliver system log data and network traffic data since such data are likely to

[16] It is possible that employers have developed specific policies to defeat employees' expectation of privacy.

[17] For more information, see at https://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/Pages/AuditDataStandards.aspx.

reveal sensitive information after deeper analysis. Given the fact that cybersecurity assurance is voluntary at this stage, data availability issue is likely to remain a challenge.

How to Handle Risks Arising from Interaction with Third Parties?

In the modern economy, firms’ information systems are usually interconnected with business partners, customers, and vendors. It is also not uncommon that firms outsource part of their information storages and processing activities to third parties. The interconnected nature of business may give rise to vulnerabilities that result in a firm’s failure to achieve its cybersecurity objectives. Even if the firm has implemented effective controls to protect information assets, hackers could gain unauthorized access to proprietary data by compromising third parties. A recent survey by SOHA (2016) reveals that 63 percent of data breaches were either directly or indirectly linked to third-party vendors, highlighting the importance of managing cybersecurity risks arising from third parties.

In evaluating risks from third parties, practitioners should obtain sufficient evidence of the operating effectiveness of the third party’s controls, which is not a trivial endeavor. A practitioner could rely on the third party’s cybersecurity report that is signed by another practitioner because it is unlikely that the third party will allow the firm’s practitioner to directly perform procedures to examine the third party’s controls unless the firm and the third party share the same practitioner. The extent and quality of examination conducted by the third party’s practitioner, however, may not be the same as the firm’s practitioner. Accordingly, the firm’s practitioner may not obtain sufficient evidence pertaining to the third party’s cybersecurity controls. While the practitioner could use big data analytics to analyze social media, news articles, and online forums to examine whether the third party is effectively managing cybersecurity, the concern is only alleviated rather than remediated. A solution that enables practitioners to analyze third parties' data while preserving its confidentiality is needed. For example, privacy-preserving data mining, which is developed in the domain of cryptography, may be leveraged to achieve the objective.

4.5 CONCLUDING REMARKS

Cybersecurity assurance is an emerging field that is important to the audit profession. This paper contributes to the literature by discussing detailed issues in cybersecurity assurance from a data analytics perspective. Particularly, I argue that data analytics can be highly effective because of the nature of the engagement. A process of using data analytics in testing cybersecurity controls is further discussed. The seven steps identified in the process, from identifying assertions to evaluating results, are explained in detail to facilitate practitioners to leverage data analytics in their practices. Illustrations using synthetic data are presented to demonstrate the usefulness of data analytics for testing cybersecurity controls. Both descriptive analytics and predictive analytics are utilized to gather evidence pertaining to the assertions identified in the hypothetical engagement. Finally, some critical issues related to the use of data analytics in cybersecurity assurance that must be clarified and addressed are discussed in this essay. One obvious limitation of the study is that I am unable to examine the proposed process in an actual engagement. However, as cybersecurity assurance is only a recent development, studies like this that advance our understanding of this matter are needed.

Based on the discussions presented in this paper and especially the critical issues highlighted in the fourth section, some potential research topics that are likely to be valuable for future researchers are summarized below.

1. What is the best way to define materiality? Is there any measure that can substitute for the dollar amount for determining materiality in traditional financial statement audit? Because materiality should be considered throughout the cybersecurity assurance engagement, the lack of clear definition of materiality could result in inconsistent procedures. Future research should identify a set of candidates, such as the time elapsed between the occurrence and the detection of an incident, and examine their usefulness to determine the best measure.
2. What types of data and data analytics techniques are most appropriate in cybersecurity assurance? Because of the richness of data that can potentially be analyzed, guidance must be developed to help practitioners select the best type of data given the specific circumstances. Similarly, procedures should be in place to help practitioners to determine the best approach among a wide range of data analytics techniques based on the nature of the data and the engagement.
3. How to prioritize exceptions after performing data analytics? Analyzing disaggregated data is likely to generate many exceptions, which may overwhelm practitioners and reduce audit efficiency. Li, Chan, and Kogan (2016) proposed a framework to prioritize exceptions from a continuous auditing system by assuming rules to identify exceptions are independent. However, many types of data in cybersecurity assurance are interconnected. Therefore, a refined framework that takes the interrelationship into consideration should be developed.
4. How can a cybersecurity assurance engagement complement the traditional financial statement audit engagement? Although the two engagements differ in

objectives, evidence collected in cybersecurity assurance can potentially be utilized as audit evidence. For example, the procedures performed to examine information and systems integrity are likely to have implications for IT controls that ensure the integrity of accounting data. To what extent auditors can rely on the evidence collected in a cybersecurity assurance engagement and whether it is economically efficient for the same auditor to take both engagements is unclear, requiring both archival and analytical research. Furthermore, if evidence for assuring cybersecurity is collected using data analytics and is subsequently utilized in the audit engagement, audit standards should be updated to accommodate and encourage such techniques.

5. How to audit third parties? In addition to relying on the assurance reports of third parties, is it possible to audit third parties while preserving their data confidentiality? For example, data can be encrypted by the third party and submitted to the practitioner, who performs data analytics on the encrypted information. Researchers could borrow techniques from cryptography to develop procedures for aggregating, summarizing, and categorizing encrypted data. Alternatively, future studies could explore what types of data that are obtained outside the third parties can be most effective in evaluating cybersecurity risks associated with the third party.

CHAPTER 5: CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

5.1 CONCLUSION

This dissertation is an attempt to broaden the understanding of cybersecurity from three perspectives: external audit, risk disclosure, and assurance service. The first essay reveals a potential relationship between the external audit and cyber incidents. Specifically, using data on reported cyber incidents for the period 2005 to 2015, I observe a significant positive association between audit fee increase and cyber incidents. Following cyber incidents, increases in audit fees are smaller for firms with prior cybersecurity risk disclosures, implying that auditors have priced material cybersecurity risk prior to the cyber-attacks. In addition, evidence in this essay demonstrates that firms with repeated cyber incidents are charged higher audit fees than firms that are only breached once. Furthermore, auditors differentiate the type of information hacked as increases in audit fees are higher for firms with cyber incidents that involve intellectual property than for firms with hacking not involving intellectual property. Finally, auditor’s concern over cyber incidents is mitigated by institutional holdings and large block holders. Collectively, results in the first essay should be valuable to regulators and academics who are interested in understanding auditor’s opinion over cyber incidents. The findings that auditors both price cybersecurity risk ex-ante and respond to cyber incidents ex-post disagree with the concern that auditors are not taking cybersecurity seriously.

In the second essay, whether cybersecurity risk disclosure is informative for future reported cyber incident is examined using two measures: the presence of cybersecurity risk disclosure and the content of cybersecurity risk disclosure as measured by adjusted length. Both the presence and content of cybersecurity risk disclosure are found to be positively associated with subsequently reported cyber incidents, suggesting that cybersecurity risk disclosure is not boilerplate. The results also demonstrate that investors are only using information conveyed by the presence of, but not the length of cybersecurity risk disclosure. Furthermore, it is shown that there is a differential effect before and after the SEC's cybersecurity disclosure guidance. The presence of cybersecurity risk disclosure is no longer associated with subsequent cyber incidents, revealing that the SEC's emphasis on cybersecurity risk disclosures results in more disclosures by firms not having material cybersecurity risks. The study did not find a significant association between firm-specific disclosure and cyber incidents, providing some relief for the regulator's concern that more firm-specific disclosures may provide information for hackers. Finally, the topic analysis indicates that firms are more concerned about business operations and financial performance when encountering cybersecurity issues. Moreover, there is a growing concern regarding reputation damage and loss of intellectual property due to cyber incidents. Collectively, results in this essay should be valuable to practitioners, regulators, and academics who are interested in the informativeness of cybersecurity risk disclosures. I stand with the SEC to emphasize the importance of cybersecurity risk disclosure, but raise a question about the potential unintended consequence resulting from the disclosure guidance.

In the third essay, issues surrounding cybersecurity assurance is discussed from a data analytics perspective. Particularly, I argue that data analytics can be effectively leveraged to cybersecurity assurance. A process consists of seven steps in testing cybersecurity controls using data analytics are further explained in detail. Illustrative examples using synthetic data are presented to demonstrate the usefulness of data analytics following the process. Specifically, both descriptive and predictive analytics are utilized to gather evidence pertaining to the identified assertions. Finally, some critical issues related to the use of data analytics in cybersecurity assurance are discussed.

5.2 LIMITATIONS AND FUTURE RESEARCH

This dissertation is not without limitation. For the first essay, the results of the study do not address how external auditors are evaluating cybersecurity risks prior to cyber incidents. A thorough investigation is necessary to advance our understanding of cybersecurity risk anticipation. For example, are auditors treating cybersecurity risk as an operational risk or dealing with it differently? In addition, although auditors should respond to cyber incidents because such incidents may be indicative deficiencies in ICFR and risks of material misstatement, there could be other reasons why external auditors would increase audit fees following a cyber incident, which requires in-depth case studies or interviews with external auditors. Moreover, investigating the contagion effect of cyber incidents would be of great interest. After a firm was breached, will audit fees of its competitors also increase? Alternatively, after a client experiences a cyber incident, will the auditor charge higher fees for all the clients? Another issue that is not addressed is the potential deterrence effect. If auditors are punishing breached firms, will they serve the monitoring role to encourage firms to make more cybersecurity investments so that future incidents may be deterred?

For the second essay, cyber incident is used as the proxy for material cybersecurity risk, which may not be always accurate as theoretically any system can be breached regardless of the security measures. Future studies could explore additional measures. For example, the number of attempted attacks identified by a firm’s intrusion detection system may be a good measurement for cybersecurity risk. Another limitation is that the study implies that firms are knowledgeable about the cybersecurity risks they face, which is an arguable assumption. Examining firms’ awareness of cybersecurity risk would be an interesting research topic. Further, the second essay does not answer the question why investors are not utilizing information conveyed in the content of cybersecurity risk disclosure. There could be at least two alternative explanations. Investors may be unaware of the informativeness of the content of cybersecurity risk disclosures. Alternatively, investors may decide not to price such risk even if they are aware of it because they believe that firms with lengthy cyber disclosures are likely to significantly increase investments to address the risk, reducing the probability of future material incidents. Another avenue for future research is to examine the time-series change of firms’ cybersecurity risk disclosures. While the essay demonstrates that such disclosure is informative in a cross-sectional setting, it is possible that the change in a firm’s disclosure from year to year may also convey useful information.

One obvious limitation of the third essay is that the proposed process is not empirically validated in an actual engagement. Future studies may identify unforeseen issues when applying the process. Additionally, the issue about how a cybersecurity assurance engagement could complement an audit engagement is not discussed. Future research could explore, for example, if cybersecurity assurance can be integrated into the financial audit. Another limitation is that the study did not offer specific guidance to help practitioners select data source and data analytics techniques. Without such guidance, practitioners may be overwhelmed by a wide range of available data and techniques.

Procedures can be developed to offer recommendations for practitioners based on the particularity of the engagement.

Despite these limitations, this dissertation contributes to our understanding of emerging cybersecurity issues, and addresses questions that have strong implications for regulators, investors, practitioners, and researchers.

BIBLIOGRAPHY

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Aguilar, L. A. (2014). Boards of Directors, Corporate Governance and Cyber-Risks.

Akresh, A. D. (2010). A risk model to opine on internal control. *Accounting Horizons, 24*(1), 65-78.

Alpaydin, E. (2014). *Introduction to machine learning*: MIT press.

American Institute of Certified Public Accountants. (1996). *Report of the Special Committee on Assurance Services (The Elliott Report)*.

American Institute of Certified Public Accountants. (1997). Consideration of Fraud in a Financial Statement Audit. *Statement on Auditing Standards No. 82*.

American Institute of Certified Public Accountants. (2016). Cybersecurity Reporting: A Backgrounder, Available at https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/cybersecurity/aicpa_brief_cybersecurity.pdf.

American Institute of Certified Public Accountants. (2017a). AT-C Section 205: Examination Engagements.

American Institute of Certified Public Accountants. (2017b). Cybersecurity risk management reporting fact sheet, Available at https://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/DownloadableDocuments/Cybersecurity-Fact-Sheet.PDF.

American Institute of Certified Public Accountants. (2017c). Reporting on an Entity's Cybersecurity Risk Management Program and Controls - Attestation Guide.

Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems, 25*, 29-44.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: research needs. *Auditing: A Journal of Practice & Theory, Forthcoming*.

Association of Corporate Counsel. (2016). SEC priorities and enforcement trends, available at http://m.acc.com/chapters/del/upload/2016-04-19_AkinGump_SEC_Trends-PPTX.pdf.

Bao, Y., & Datta, A. (2014). Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. *Management Science, 60*(6), 1371-1391. doi:10.1287/mnsc.2014.1930

Beatty, A., Cheng, L., & Zhang, H. (2015). Sometimes Less is More: Evidence from Financial Constraints Risk Factor Disclosures. *Working Paper*.

Benaroch, M., Chernobai, A., & Goldstein, J. (2012). An internal control perspective on the market value consequences of IT operational risk events. *International Journal of Accounting Information Systems, 13*(4), 357-381.

Bennett, C. (2015). SEC weighs cybersecurity disclosure rules. *The Hill*.

Berezina, K., Cobanoglu, C., Miller, B. L., & Kwansa, F. A. (2012). The impact of information security breach on hotel guest perception of service quality, satisfaction, revisit intentions and word-of-mouth. *International journal of contemporary hospitality management, 24*(7), 991-1010. doi:10.1108/09596111211258883

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics, 50*(2-3), 296-343. doi:10.1016/j.jacceco.2010.10.003

Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. *Accounting Horizons, 29*(2), 451-468.

Brown, S. V., Tian, X. S., & Tucker, J. W. (2015). The spillover effect of SEC comment letters on qualitative corporate disclosure: Evidence from the risk factor disclosure. *Working Paper*.

Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research, 49*(2), 309-346.

Bryan Cave. (2016). 2016 Data Breach Litigation Report.

Brynjolfsson, E., Hammerbacher, J., & Stevens, B. (2011). Competing through data: Three experts offer their game plans. *McKinsey Quarterly, 4*, 36-47.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly, 34*(3), 523-548.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H.-m., & Steele., L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies, 19*(1), 396-455.

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security, 11*(3), 431-448.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement audits. *Accounting Horizons, 29*(2), 423-429.

Cardenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy, 11*(6), 74-76.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce, 9*(1), 70-104.

Center for Audit Quality. (2014). CAQ Member Alert: Cybersecurity and the External Audit.

Chai, S., Kim, M., & Rao, H. R. (2011). Firms' information security investment decisions: Stock market evidence of investors' behavior. *Decision Support Systems, 50*(4), 651-661. doi:10.1016/j.dss.2010.08.017

Chang, C. J., & Hwang, N.-C. (2003). The impact of retention incentives and client business risks on auditors' decisions involving aggressive reporting practices. *Auditing: A Journal of Practice & Theory, 22*(2), 207-218. doi:DOI 10.2308/aud.2003.22.2.207

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165-1188.

CISCO. (2017). Annual Cybersecurity Report.

Cohen, J. R., & Simnett, R. (2014). CSR and assurance services: A research agenda. *Auditing: A Journal of Practice & Theory, 34*(1), 59-74.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach. *Information Systems Research, 20*(1), 79-98.

Das, S., Mukhopadhyay, A., & Anand, M. (2012). Stock market response to information security breach: A study using firm and attack characteristics. *Journal of Information Privacy and Security, 8*(4), 27-55.

Debreceny, R. S. (2013). Research on IT Governance, Risk, and Value: Challenges and Opportunities. *Journal of Information Systems, 27*(1), 129-135.

Debreceny, R. S. (2014). Aggravated Cybersecurity Risks Implications for Accounting and Auditing Research and Practice. *JIS Senior Editors' Blog-Journal of Information Systems*.

Debreceny, R. S., Gray, G. L., & Rahman, A. (2003). The determinants of Internet financial reporting. *Journal of Accounting and Public Policy, 21*(4), 371-394.

Desir, R., Casterella, J. R., & Kokina, J. (2013). A reexamination of audit fees for initial audit engagements in the post-SOX period. *Auditing: A Journal of Practice & Theory, 33*(2), 59-78.

Dilla, W., Janvrin, D. J., & Raschke, R. (2010). Interactive data visualization: New directions for accounting information systems research. *Journal of Information Systems, 24*(2), 1-37.

Doogar, R., Sivadasan, P., & Solomon, I. (2015). Audit fee residuals: costs or rents? *Review of Accounting Studies, 20*(4), 1247-1286.

Doyle, J., Ge, W., & McVay, S. (2007). Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics, 44*(1), 193-223.

Edmans, A. (2014). Blockholders and corporate governance. *Annual Review of Financial Economics, Vol 6, 6*, 23-50. doi:10.1146/annurev-financial-110613-034455

Elliott, J. A., Ghosh, A., & Peltier, E. (2013). Pricing of risky initial audit engagements. *Auditing: A Journal of Practice & Theory, 32*(4), 25-43. doi:10.2308/ajpt-50523

Ettredge, M. L., & Richardson, V. J. (2003). Information transfer among internet firms: the case of hacker attacks. *Journal of Information Systems, 17*(2), 71-82.

Ferraro, M. F. (2013). Groundbreaking'or Broken? An Analysis of SEC Cyber-Security Disclosure Guidance, Its Effectiveness, and Implications.

Filzen, J. J. (2015). The information content of risk factor disclosures in quarterly reports. *Accounting Horizons, 29*(4), 887-916.

Filzen, J. J., McBrayer, G., & Shannon, K. (2016). Risk Factor Disclosures: Do Managers and Markets Speak the Same Language? *Working Paper*.

Financial Accounting Standards Board. (1975). Statement of Financial Accounting Standards No. 5: Accounting for Contingencies

Francis, J. R., & Wang, D. (2005). Impact of the SEC's public fee disclosure requirement on subsequent period fees and implications for market efficiency. *Auditing: A Journal of Practice & Theory, 24*(1), 145-160.

Free, C., Salterio, S. E., & Shearer, T. (2009). The construction of auditability: MBA rankings and assurance in practice. *Accounting, Organizations and society, 34*(1), 119-140.

Gatzlaff, K. M., & McCullough, K. A. (2010). The Effect of Data Breaches on Shareholder Wealth. *Risk Management and Insurance Review, 13*(1), 61-83.

Gaulin, M. (2017). Risk Fact or Fiction: The information content of risk factor disclosures. *Working Paper*.

Gelinne, J., Fancher, J. D., & Mossburg, E. (2016). The hidden costs of an IP breach: Cyber theft and the loss of intellectual property. *Deloitte Review*(19).

Glasser, J., & Lindauer, B. (2013). Bridging the gap: A pragmatic approach to generating insider threat data. *IEEE Security and Privacy Workshops (SPW)*, 98-104.

Goel, S., & Shawky, H. A. (2009). Estimating the market impact of security breach announcements on firm values. *Information & Management, 46*(7), 404-410. doi:10.1016/j.im.2009.06.005

Goldstein, J., Chernobai, A., & Benaroch, M. (2011). An event study analysis of the economic impact of IT operational risk and its subcategories. *Journal of the Association for Information Systems, 12*(9), 606-631.

Gordon, L. A., Loeb, M. P., & Sohail, T. (2010). Market value of voluntary disclosures concerning information security. *MIS Quarterly, 34*(3), 567-594.

Gordon, L. A., Loeb, M. P., & Zhou, L. (2011). The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security, 19*(1), 33-56.

Grant, G. H., & Grant, C. T. (2014). SEC cybersecurity disclosure guidance is quickly becoming a requirement. *The CPA Journal, 84*(5), 69.

Hardies, K., Breesch, D., & Branson, J. (2015). The Female Audit Fee Premium. *Auditing: A Journal of Practice & Theory, 34*(4), 171-195.

Harish, B. S., Guru, D. S., & Manjunath, S. (2010). Representation and classification of text documents: A brief review. *IJCA, Special Issue on RTIPPR, 2*, 110-119.

Higgs, J. L., Pinsker, R., Smith, T., & Young, G. (2014). The Relationship Between Board-Level Technology Committees and Reported Security Breaches. *Journal of Information Systems*.

Hilary, G., Segal, B., & Zhang, M. H. (2017). Cyber-Risk Disclosure: Who Cares? *Working Paper*.

Hinz, O., Nofer, M., Schiereck, D., & Trillig, J. (2015). The influence of data theft on the share prices and systematic risk of consumer electronics companies. *Information & Management, 52*(3), 337-347. doi:10.1016/j.im.2014.12.006

Hoag, M. L., & Hollingsworth, C. W. (2011). An intertemporal analysis of audit fees and Section 404 material weaknesses. *Auditing: A Journal of Practice & Theory, 30*(2), 173-200. doi:10.2308/ajpt-50005

Hogan, C. E., & Wilkins, M. S. (2008). Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies? *Contemporary Accounting Research, 25*(1), 219-242.

Hoitash, R., Hoitash, U., & Bedard, J. C. (2008). Internal control quality and audit pricing under the Sarbanes-Oxley Act. *Auditing: A Journal of Practice & Theory, 27*(1), 105-126. doi:DOI 10.2308/aud.2008.27.1.105

Hope, O.-K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies, Forthcoming*.

Hu, Q., Dinev, T., Hart, P., & Cooke, D. (2012). Managing employee compliance with information security policies: the critical role of top management and organizational culture. *Decision Sciences, 43*(4), 615-660.

Huang, H.-W., Raghunandan, K., & Rama, D. (2009). Audit fees for initial audit engagements before and after SOX. *Auditing: A Journal of Practice & Theory, 28*(1), 171-190. doi:10.2308/aud.2009.28.1.171

IBM. (2013). Descriptive, predictive, prescriptive: transforming asset and facilities management with analytics. *Thought Leadership White Paper*.

Ifinedo, P. (2014). Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. *Information & Management, 51*(1), 69-79. doi:10.1016/j.im.2013.10.001

ISACA. (2006). Information Security Governance Guidance for Boards of Directors and Executive Management, 2nd Edition

Issa, H. (2013). Exceptional exceptions. *Doctoral dissertation, Rutgers University-Graduate School-Newark*.

Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review, 89*(5), 1751-1773.

Johnson, S. (2010). SEC pushes companies for more risk information. *CFO Magazine*.

Johnstone, K. M. (2000). Client-acceptance decisions: Simultaneous effects of client business risk, audit risk, auditor business risk, and risk adaptation. *Auditing: A Journal of Practice & Theory, 19*(1), 1-25.

Ke, B., Huddart, S., & Petroni, K. (2003). What insiders know about future earnings and how they use it: Evidence from insider trades. *Journal of Accounting and Economics, 35*(3), 315-346.

Khalil, S., & Mazboudi, M. (2016). Client Acceptance and Engagement Pricing following Auditor Resignations in Family Firms. *Auditing: A Journal of Practice & Theory, 35*(4), 137-158. doi:10.2308/ajpt-51489

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995-1003.

Kogan, A., Alles, M. G., Vasarhelyi, M. A., & Wu, J. (2014). Design and evaluation of a continuous data level auditing system. *Auditing: A Journal of Practice & Theory, 33*(4), 221-245.

Konrad, J. (2000). Motion detection and estimation. *Handbook of Image and Video Processing, 207*, 225.

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *Journal of Accounting and Economics, 84*(5), 1639-1670.

Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do managers withhold bad news? *Journal of Accounting Research, 47*(1), 241-276.

Krahel, J. P., & Titera, W. R. (2015). Consequences of big data and formalization on accounting and auditing standards. *Accounting Horizons, 29*(2), 409-422.

Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies, 18*(4), 1088-1122.

Kwon, J., Ulmer, J. R., & Wang, T. (2013). The association between top management involvement and compensation and information security breaches. *Journal of Information Systems, 27*(1), 219-236.

Lawrence, A., Minutti-Meza, M., & Vyas, D. (2016). Is Operational Control Risk Informative of Undetected Financial Reporting Deficiencies? *Working Paper*.

Lennox, C. S., Francis, J. R., & Wang, Z. (2011). Selection models in accounting research. *The Accounting Review, 87*(2), 589-616.

Li, H., No, W. G., Wang, T., & Vasarhelyi, M. A. (2017). Cybersecurity assurance and cybersecurity disclosure guidance. *Working Paper*.

Li, P., Chan, D. Y., & Kogan, A. (2016). Exception Prioritization in the Continuous Auditing Environment: A Framework and Experimental Evaluation. *Journal of Information Systems, 30*(2), 135-157.

Lin, J. W., Hwang, M. I., & Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, *18*(8), 657-665.

Loop, P. (2016). Cybersecurity and the Board: 8 Issues Keeping Directors up at Night. *The Wall Street Journal*.

Mayhew, B. W. (2005). Discussion of impact of the SEC's public fee disclosure requirement on subsequent period fees and implications for market efficiency. *Auditing-a Journal of Practice & Theory, 24*, 161-169. doi:DOI 10.2308/aud.2005.24.s-1.161

Mitra, S., Hossain, M., & Deis, D. R. (2007). The empirical relationship between ownership characteristics and audit fees. *28*(3), 257-285.

Mitra, S., & Ransbotham, S. (2015). Information Disclosure and the Diffusion of Information Security Attacks. *Information Systems Research*, *26*(3), 565-584.

Munsif, V., Raghunandan, K., Rama, D. V., & Singhvi, M. (2011). Audit fees after remediation of internal control weaknesses. *Accounting Horizons, 25*(1), 87-105. doi:10.2308/acch.2011.25.1.87

No, W. G., & Vasarhelyi, M. A. (2017). Cybersecurity and continuous assurance. *Journal of Emerging Technologies in Accounting*, Forthcoming.

O’Dwyer, B. (2011). The case of sustainability assurance: Constructing a new assurance service. *Contemporary Accounting Research, 28*(4), 1230-1266.

Ponemon Institute. (2016). 2016 Cost of Data Breach Study: United States.

Power, M. (1997). Expertise and the construction of relevance: accountants and environmental audit. *Accounting, Organizations and society, 22*(2), 123-146.

PricewaterhouseCoopers. (2016). The Global State of Information Security.

Public Company Accounting Oversight Board. (2004). Auditing Standard No. 2: An audit of internal control over financial reporting performed in conjuction with an audit of financial statements.

Public Company Accounting Oversight Board. (2011). AU Section 350: Audit Sampling.

Public Company Accounting Oversight Board. (2014). Standing advisory group meeting: cybersecurity. Available at http://pcaobus.org/News/Events/Documents/0624252014_SAG_Meeting/06252014_Cybersecurity.pdf

Public Company Accounting Oversight Board. (2015). Staff inspection brief.

Public Company Accounting Oversight Board. (2016). Staff inspection brief.

Ransbotham, S., & Mitra, S. (2009). Choice and chance: A conceptual model of paths to information security compromise. *Information Systems Research*, *20*(1), 121-139.

Reed, J. (2017). Cybersecurity Disclosure Act of 2017. *Available at* https://www.congress.gov/bill/115th-congress/senate-bill/536/text?q=%7B%22search%22%3A%5B%22Cybersecurity+Disclosure+Act+of+2017%22%5D%7D&r=1.

Reuters. (2005). Refco risks boiler-plate disclosure. By Scott Malone.

Reuters. (2015). Australian metal detector company counts cost of Chinese hacking.

Robbins, R. B., & Rothenberg, P. L. (2005). Writing effective risk factor disclosure in offering documents and exhange act reports. *Insights: The Corporate & Securities Law Advisor, 19*(5).

Romanosky, S., Hoffman, D., & Acquisti, A. (2014). Empirical Analysis of Data Breach Litigation. *Journal of Empirical Legal Studies, 11*(1), 74-104. doi:10.1111/jels.12035

Schneider, G. P., Dai, J., Janvrin, D. J., Ajayi, K., & Raschke, R. L. (2015). Infer, predict, and assure: Accounting opportunities in data analytics. *Accounting Horizons, 29*(3), 719-742.

Schubert, D. F., Cedarbaum, J. G., & Schloss, L. (2015). The SEC's Two Primary Theories in Cybersecurity Enforcement Actions. *The Cybersecurity Law Report*.

Securities and Exchange Commission. (2005). Release #33-8591: Securities offering reform (Section VII: Additional Exchange Act disclosure provisions).

Securities and Exchange Commission. (2010). 17 CFR PARTS 211, 231 and 241. Release Nos. 33-9106; 34-61469; FR-82.

Securities and Exchange Commission. (2011). CF Disclosure Guidance: Topic No. 2: Cybersecurity.

Securities and Exchange Commission. (2014). Cybersecurity Roundtable. Available at https://www.sec.gov/spotlight/cybersecurity-roundtable.shtml.

Shabtai, A., Elovici, Y., & Rokach, L. (2012). *A survey of data leakage detection and prevention solutions*: Springer Science & Business Media.

Shapiro, B., & Baker, C. R. (2001). Information technology and the social construction of information privacy. *Journal of Accounting and Public Policy, 20*(4), 295-322.

Sharma, V. D. (2004). Board of director characteristics, institutional ownership, and fraud: Evidence from Australia. *Auditing-a Journal of Practice & Theory, 23*(2), 105-117. doi:DOI 10.2308/aud.2004.23.2.105

Sheneman, A. G. (2017). The Effect of Operating Control Failures on the Cost of Capital-Evidence from Data Breaches. *Working Paper*.

Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity score matching in accounting research. *The Accounting Review, 92*(1), 213-244. doi:10.2308/accr-51449

Shumsky, T. (2016). Corporate Judgment Call: When to Disclose You've Been Hacked. *The Wall Street Journal*.

Skinner, D. J. (1994). Why firms voluntarily disclose bad news. *Journal of Accounting Research, 32*(1), 38-60.

Soha System. (2016). Third Party Access Is A Major Source Of Data Breaches, Yet Not An IT Priority, Available at http://go.soha.io/hubfs/Survey_Reports/Soha_Systems_Third_Party_Advisory_Group_2016_IT_Survey_Report.pdf?t=1467123126371.

Soomro, Z. A., Shah, M. H., & Ahmed, J. (2016). Information security management needs more holistic approach: A literature review. *International Journal of Information Management, 36*(2), 215-225. doi:10.1016/j.ijinfomgt.2015.11.009

Stanley, J. D. (2011). Is the audit fee disclosure a leading indicator of clients' business risk? *Auditing: A Journal of Practice & Theory, 30*(3), 157-179.

Steinbart, P. J., Raschke, R., Gal, G., & Dilla, W. N. (2016). The organizational benefits of a good relationship between the internal audit and information security functions. *Working Paper*.

Steinbart, P. J., Raschke, R. L., Gal, G., & Dilla, W. N. (2012). The relationship between internal audit and information security: An exploratory investigation. *International Journal of Accounting Information Systems, 13*(3), 228-243. doi:10.1016/j.accinf.2012.06.007

Steinbart, P. J., Raschke, R. L., Gal, G., & Dilla, W. N. (2013). Information security professionals' perceptions about the relationship between the information security and internal audit functions. *Journal of Information Systems, 27*(2), 65-86.

Tysiac, K. (2017). A new cybersecurity risk management reporting framework for management and CPAs. *Journal of Accountancy*.

United States Senate. (2013). A "Kill Chain" Analysis of the 2013 Target Data Breach.

Vance, A., Lowry, P. B., & Eggett, D. (2013). Using accountability to reduce access policy violations in information systems. *Journal of Management Information Systems, 29*(4), 263-289. doi:10.2753/Mis0742-1222290410

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons, 29*(2), 381-396.

Verizon. (2016). 2016 Data Breach Investigations Report.

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics, 32*(1), 97-180.

Von Solms, B. (2005). Information Security Governance–compliance management vs operational management. *Computers & Security, 24*(6), 443-447.

Wallage, P. (2000). Assurance on sustainability reporting: an auditor's view. *Auditing: A Journal of Practice & Theory, 19*(1), 53-65.

Wang, T., & Cuthbertson, R. (2015). Eight issues on audit data analytics we would like researched. *Journal of Information Systems, 29*(1), 155-162.

Wang, T., Kannan, K. N., & Ulmer, J. R. (2013). The association between the disclosure and the realization of information security risk factors. *Information Systems Research, 24*(2), 201-218. doi:10.1287/isre.1120.0437

Wang, T., Ulmer, J. R., & Kannan, K. (2013). The textual contents of media reports of information security breaches and profitable short-term investment opportunities. *Journal of Organizational Computing and Electronic Commerce, 23*(3), 200-223. doi:10.1080/10919392.2013.807712

Warren Jr, J. D., Moffitt, K. C., & Byrnes, P. (2015). How Big Data will change accounting. *Accounting Horizons, 29*(2), 397-407.

Yan, Z., Moffitt, K. C., & Vasarhelyi, M. A. (2017). Automate contract analysis in auditing. *Working Paper*.

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

Yoon, K., Hoogduin, L., & Zhang, L. (2015). Big data as complementary audit evidence. *Accounting Horizons, 29*(2), 431-438.

Zafar, H., Ko, M. S., & Osei-Bryson, K.-M. (2015). The value of the CIO in the top management team on performance in the case of information security breaches. *Information Systems Frontiers*, 1-11.

Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective Big Data analysis in continuous auditing. *Accounting Horizons, 29*(2), 469-476.

Zhang, L., Pawlicki, A. R., McQuilken, D., & Titera, W. R. (2012). The AICPA assurance services executive committee emerging assurance technologies task force: The audit data standards (ADS) initiative. *Journal of Information Systems, 26*(1), 199-205.

APPENDICES

Appendix A: Variable Definitions for Chapter 2

Variable	*Definition*
LogAudit	Natural log of audit fees for the fiscal year of the cyber incident;
Cyber-Incident	Indicator variable, equal to 1 if the firm experiences a cyber incident during fiscal year t, and 0 otherwise;
Lnassets	Natural log of total assets in millions;
InvRec	Sum of inventory and accounts receivable divided by total assets;
Segments	Number of business segments;
Foreign	Indicator variable, equal to 1 if the firm has foreign operations (based on FCA), and 0 otherwise;
Merger	Indicator variable, equal to 1 if the firm was involved in merger activity during the fiscal year (based on AQP), and 0 otherwise;
Special	Indicator variable, equal to 1 if the firm was involved in merger activity during the fiscal year (based on SPI), and 0 otherwise;
Loss	Indicator variable, equal to 1 if the firm reported negative net income, and 0 otherwise;
Growth	One-year growth rate in sales;
Btm	Book value of common equity divided by market value of common equity;
Big4	Indicator variable, equal to 1 if the auditor is a member of the Big 4, and 0 otherwise;
GCO	Indicator variable, equal to 1 if the auditor issues a going-concern audit opinion in year t, and 0 otherwise;
Initial	Indicator variable, equal to 1 if an auditor change occurred during the fiscal year, and 0 otherwise;
ROA	Operating income after depreciation divided by total assets;
Leverage	Total liabilities divided by total assets;
Quick	Current assets minus inventories divided by current liabilities;
ICW	Indicator variable, equal to 1 if the auditor reports an internal control weakness, and 0 otherwise;
Busy	Indicator variable, equal to 1 if the auditee’s fiscal year ends in December, and 0 otherwise;
Residual	Represents the prior-period unexpected audit fees measured as the residual from yearly estimations of the basic audit fees model (Equation (2))
Δ*Cyber-Incident*	Indicator variable, equal to 1 if the firm experiences a cyber incident during fiscal year t but not in year t-1, and 0 otherwise;
Δ*Non_Cyber-Incident*	Indicator variable, equal to 1 if the firm experiences a data breach (not involving hacking) during fiscal year t but not in year t-1, and 0 otherwise;
Disclosure	Indicator variable, equal to 1 if the firm has cybersecurity risk disclosure in year t-1, and 0 otherwise;
Past_Breach	Indicator variable, equal to 1 if the firm had any cyber incident prior to year t, and 0 otherwise;
IP	Indicator variable, equal to 1 if the cyber incident involves intellectual property, and 0 otherwise;
INST	Percentage of institutional ownership of shares outstanding;
NUM	Number of block institutional ownerships that have larger than 5% shares outstanding; and
Δ	One-year change in the level of each variable.

Appendix B: Keywords for Identifying Cybersecurity Risk Disclosure

encryption
computer (virus|breach|break-in|attack|security)
security (breach|incident)
(information|network|computer) security
intrusion
hacking|hacker
denial of service
cyber(-|)(attack|fraud|threat|risk|terrorist|incident|security)
cyber-based attack
cybersecurity
infosec
system security
information technology (security|attack)
data theft
phishing
malware
data confidentiality
confidentiality of data
confidential data
unauthorized access
data corruption
corruption of data
network break-in
espionage
cyber(-|)insurance
data breach
crimeware
ransomware
keylogger
keystroke logging
social engineering

Appendix C: Variable Definitions for Chapter 3

Variable	*Definition*
Breach	Indicator variable, equal to 1 if the firm experiences cyber incident(s) during fiscal year t, 0 otherwise;
Past_breach	Indicator variable, equal to 1 if the firm experiences cyber incident(s) in any year preceding fiscal year t, 0 otherwise;
Disclosure	Indicator variable, equal to 1 if the firm has cybersecurity risk disclosure in fiscal year t, 0 otherwise;
Length	Total number of words in cybersecurity risk disclosure in fiscal year t, normalized by the average number of words in individual risk factors;
Size	Natural log of total assets in millions in fiscal year t;
LN_Segments	Natural log of number of business and geographic segments in fiscal year t;
Age	Number of year firms are included in Compustat in fiscal year t;
Loss	Indicator variable, equal to 1 if the firm reported negative net income in fiscal year t, 0 otherwise;
LN_Analyst	Natural log of number of analysts following in fiscal year t;
Foreign	Indicator variable, equal to 1 if the firm has foreign operations (based on FCA) in fiscal year t, 0 otherwise;
Merger	Indicator variable, equal to 1 if the firm was involved in merger activity in fiscal year t (based on AQP), 0 otherwise;
Growth	One-year growth rate in sales in fiscal year t;
ICW	Indicator variable, equal to 1 if the auditor reports an internal control weakness in fiscal year t, 0 otherwise;
Finance	Indicator variable, equal to 1 if the firm operates in finance industry (i.e. SIC between 6000 and 6999);
Consumer	Indicator variable, equal to 1 if the firm operates in consumer goods industry (i.e. SIC between 5200 and 5999);
Guidance	Indicator variable, equal to 1 after 2011, 0 otherwise;
Market_cap	Natural log of market capitalization of common stock in fiscal year t;
Severity	Indicator variable, equal to 1 if the cyber incident involve hacking by third parties, 0 otherwise;
Leverage	Total liabilities divided by total assets in fiscal year t;
Btm	Book value of common equity divided by market value of common equity in fiscal year t;
Score	One minus the cosine similarity score between firm's cybersecurity risk disclosure and industry's average disclosure for fiscal year t, adjusted by length using Taylor expansion proposed by Brown and Tucker (2011)
Informativeness	Percentage of unique words that are not used by any other firms in the same industry for the same fiscal year

Appendix D: Risk Factor Extraction

All available 10-K filings filed between January 2005 to December 2015 from the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system are first downloaded because risk factor disclosures were mandated on December 2005. Similar to Campbell et al. (2014) and Gaulin (2017), the procedure for extracting risk factor disclosure (i.e., ITEM 1A) is based on the assumption that 10-K filings in HyperText Markup Language (HTML) format contain visual clues (e.g., emphasis or whitespace separation) for readers to easily recognize item boundaries. The HTML filings are parsed into a tree structure using Beautifulsoup package in Python. The leaf nodes of the tree are textual information while the internal nodes of the three are HTML tags that can be used for identifying headings. For example, tag <p> defines a paragraph that is visually separated and isolated from text below and above. By assuming items are presented in order, I iterate all the HTML tags that contain the text "ITEM 1A", "ITEM 1B", "ITEM 2" (case insensitive). From all the candidates, the ones that are emphasized are identified (i.e., the ones include tag 'b', 'em', 'strong', 'h1', 'h2', 'h3', 'h4', 'h5', 'h6', 'u', 'p', 'font', 'div', 'span', or 'li' if using HTML emphasis tags, or 'bold', 'italic', or 'underline' if using Cascading Style Sheets within HTML tags). For all the candidates that satisfy the emphasis criteria, I identify their first parent node that is one of the following: 'h1', 'h2', 'h3', 'h4', 'h5', 'h6', 'p', 'div', 'ul', 'ol', and 'table'. For the ones that are not separated by 'table', I obtain the plain text in the separated paragraph which contains the phrase "ITEM 1A RISK FACTOR" without any other words. For the ones that are separated by 'table', I gather the entire row and obtain the plain text in the entire row which contains the phrase "ITEM 1A RISK FACTOR" without any other words.

Following the procedure, a list of elements that contain the headers for Item 1A, Item 1B or Item 2 is obtained. Risk factor disclosures are identified by extracting all the contents between the first Item 1A header and the first Item 1B or Item 2 header (in case there is no Item 1B). Individual risk factors are also extracted using HTML tags, similar to the approach used in Gaulin (2017). The SEC requires that each risk factor should be preceded by a subcaption that summarizes the risk. These subcaptions are identified based on such requirement: i.e., they are emphasized (bold, underline, or italic), and are at the beginning of each paragraph or isolated on its own line. The identified subcaptions are further filtered by applying a threshold (i.e., there are at least 10 words below that subcaptions). Contents between subcaptions represent individual risk factors.

Appendix E: Keywords and Phrases

Keywords to Identify Cybersecurity Risk Disclosures

encryption
computer (virus|breach|break-in|attack|security)
security (breach|incident)
(information|network|computer) security
intrusion
hacking|hacker
denial of service
cyber(-|)(attack|fraud|threat|risk|terrorist|incident|security)
cyber-based attack
cybersecurity
infosec
system security
information technology (security|attack)
data theft
phishing
malware
data confidentiality
confidentiality of data
confidential data
unauthorized access
data corruption
corruption of data
network break-in
espionage
cyber(-|)insurance
data breach
crimeware
ransomware
keylogger
keystroke logging
social engineering

Phrases to Identify Topics (Stemmed)

Lawsuit and Litigation 'addit-regulatori', 'applic-law', 'civil-crimin', 'civil-litig', 'compli-applic', 'compli-law', 'complianc-cost', 'contractu-oblig', 'crimin-penalti', 'enforc-action', 'expo-civil', 'expo-litig', 'fail-compli', 'failur-compli', 'feder-state', 'fine-penalti', 'govern-regul', 'law-govern', 'law-protect', 'law-regul', 'legal-claim', 'legal-liabil', 'legisl-regulatori', 'liabil-claim', 'liabil-law', 'litig-liabil', 'litig-regulatori', 'loss-litig', 'possibl-liabil', 'potenti-liabil', 'privaci-law', 'regulatori-action', 'regulatori-approv', 'regulatori-environ', 'regulatori-interv', 'regulatori-penalti', 'regulatori-requir', 'regulatori-scrutini', 'result-legal', 'result-litig', 'secur-law', 'signific-legal', 'state-feder', 'state-law', 'state-local', 'subject-litig', 'violat-applic'

Business Operations	'abil-conduct', 'abil-oper', 'abil-perform', 'act-vandal', 'affect-oper', 'busi-continu', 'busi-damag', 'busi-disrupt', 'busi-failur', 'busi-harm', 'busi-interrupt', 'caus-disrupt', 'caus-interrupt', 'compromis-network', 'compromis-secur', 'comput-equip', 'comput-hardwar', 'comput-network', 'comput-telecommun', 'conduct-busi', 'continu-oper', 'continu-plan', 'creat-disrupt', 'critic-busi', 'damag-disrupt', 'damag-failur', 'damag-interrupt', 'deliv-product', 'denial-servic', 'disast-power', 'disast-recoveri', 'disast-terror', 'disast-terrorist', 'disrupt-busi', 'disrupt-compani', 'disrupt-inform', 'disrupt-oper', 'disrupt-servic', 'disrupt-shutdown', 'effect-oper', 'electr-telecommun', 'enterpri-resourc', 'experi-interrupt', 'failur-disrupt', 'failur-interrupt', 'failur-network', 'hardwar-failur', 'harm-oper', 'impact-oper', 'infrastructur-vulner', 'intern-control', 'intern-oper', 'internet-telecommun', 'interrupt-busi', 'interrupt-failur', 'interrupt-malfunct', 'interrupt-oper', 'interrupt-power', 'interrupt-servic', 'jeopard-secur', 'loss-telecommun', 'malfunct-oper', 'materi-disrupt', 'network-disrupt', 'network-failur', 'network-infrastructur', 'oper-disrupt', 'oper-failur', 'oper-infrastructur', 'oper-interrupt', 'penetr-network', 'power-loss', 'power-outag', 'properti-damag', 'resourc-plan', 'result-disrupt', 'result-interrupt', 'servic-attack', 'servic-disrupt', 'servic-interrupt', 'signific-disrupt', 'signific-interrupt', 'similar-disrupt', 'softwar-hardwar', 'softwar-network', 'subject-disrupt', 'suppli-chain', 'technolog-disrupt', 'technolog-fail', 'technolog-failur', 'technolog-infrastructur', 'technolog-network', 'telecommun-failur', 'telecommun-outag', 'transmiss-distribut', 'uninterrupt-oper'
Reputation	'abil-attract', 'affect-reput', 'attract-new', 'attract-retain', 'busi-reput', 'compani-reput', 'custom-relationship', 'damag-brand', 'damag-reput', 'effect-reput', 'harm-reput', 'impact-reput', 'negat-public', 'relationship-custom', 'relationship-manag', 'reput-brand', 'reput-damag', 'reput-expo', 'reput-financi', 'reput-harm', 'reput-loss', 'reput-suffer'
Intellectual Property	'competit-posit', 'intellectu-properti', 'proprietari-busi', 'research-develop', 'trade-secret'
Financial Performance	'addit-cost', 'addit-resourc', 'affect-financi', 'capac-constraint', 'capit-expenditur', 'capit-resourc', 'cash-flow', 'common-stock', 'compen-loss', 'decreas-revenu', 'effect-financi', 'financi-condit', 'financi-liabil', 'financi-loss', 'financi-oper', 'financi-perform', 'financi-posit', 'financi-result', 'impact-financi', 'increas-cost', 'increas-expen', 'incur-liabil', 'loss-liabil', 'loss-revenu', 'lost-revenu', 'oper-cash', 'oper-cost', 'oper-expen', 'oper-financi', 'proceed-liabil', 'reduc-revenu', 'remedi-cost', 'revenu-profit', 'signific-capit', 'signific-cost', 'signific-expen', 'signific-invest', 'signific-liabil', 'signific-loss', 'substanti-cost', 'suffer-loss'

Appendix F: Variable Definitions for Chapter 4

Variable	*Definition*
abn_email	Indicator variable, equal to 1 if the number of emails exchanged on that day is large than 1.5 times the average number of email exchanged over the past 30 days, 0 otherwise;
abn_file	Indicator variable, equal to 1 if the number of files transferred to external device on that day is large than 1.5 times the average number of files transferred to external device over the past 30 days, 0 otherwise;
abn_hours	Indicator variable, equal to 1 if the total hours of logon activity on that day is large than 1.5 times the average total hours of logon activity over the past 30 days, 0 otherwise;
abn_device_in	Indicator variable, equal to 1 if the number of external device connect activity on that day is large than 1.5 times the average number of external device connect activity over the past 30 days, 0 otherwise;
abn_attachment	Indicator variable, equal to 1 if the number of email attachments on that day is large than 1.5 times the average number of email attachments over the past 30 days, 0 otherwise;
abn_after_log	Indicator variable, equal to 1 if the user has logon activity in after hours (after 6:30 pm) on that day and has less than 10 logon activities in after hours over the past 30 days, 0 otherwise;
abn_pc_log	Indicator variable, equal to 1 if the user has logon activity in other's workstation on that day and has less than 10 logon activities in other's workstation over the past 30 days, 0 otherwise;
abn_after_device	Indicator variable, equal to 1 if the user has external device connect activity in after hours (after 6:30 pm) on that day and has less than 10 external device connect activities in after hours over the past 30 days, 0 otherwise;