THE UNIVERSITY OF CHICAGO

DO AUDITORS HELP PREVENT DATA BREACHES?

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

LISA YAO LIU

CHICAGO, ILLINOIS

JUNE 2020

Table of Contents

List of Figures

List of Tables

Acknowledgements

I greatly appreciate the guidance and support of my dissertation committee: Philip G. Berger, Hans Christensen (chair), Christian Leuz, Mark Maffett, and Mike Minnis. I thank Ionela Andreicovici, Christopher Armstrong, Ray Ball, John Barrios, Jonathan Bonham, Matthias Breuer, Anna Costello, Dhammika Dharmapala, Raphael Duguay, John Gallemore, João Granja, Luzi Hail, Chris Hansen, Katharina Hombach, Jonas Jonasson, Emir Kamenica, Sehwa Kim, Anya Kleymenova, Kalin Kolev, Eva Labro, Rebecca Lester, Jinzhi Lu, Charlie McClure, Maximilian Muhn, Anya Nakhmurina, Valeri Nikolaev, Matthew Plosser, Thomas Rauter, Doug Skinner, Abbie Smith, Lior J. Strahilevitz, Andrew Sutherland, Chad Syverson, Rimmy Tomy, Felix Vetter, Anastasia Zakolyukina, Anthony Lee Zhang, Luigi Zingales, and participants at the 2019 AAA/Deloitte Foundation/J. Michael Cook Doctoral Consortium, the 2019 CMU Accounting Mini Conference emerging scholar session, Columbia University, London Business School, London School of Economics, New York University, Northwestern University, Rochester University, Stanford University, the University of Chicago, the University of Pennsylvania, and University of Toronto for helpful comments and suggestions. I also benefited from discussions with fellow PhD students at the University of Chicago: Ehsan Azarmsa, Mihir Gandhi, Kalash Jain, Miao Liu, Shirley Lu, Yao Lu, Johanna Shin, Gurpal Sran, Ana-Maria Tenekedjieva, James Traina, and Lauren Vollon. I thank Kamay Lafalaise (the attorney in the Office of the General Counsel from the Federal Trade Commission) for providing me with the FTC data information. I thank Martha Van Haitsma (the co-director of the University of Chicago Survey Lab) and Sona Margaryan (director of Strategic Initiatives at the University of Chicago) for their help with my survey design. I thank Brandon Gipper for sharing the PCAOB inspection data with me. I acknowledge research support from the Accounting Research Center at the University of Chicago Booth School of Business and thank Daniel Chavez, Cagdas Okay, Georgios Tzortzis, Benjamin

Levine, Chanho Moon, and Dhuv Baid for double checking my data matching process. I gratefully acknowledge financial support from the University of Chicago Booth School of Business, the Deloitte Foundation, and the Bradley Fellowship awarded by the Stigler Center for the Study of the Economy and the State. This study was exempt from further review by the Institutional Review Board at the University of Chicago (IRB19-1066), under Federal Regulation (45 CFR 46.101(b)).

I also appreciate discussions with various industry professionals, including: Mark Lavalle, audit partner at KPMG; Len Jui, board member of IAASB and partner at KPMG; Jodilia Vasanji, managing director at Deloitte & Touche LLP; Paulo Blanc, IT audit supervisor at ArcelorMittal; Rashesh Patel, principal examiner, IT risk and controls at FINRA; Vishal Dalal, consulting, internal audit, and SOX professional at Vonya Global; Rong Liu, IT audit, compliance, risk and internal control at Wolters Kluwer; Vincent Banks, CPA, CGMA, vice president internal audit at GreenSky®; Chris G Nicholson, FCA CPA, audit committee member; Dan Gaffney, MBA, CPA, CIA, CISA, internal audit and IT audit consultant; and many anonymous audit partners, internal auditors, external auditors, and legal counsels.

Abstract

With the increase in digital automation of financial statements and computer-based audit evidence, I examine whether and how auditors help prevent data breaches. I use two plausibly exogenous shocks (regulation based and learning-experience based) and find that improvements in auditing reduce the likelihood of data breaches. I then conduct interviews and an anonymous survey to collect information on mechanisms that are not captured by the empirical analyses. Consistent with a complementary relationship between auditors and their clients in preventing data breaches, I find the effect is larger in firms with more integrated data systems, a greater percentage of board members on the audit committee, and stronger internal controls. Overall, these results support a disciplining effect of auditing processes (e.g., audit procedures and testing) on reducing a new kind of agency friction between firms and data providers.

1. Introduction

With the advancement of information and communication technology (especially AI and big data), firms are motivated to collect more data due to the declining costs of data collection and storage (Goldfarb and Tucker 2012). This also creates new risks: specifically, data breaches.[1] As documented in prior research, breaches harm a variety of economic agents, such as financial institutions (Duffie and Younger 2019), consumers (Liu 2019), and the industry peers of breached companies (Haislip et al. 2019). On the firm side, data breaches are associated with reputation loss (e.g., Gwebu et al. 2018; Syed 2019), increased cost of debt (Sheneman 2018), and financial reporting deficiencies (Lawrence et al. 2018). Given the *ex post* consequences of data breaches on the aggregate economy, efforts to understand factors that prevent data breaches are of considerable practical and economic importance. In this paper, I document an *ex ante* disciplining effect of the auditing process and investigate the following question: Do auditors help prevent data breaches?

Auditors are not only concerned with final financial numbers, but also with how these numbers are generated and with the underlying data.[2] With digital automation (DeFond and Zhang 2014; Yoon et al. 2015) and with the substantial growth of intangible assets in the "New Economy" (Haskel and Westlake 2018; De Ridder 2019), many financial numbers are automated by information systems and much of the audit evidence is computer-based (e.g., Efendi, Mulig, and Smith 2006; Alves 2010; Yoon et al. 2015; Brands and Smith 2016). The auditing standard AS5

[1] For instance, the 2013 Yahoo! Breach affected three billion users, and the 2014 Yahoo! hack impacted roughly 500 million people. Similarly, a 2013 data breach at Target exposed 40 million customers' credit card information. In 2017, an Equifax data breach compromised about 143 million US consumer accounts; Ponemon Institute (June 2017) estimate that the average global cost of data breach per lost or stolen record was $141 in 2017. Insiders also sold shares before the Equifax breach was revealed to investors (Robert J. Jackson 2018).

[2] For example, *Assure Professional* (April 29, 2014) discussed the relationship between audit and target data breach, saying that "[T]he massive data breach that Target incurred this winter was a textbook example of why audits are so important, especially when it comes to financial data." *MarketWatch* (October 3, 2017) reported that "Equifax auditors are on the hook for data security risk controls" and that "[B]efore an auditor reviews numbers, it must make sure that execs set the risk 'tone at the top' on controls, including of IT systems."

explicitly requires that "the *identification of risks and controls within IT is not a separate evaluation*. Instead, it is *an integral part of the top-down approach* [emphasis added]" in an integrated audit. In the risk-based audit procedure, IT auditors assist the financial statement audit to verify financial statements, including the IT general controls (applying to all aspects of the IT function and the overall integrity of the system) and the IT applications controls (applying to business processing transactions in individual computer application programs).

In the process of verifying financial statements, auditors examine data on firms' economic transactions in order to ensure that it is linked to information in financial statements. For example, auditors care about whether there is unauthorized access to the underlying economic transaction data. If the access controls are not robust, transactions are vulnerable to being manipulated, endangering the accuracy and credibility of financial statements. Through the *ex post monitoring*, external auditors can also create *ex ante incentives* for firms to adopt high-quality internal controls (e.g., Hogan and Wilkins 2008; Altamuro and Beatty 2010; Barrios, Lisowsky, and Minnis 2018).[3]

Although the arguments above support the hypothesis that auditors help reduce the likelihood of data breaches, the effect could be too small to be empirically captured. For example, although auditors should test IT general and application controls, these tests may not be specific or technical enough, or auditors may not perfectly enforce the auditing standards. Additionally, some breaches are complex and sophisticated (e.g., hackers access systems), precluding defense against at least some attacks. Moreover, if firms realize the benefits of auditors, more effective data protection mechanisms may be crowded out or substituted, which diminishes the effects even

[3] Internal control has four components: (1) control environment; (2) risk assessment, including a company's ability to maintain proper records and accurate financial data; (3) control activities, such as information processing (e.g., IT environment that includes access controls and passwords verification) and physical controls over the access to assets; (4) information and communication. When auditors document internal controls, auditors look for components of internal controls such as information processing, physical controls, recording, authorization, independent checks, and segregation of duties.

further. Thus, the degree and the magnitude to which auditors help prevent data breaches remains an empirical question.

To explore whether auditors help prevent data breaches, I first provide descriptive evidence on the negative relation between the likelihood of data breaches and audit quality. While this descriptive evidence is consistent with my hypothesis, endogeneity concerns could undermine my findings. For example, firms that are concerned with preventing breaches may select better auditors, and firms that differ with respect to audit quality are likely to vary along other unobservable dimensions. The key to identification is disentangling firms' responses to auditing shocks from their endogenous responses. An ideal natural experiment would be to randomly vary the quality of audit services while holding the firm constant.

To mitigate endogeneity concerns, I use two shocks (regulation based and learning-experience based) to exploit plausibly exogenous variation on the supply side of audit services. The regulation-based shock is the first-time inspection fieldwork conducted by the Public Company Accounting Oversight Board (PCAOB). Prior research finds that PCAOB inspections improve the quality of internal control audits, facilitate auditors' learning, and improve audit quality (e.g., DeFond and Lennox 2017; Aobdia and Shroff 2017; Aobdia 2018; Gipper, Leuz, and Maffett 2019; Hanlon and Shroff 2019). Inspections assess the audit process (e.g., they test internal controls) (DeFond and Lennox 2017); one important aspect of these inspections is assessing auditors' testing of their clients' IT general controls (PCAOB 2010 and 2013).[4] The initial PCAOB inspections are staggered across different auditors over time. I find that after auditors are inspected

[4] PCAOB specifically requires that auditors have an "understanding of how the organization is dependent on or enabled by *information technologies; and the manner in which information systems are used to record and maintain financial information* [emphasis added]" (PCAOB, QC Section 40, 2003). The PCAOB (2013) Stuff Audit Practice Alert No. 11 lists information technology (IT) considerations (such as system-generated data and reports) as a significant, frequently cited auditing deficiency in PCAOB inspection reports.

by the PCAOB, their clients are less likely to have data breaches as compared to those not yet inspected.

The second is a learning experience-based shock: auditors learning from their mistakes. If an auditor's client experiences an incident such as restatements or data breaches, the auditor will update its belief about its own examination process as well as the consequences of these incidents, which, in turn, impacts auditing (e.g., Weber et al. 2008; Li et al. 2017; Haislip et al. 2019). Thus, auditors learn from these incidents and strengthen their audits in order to avoid future failures. Different auditors will experience this "learning" at different times, creating a staggered variation. For this empirical analysis, I use auditor-client data and find, within the same auditor, that an audit failure leads to fewer data breaches among other clients.

An important caveat is that researchers cannot directly observe firms' data protection processes. Thus, I complement prior analyses by conducting 36 interviews (11 with accounting firm partners) and by creating an anonymous survey for industry professionals to obtain institutional insights and to collect information on mechanisms that are not captured by empirical analyses. (The remaining interviewees include five non-partner external auditors, nine internal auditors, one audit committee member, five corporate legal counsels/experts, and five regulators. See the appendix for a brief summary of interviews and surveys.) The ample anecdotes provided by interviewees support auditors' role in preventing data breaches: Providing relevant information and incentives for internal controls. In the survey response, ninety percent of participants think IT audit and internal control tests help protect firms' financial reporting data. Seventy-seven percent of auditors think these tests also help protect firms' personal data (e.g., consumer and employee information). Auditors also believe that their processes (e.g., testing to assess management controls and requests to change management controls) could help companies discover vulnerabilities in

their information systems.

To formalize the framework for the mechanism, I outline a simple stylized model (in the appendix) with two cross-sectional predictions: (1) in more integrated data systems, information transfers from audited financial data systems to unaudited data systems are more likely; (2) when firms are more receptive to auditors' ex post monitoring (as proxied by firms' audit committees and internal control strength), ex ante incentives for internal controls are stronger, which suggests a complementary relationship between auditors and their clients in preventing data breaches.[5] Consistent with this relationship, I find that the reduction in the likelihood of data breaches is larger in firms with more integrated data systems, a greater percentage of board members on the audit committees, and stronger internal controls. While these tests are imperfect and subject to endogeneity concerns, the sample size is large and the tests provide informative descriptive evidence.

This paper makes three distinct contributions. First, this paper contributes to the emerging literature on data breaches. Prior literature focuses mainly on the *ex post* consequences of data breaches. Kamiya et al. (2018) show that breached firms respond to online attacks by reducing CEO risk-taking incentives and by strengthening risk management. Duffie and Younger (2019) discuss the negative economic impact of cyber attacks on deposits at financial institutions; these attacks could lead to cyber runs and could turn an operational event into a liquidity event. Sheneman (2018) finds that data breaches increase firms' cost of debt. In regard to investors, Kannan et al. (2007) and Campbell et al. (2003) find mixed evidence regarding market reactions

[5] Because of this incentive alignment between firms and auditors, the effect and magnitudes could be larger than for the other misconducts (e.g., embezzlement and bribery). Additionally, data breaches are empirically observable; unlike outcomes such as investment efficiency, we do not need an empirical estimation. It is also mostly objective (as compared to outcomes such as accounting estimates), making my results less susceptible to concerns about incentive misalignment between firms and auditors.

to data breaches, but their tests could have low power due to small sample size. Haislip et al. (2019) find that non-breached industry peers experience negative capital market reactions and have higher audit fees. I contribute to the literature by documenting whether and how auditors help prevent data breaches, an *ex ante* disciplining effect.

Second, it adds to the literature on the boundaries of auditing services. The evidence in this paper shows that auditing benefits the stakeholders more than has been previously documented. In addition to the audited outcome (i.e., information value), the audit process (i.e., audit procedures and testing) can also benefit economic stakeholders. Prior literature focuses mainly on how stakeholders use verified financial statements to make economic decisions (e.g., Mansi et al. 2004; Minnis 2011; Duguay 2019); the effect I study, however, is not concerned with the use of financial statements but with improvements in the data systems and controls (the benefit of the verification *process*). These improvements also help reduce an important agency friction between firms and *data providers* (e.g., consumers and employees): specifically, that firms may commit to a consumer-friendly data usage policy ex ante that is difficult to enforce ex post (Jin 2018).[6]

Finally, by showing that the auditing process enhances the security of the data systems and controls, I complement the literature on the benefit of internal controls. The literature studies how the improvement in internal controls can benefit internal managerial decisions, leading to more accurate management forecasts (Feng et al. 2009), increased investment efficiency (Cheng et al. 2013), and effective inventory management (Feng et al. 2015). I contribute to the literature by showing that the auditing process directly disciplines other control systems (e.g., IT infrastructure)

[6] Consistent with data providers' negative reactions to data breaches, Liu (2019) finds a chilling effect on consumer spending; for example, Target's 2013 breach undercut quarterly sales and ultimately cost the CEO his job. However, another possibility is that some consumers understand the risk that their personal data might be compromised (with a certain percentage probability) when they provide it to a given set of firms. If this is the case, ex post breaches may match consumers' ex ante expectations. Auditing would still likely reduce data breaches in such a scenario.

to benefit economic stakeholders. Additionally, I document a causal effect of auditing on preventing data breaches and estimate the magnitude of these effects.

2. Institutional Background and Conceptual Development

2.1 Auditing Processes and Data Protection

With many manual processes and documentation moving to the digital world, more financial numbers are now automated by the information systems and much of the audit evidence is becoming computer-based (e.g., Efendi, Mulig, and Smith 2006; Alves 2010; Yoon et al. 2015; Brands and Smith 2016). Thus, auditors must understand and test these data controls before they can conclude that the automated information is reliable.[7] Auditors have long been concerned with physical assets (e.g., counting inventory); facing the increasing digital technology, auditors now also care about intangible and digital assets (e.g., customer lists).

The auditing standard AS5 requires that "the *identification of risks and controls within IT is not a separate evaluation*. Instead, it is *an integral part of the top-down approach* [emphasis added]" in an integrated audit.[8] In the risk-based audit procedures, IT auditors assist the financial statement audit with verifying financial statement values. These auditors focus on *IT general controls* (which relate to the overall integrity of the system and apply to all aspects of the IT function, such as file security, access controls, data/program access changes, new system developments, current system changes, and computer operations) and *IT application controls* (which apply to business processing transactions like sales or cash receipts and which test the

[7] See https://www.cpapracticeadvisor.com/home/article/10263076/the-evolution-of-technology-for-the-accounting-profession. Information systems consist of the methods and records used to record, process, summarize and report company's transactions and to maintain accountability for the related accounts. AICPA issued SAS No. 48 (*The Effects of Computer Processing on the Examination of Financial Statements*) in 1984 because IT could impact the nature, timing, and extent of audit procedures (Yang and Guan 2004; Hoffman et al. 2018). Please also see the appendix for concrete auditing procedures on IT controls. I provide additional institutional discussions and empirical evidence in Appendix A3 to demonstrate that auditors have the relevant skills to test data protection controls. If a client's IT systems are too complex and specialized, IT specialists are invited to assist the auditing process.

[8] An integrated audit combines a financial statement audit with an audit of internal controls.

performance of individual computer application programs, such as accepting authorized input, correct processing, and generating the appropriate output).[9] IT general controls need to be effective so that auditors can rely on the IT application controls. Auditors then design and implement the scope of substantive test procedures based on the results of the internal control tests. If the system does not operate effectively, the need for substantive procedures will increase in order to decrease the detection risk.

Prior research also finds that external auditors' IT expertise helps firms strengthen internal controls and discover vulnerabilities in their information systems (e.g., Haislip et al. 2016). When assessing the risk of material misstatement in financial statements, auditors are required to consider a company's IT systems and controls, including the IT risks stemming from unauthorized access (e.g., Auditing Standards No. 12 Appendix B; Center for Audit Quality 2016 and 2017; Li et al. 2012; Schroeder and Shepardson 2015); the rationale is that if a firm's data access control is not robust, transactions are vulnerable to being manipulated and the reliability of the financial statements is compromised. Data controls that are intended to make numbers accurate and reliable can also help make the numbers secure.[10] This means that auditors are not only concerned with the final financial numbers, but also with the process of generation and the data that underlie these financial numbers.

Generally, when firms have a significant amount of information that is authorized,

[9] The transaction processing the immediate processing of transactions and batch processing (e.g., gathering information as a group to the computer periodically). Access controls include those designed to protect the information from unauthorized access. Auditors could specifically test passwords and firewalls to prevent outside threats.

[10] Small firms (e.g., non-accelerated filers) have less stringent internal control testing by auditors (SOX 404b), but their scope for improvement is larger. Non-financial information systems would also be affected. For example, in the analytical procedure, when auditors verify payroll expenses, they also test the HR systems to confirm the number of employees and salary. Even if companies outsource data services, external auditors will examine their clients' physical security and request that a System and Organization Controls (SOC) report be obtained from the vendor. This ensures that vendors have data protection controls and that those controls map onto clients' control systems. Schoenfeld (2020) provides descriptive evidence on SOC audits in the era of big data.

recorded, processed, or reported electronically, substantive procedures alone are not sufficient, requiring tests of the operating effectiveness of controls. Even for substantive tests in the financial statement audit, auditors use the computer to perform tests when they have access to the client's data. The computer can be used to examine a client's data for validity, completeness, and accuracy; select client data for audit samples; and to compare similar data in client's files to identify discrepancies; and compare the results of audit procedures to the client's data. Additionally, in a financial statement audit, auditors need to understand the design of and to assess the internal controls, and must also search for strengths of controls for auditors to rely on. If auditors plan to rely on a specific control, they should test the control to determine whether it operates effectively.

Given the considerable economic consequences of data breaches (e.g., Duffie and Younger 2019, Haislip et al. 2019, Sheneman 2018, Lawrence et al. 2018), can auditing procedures and tests have a disciplining effect that helps prevent data breaches? My goal in this paper is to document a causal effect of auditing on preventing data breaches and to estimate the magnitude of such effects.

2.2 Cross-sectional Mechanism Discussions

During the auditing process (ex post monitoring), auditors provide relevant and useful information through the intertwined relation in firms' data systems and through interactions with firms' audit committees. Facing auditors' ex post monitoring, firms' ex ante incentives to adopt high-quality internal controls are strengthened. These high-quality internal controls serve the entire organization, including information process and IT environment.

Because data systems are intertwined, auditors can provide relevant and useful information that alerts firms to potential vulnerabilities in their IT infrastructure. For example, auditors may not assess IT general controls in isolation but in combination with other data control systems,

especially if those controls affect related accounts (PCAOB 2013). Although auditors focus on material risks, given the same control system, the data below the materiality threshold could also be protected through auditing.

The Sarbanes-Oxley Act requires that audit committees oversee compliance, risk management, and internal controls for companies' financial reporting. Two salient examples are the reactions of proxy advisors ISS and Glass Lewis to the Target breaches and to Facebook's use of data, respectively. At the annual shareholder meeting in 2014, ISS vetoed the election of all Target audit committee members because they failed to fulfill the responsibility of risk assessment. Glass Lewis and other institutional investors argued that Facebook's audit committee neglected to oversee the risk and compliance related to Facebook users' data. Several communication channels exist between audit committees and external auditors. An audit committee engages with the auditor throughout the entire audit—during planning, at interim reporting periods, and at year-end. Additionally, audit procedures should adapt to each company's unique IT environment and an auditor would discuss these changes with the audit committee and with management. Specifically, Auditing Standard No. 5 requires that auditors evaluate the severity of a control deficiency and communicate these deficiencies to the audit committee and to management in an integrated audit (Auditing Standards No. 13).

Auditors could also provide *ex ante incentives* for firms to adopt high-quality internal control systems and procedures through *ex post monitoring* (e.g., detecting deficiencies in internal controls, assessing the design of internal controls, and disclosing the quality of internal controls to external parties). When high-quality internal controls are in place, a firm's general data and information systems are strengthened (e.g., Hogan and Wilkins 2008; Feng et al. 2009; Feng et al. 2014; Altamuro and Beatty 2010; Barrios, Lisowsky, and Minnis 2018).

To provide a framework and formalize the intuition for the mechanism, a simple stylized model (in the appendix) illustrates two cross-sectional predictions: (1) in more integrated data systems, information transfer from audited financial data to other systems is more likely; (2) when firms are more receptive to auditors' ex post monitoring, the ex ante incentive for strengthening internal controls is stronger. Based on these predictions, I use proxy variables to conduct cross-sectional analyses on integrated systems, audit committees, and internal control weaknesses. In addition, I measure these proxy variables ex ante in order to capture the initial conditions and to mitigate the concern that these variables may be correlated with the shocks I exploit.

2.3 The Demand Side of Audit Services

Firms' alternative governance mechanisms (e.g., IT and risk committees; insurance companies) may arise endogenously to help mitigate the harm of data breaches, thus one underlying condition in this paper is that these alternative governance mechanisms alone do not fully resolve the issue. That is, firms' investment in controls and IT systems may not be significant enough to detect data issues in the absence of audit services. This could be possible for several reasons, such as information constraints and coordination failure.

First, frictions and constraints could prevent firms from providing sufficiently robust data protection. Bloom et al. (2013) argue that informational barriers are the main reason why firms fail to adopt quality control systems. Data breaches may be complex and sophisticated (e.g., if hackers have advanced skills), precluding a comprehensive defense. On the other hand, many practitioners find that "sometimes the root causes of a breach are weaknesses in basic security procedures," and that "it usually is the small things, such as not following simple procedures, that provide an intruder a door to gain access" (Internal Auditor (2014). For example, after an investigation into cyber frauds ("Business email compromises"), the SEC explained that these

frauds were not sophisticated in design or in use of technology; instead, it was "weaknesses in policies and procedures and human vulnerabilities that rendered the control environment ineffective" (2018 SEC Investigative Report). It is also important to differentiate between opportunistic attacks and targeted threats. While preventing targeted threats may be difficult, companies should be able to prevent opportunistic attacks. For example, Verizon's investigative team found malicious activity on the CSO's smartphone and laptop; this appeared to be an opportunistic rather than a targeted attack (CPA PracticeAdvisor 2007).[11] Thus, auditors' internal control tests (i.e., third-party interventions) could provide informative signals and could help firms discover vulnerabilities in their information systems.

Second, even if firms do not have any informational constraints, there could be a wedge between the social and the private costs of data breaches. For example, a coordination failure could exist because an economic harm (e.g., consumers losing trust) is spread across the whole market, giving individual firms little incentive to internalize system-wide risks (Kashyap and Wetherilt 2019). Kashyap and Wetherilt (2019) explain that the pricing and contracting mechanisms may not work for individual firms when these firms are exposed to common risks (including cyber risks). Therefore, compared to social planners, an individual firm has fewer incentives to provide sufficient data protections.

Because firms are ultimately responsible for data protection, they (the demand side) implement any of the changes recommended by auditors that provide direct or indirect benefits. The question is whether the requirements for IT testing from auditors (supply side) are enough to discipline firms' data systems and controls (demand side). As firms disclose the risk of data breaches in their financial statements (e.g., in the sections of risk factors, MD&A, and disclosure

[11] See: https://www.cpapracticeadvisor.com/news/12325862/how-accountants-can-help-clients-avoid-data-breaches

controls and procedures), auditors verify financial statement disclosures and consider whether the information therein is materially consistent with the financial statements or whether it is a material misstatement of fact.[12] This verification can "harden" information (e.g. Minnis 2011) and increase the cost of data breaches, thus helping discipline firms' data protection behavior. However, the inclusion of this information in financial reports may also change firm behavior through other alternative governance mechanisms (e.g., Leuz and Wysocki 2016; Christensen et al. 2017). Therefore, it is crucial to disentangle a firm's response to auditing shocks from a firm's organic responses (Heckman 1981). In this paper, I use two shocks (regulation based and learning-experience based) to exploit plausibly exogenous variation on the supply side of audit services.

3. Data

My information on data breaches comes from the Privacy Rights Clearinghouse (PRC). The data are available from 2005 and include data breaches and the number of compromised records (as reported by government agencies or verifiable media sources).[13] Although Kamiya et al. (2018) selects a random sample and verifies PRC's data in order to double check the reporting

[12] Because the regulatory enforcement of the 2011 SEC Disclosure Guidance (especially on expected breaches) is not strict and because disclosures only include cyber risk (and not insiders' unauthorized access), whether or not auditors can play a role in this disclosure channel remains unclear. Specifically, in 2011, the SEC issued a CF Disclosure Guidance on cyber-security risk disclosures. Prior to the Guidance, publicly traded companies were not required to report cyber risks in their SEC filings (though banks were required to report these risks to the Department of the Treasury and healthcare companies were required to report to the Department of Health and Human Services). However, this guidance was non-binding and the enforcement power was not significant. Thus, in 2018, the SEC issued new guidance and lowered the threshold for disclosure, requiring, for example, the disclosure of uncertainties and the likelihood of future incidents. In addition, the company's CEO and CFO must certify the disclosure controls and procedures or face the risk of enforcement action. In regards to the financial value treatment, firms are required to recognize and disclose the contingent liability after a breach (recognition if the loss is probable and the amount can be reasonably estimated; disclosure if it is reasonably possible—less than probable but more than remote—that a loss or expense will occur). For example, after the 2013 Target breaches, Target accrued liability related to credit card reissuance and other operating expenses. With respect to fines (or other penalties) from regulatory bodies, Target disclosed that the losses were reasonably possible but that the amount could not be reasonably estimated. However, this accounting treatment would occur *after* a breach happens (an actual breach) while the focus of this paper is *preventing* data breaches (an expected breach).

[13] http://www.privacyrights.org/data-breach. See https://www.privacyrights.org/data-breach-FAQ for detailed data information. One federal government source is the Department of Health and Human Services Office for Civil Rights (https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf) because it is the most well-structured, up to date information available (Privacy Rights Clearinghouse 2019).

accuracy, potential concerns, such as materiality thresholds for breach disclosures and breach detections, may still exist, leading to the over-/under-estimation of treatment effects. For example, if breaches are immaterial in the later period without disclosure, I may overestimate the treatment effect. Although breaches are defined annually in the paper, if a breach is not detected within a year and if more data breaches are detected over time, this would bias against my finding the treatment effect. Therefore, I discuss several cross-checks to alleviate these concerns. First, this data is collected by a third-party non-profit organization that is incentivized to actively search for and report data breaches in order to encourage public scrutiny and action, instead of misreporting or hiding breaches. Second, a simple descriptive assessment is used to ensure that the smallest firm size for breached and not-breached companies is similar. When I implement this restriction, only around 1.4% observations are dropped, and the results have stronger statistical power. Third, I match firms on variables correlated with data breaches and find that my results still hold (see the appendix for more information). Finally, my descriptive results are robust to the FTC datasets (which I accessed by invoking the Freedom of Information Act (FOIA)), which are comprised of customers' self-reports to the FTC about their identity theft. It is difficult for companies to hide these cases because they are reported by data providers who have incentives to report to the FTC shortly.[14]

Data breaches include hacking or malware by outsiders (as shown in Figure 1, this accounts for 25% of data breaches) and data mishandling by insiders (e.g., data digitally sent to the wrong party; intentional breach of information by someone with legitimate access; payment card fraud;

[14] The observed data breaches are the joint probability of data breaches occurring and being detected, which is equal to the unconditional probability of a data breach occurring multiplied by the probability of detection conditional on an occurrence. Therefore, I examine a case when breaches certainly getting detected (i.e., the probability of being detected conditional on an occurrence is equal to one) to help empirically assess concerns about detection and reporting biases. In addition, my results are robust to including the Audit Analytics dataset (Haislip et al. 2019).

physical loss of paper documents or portable device; stationary computers lost, inappropriately accessed, discarded, or stolen). If a firm's information system is vulnerable to outsiders, it is also vulnerable to insiders. Thus, both of types of data breaches are related to the vulnerability of firms' information systems and internal controls.[15] The PRC dataset is my primary data source. I manually match the data from PRC with public companies to get 1,214 observations with 524 unique firms. The simple descriptive statistics in Table 1 show that breached companies are relatively larger, less likely to experience a loss, and have higher asset intangibility (proportion of assets that are not PP&E) than do the other companies.

My firm-level financial data are from Compustat.[16] I manually match firms in the PRC dataset to public firm databases (e.g., Compustat, CRSP, and SEC filings) using firm names and also use a variety of other platforms (e.g., Bloomberg) to ensure that organizations are listed during the sample period.[17] My auditor data are from Audit Analytics. The sample consists of all US-listed firms in the Audit Analytics database from 2004 to 2016. For each firm-year observation, I collect the firm's auditor and audit office, audit fees, audit opinion information, and restatement information. Audit-related fees are important because they include fees for the information technology security review.[18] I then merge financial, IPO, and delisting data from the Compustat

[15] For example, as discussed by SQN Banking Systems in "Manipulated Data: The New Bank Hack," "[A]ccess control" "helps to prevent and reduce *internal data manipulation*, it also reduces the number of avenues through which *hackers can gain entry to manipulate the data*." (see https://sqnbankingsystems.com/blog/manipulated-data-new-bank-hack/). For subsequent empirical analyses on the effect of auditing on the likelihood of data breaches, I split the sample into hacking and non-hacking cases. While I find a slightly larger reduction for non-hacking cases, its higher power (due to its larger sample size) may confound the interpretation that auditors play a bigger role in reducing non-hacking breaches.

[16] I collect data for all US firms that are listed on the NYSE, Amex, and Nasdaq. All continuous variables are trimmed at the 1% level.

[17] I also check to see whether an organization's parent company is publicly listed. I compare the matching rate with prior literature and find similar matching rates for online hacks. Specifically, Kamiya et al. (2018) find 307 online attacks and 224 unique public firms. I find 304 online hacks and 211 unique public firms. In total, I find 1,214 data breaches and 524 unique firms.

[18] In the contract between EY and Equifax, for example, "EY charged $4.3 million of that total for audit-related services including service auditor examinations, or SOC reports, provided to banks and other financial firm customers

Annual file. I use SEC 10-K filings to extract firms' business addresses and get the PCAOB inspection dataset from the PCAOB website.[19] I use BoardEx data to compute the percentage of audit committee members on a firm's board of directors.

4. Do Auditors Help Prevent Data Breaches?

4.1 Descriptive Evidence

I first estimate a regression model to examine potential firm characteristics that are related to data breaches. The goal of this analysis is two-fold. First, it provides descriptive evidence on whether auditing-related variables are associated with the likelihood of data breaches. Second, it helps determine what control variables should be included in the subsequent analyses. I estimate the following model at a firm-year level (suppressing time and firm subscripts):

$$Breach = \sum \beta_i Determinants_i + \sum \alpha Fixed\ Effects + \epsilon \quad (1)$$

Breach is an indicator variable equal to one if a firm experiences data breaches in a given year, and zero otherwise. I include auditing-related variables to assess whether the likelihood of data breaches is associated with firms' auditing-related resources. Auditing-related variables include the percentage of audit committee members (measured as the number of audit committee members divided by the total number of board members), internal control weakness, and big auditors. I also include firm performance (measured as a loss indicator), size (measured as the natural log of total assets), and asset intangibility (measured as the proportion of assets that are not PP&E) following prior literature (e.g., Kamiya et al. 2018).

Big firms may have the resources to establish strong internal control systems, thereby reducing the likelihood of data breaches. However, big firms are also more visible and are therefore

to prove First Data's controls over data security, availability, processing integrity, confidentiality and privacy meet legal and regulatory requirements."

[19] I thank Brandon Gipper for sharing the detailed PCAOB inspection dataset with me (Gipper et al. 2019).

more likely to be targeted by hackers. Their organizational structures are also so complex that insider disclosure and the physical loss of information are both more likely to occur, making it exponentially more costly to defend against data breaches. Thus, the effect of firm size on data breaches is unclear. Poor-performance firms may lack the necessary resources to strengthen internal controls, making them vulnerable to data breaches. The predictions for firm asset intangibility are also unclear. Because intangible assets are important, firms are likely to be aware of best practices for data protection and have many effective defense mechanisms in place. Additionally, external auditors may focus specifically on intangible assets. On the other hand, firms with high asset intangibility will likely have more data (e.g., more customer and employee information) and are thus more likely to have data breaches (e.g., Kamiya et al. 2018). Data breaches are correlated with industry characteristics (as shown in Appendix Figure A1), so I include industry fixed effects to control for static industry differences in Column (1) of Table 2. In addition to industry fixed effects, in Column (2) of Table 2, I include year fixed effects to account for general trends in data protection technologies. I cluster standard errors by (two-digit) industry to account for cross-sectional correlation within industries. Note that while observable characteristics may be related to other important determinants (e.g., size is likely correlated with the number of employees and asset intangibility may be associated with customer lists and information), this determinant model is coarse and other important yet unobservable determinants could play a role.

I report descriptive statistics in Table 1. Breaches are not frequent events: around 2% of firm-year observations have data breaches. Specifically, around 77% of unique sample firms have no data breaches, 8% have one over the sample period, and 15% have more than one data breach. On average, around 75% of observations have big auditors (defined as defined as the Big Four),

and around 24% of firms' net income is negative. In Table 2, I report the results of estimating Model (1), and find that the likelihood of data breaches is positively related to bad performance (e.g., a loss indicator), asset intangibility (i.e., the portion of assets coming from items other than PP&E), and visibility (e.g., size), and that data breaches are negatively associated with audit-related variables (e.g., the percentage of audit committee members, internal control strength, and big auditors) across two specifications (Column 1 has year fixed effects and Column 2 has year as well as industry fixed effects). These are all statistically significant determinants of data breaches at (at least) the 5% level (except the loss indicator in both columns, which is significant at the 10% level). The evidence in this table suggests that firm characteristics and auditing resources are related to data breaches.[20]

Table 3 presents descriptive results related to audit quality. To further explore the negative coefficient on big auditors, I examine firms' likelihood of data breaches in two cases: (1) when firms switch from big to non-big auditors and (2) when they switch from non-big to big auditors. Big auditors are a proxy for higher audit quality and stronger auditor incentives stemming from reputation and litigation concerns (e.g., DeFond and Zhang 2014). Thus, we should expect a positive and significant relation with the likelihood of data breaches when firms switch from big to non-big auditors (as shown in Column (2) of Table 3); conversely, we should see a negative and significant association with the likelihood of data breaches when firms switch from non-big to big auditors (as shown in Column (1) of Table 3). I include firm fixed effects to control for time-invariant firm differences in auditor choice and include year fixed effects to account for general trends in firms' auditor choice. Based on the results in my determinant table, I use the following control variables (in this and the subsequent tables), non-auditing related but associated with the

[20] I lose observations here because I have a limited sample period and matching rate when I use the BoardEx data.

likelihood of data breaches, in order to gauge the importance of firm characteristics in explaining the variation in auditing services I exploit: firm size (measured as the natural log of total assets), firm performance (measured as a loss indicator), and asset intangibility (measured as the portion of assets coming from items other than PP&E). Because the variations I explore in the following tables are related to auditing services, including auditing-related variables would incur "bad control problems" to the extent that firms' auditing-related resources vary with the shocks I exploit (Angrist and Pischke 2009). For firm control variables, in Table 3, I find that size, bad performance, and asset intangibility are positively and statistically correlated with the likelihood of data breaches, which is consistent with the determinant table.

Although the analyses above provide consistent descriptive evidence on the association between auditing and the likelihood of data breaches, these correlations may be potentially misleading due to endogeneity concerns. For instance, if firms switch from a big to non-big auditor because of financial constraints, it is likely that this constraint also affects firms' data protection technologies. Firms that are concerned with data protection are willing to spend more on assurance and will switch to big auditors. Thus, in the next section, I exploit plausibly exogenous shocks in order to provide more compelling evidence.

4.2 The Effect of Auditing on Reducing the Likelihood of Data Breaches

To explore the effect of auditing on the likelihood of data breaches, I exploit two shocks to the supply of audit services that (I argue) do not affect the demand for audit services. These shocks are not perfect and are subject to limitations (which I discuss below), but they complement each other. My baseline regression, suppressing time and firm subscripts, is

$$Breach = \alpha_1 Shocks + \sum \alpha_i Fixed\ Effects + \gamma_i Controls_i + \epsilon \qquad (2)$$

Breach is an indicator variable equal to one if a firm experiences data breaches in a given year,

and zero otherwise. *Shocks*, the variable of interest, is an indicator coded as one for the different shocks I describe below.[21] I include year fixed effects to control for changes in data technology and policy over time and to account for a time trend for reported breaches. To mitigate the concern that firms worried about data breaches switch to "better" auditors, instead of using firm fixed effects, I include firm×auditor fixed effects (rather than firm fixed effects) to control for differences in data protection, audit services, and other time-invariant factors among firms and auditors in order to isolate the variation in the firm-auditor relationship over time (excluding the variation of changing auditors).[22] I explore, within the firm-auditor relationship, how auditing shocks affect the likelihood of data breaches for auditors' clients. Because these shocks are on the auditor side, I cluster standard errors by auditor (instead of industry) to account for cross-sectional correlation in firms with the same auditor. One concern with this level of clustering is that some clusters may be unbalanced due to differences in auditors' client portfolios (Conley et al. 2018). To empirically assess how this concern affects my results, I verify that my results are robust to clustering by firm, state, industry, or year.

The first shock is regulation-based: the PCAOB's first-time inspection fieldwork (e.g., DeFond and Lennox 2017; Aobdia and Shroff 2017; Gipper, Leuz, and Maffett 2019; Shroff and Hanlon 2019; Krishnan, Krishnan, and Song 2014; Lamoreaux 2016). A PCAOB inspection provides public oversight of auditing and strengthens auditor attestation of firms' internal control

[21] I choose to run a linear probability model (LPM) for two reasons: (1) estimating a stringent fixed effects model for non-linear regressions (e.g., logit or probit) can be problematic due to the incidental parameter problems, but it is less of a concern for LPM. (2) I am interested in marginal effects, which are robust for LPM (Angrist and Pischke 2009; Wooldridge 2010).

[22] Across different shocks, my results are robust to the following fixed effect structures: (1) year, firm, and auditor fixed effects; (2) industry×year fixed effects that control for time-varying industry conditions (e.g., industry trends and determinants of changes in IT controls); (3) state×year fixed effects that control for time-varying changes in economic conditions within a state (except for the shock of auditors learning from data breaches: The magnitude more than doubles (-0.018) but the statistical significance is smaller (t-stats: -1.18)); (4) state×industry×year fixed effects that account for time-varying changes in states and industries in order to isolate the variation of treated auditors (in the shock of auditors learning from data breaches, the magnitude almost triples (-0.022) though the statistical power is lower (t-stats: -1.15)).

systems (e.g., Gipper, Leuz, and Maffett 2019). Gipper, Leuz, and Maffett (2019) argue that the new regime leads to improvements in auditing because audit deficiencies are identified and larger penalties and stricter enforcement are implemented by the PCAOB. They also provide details of the remediation process following the PCAOB inspections and illustrate how the PCAOB regime leads to changes in the way audits are conducted (see Section 1 Part 4 in the Internet Appendix of Gipper, Leuz, and Maffett (2019)). An important aspect is assessing auditors' tests of their clients' IT general controls (PCAOB 2010 and 2013). DeFond and Lennox (2017) find that the PCAOB inspections improve the quality of internal control audits and that auditors conduct more rigorous tests and evaluations of clients' internal control weaknesses after these inspections.

The PCAOB first-time inspections, staggered across different auditors at different times, provide variation on the auditor supply side.[23] Using the variation in the audit quality of inspected auditors, I examine whether those auditors' clients are less likely to have data breaches. Specifically, I exploit the fact that the PCAOB completes the first-time inspection at different timing for different auditors, which allows me to use firms whose auditors have not yet been inspected by the PCAOB as the control group. The treatment group is comprised of the clients of auditors inspected for the first time. *PCAOB First-Time Inspection*, the variable of interest, is an indicator coded as one after PCAOB first-time inspection for firms audited by an inspected auditor, zero otherwise. Thus, I identify the effects based solely on differences in the timing of the inspections. The staggered introduction mitigates concerns about concurrent economic and regulatory changes.

Table 4 shows that firms audited by higher quality auditors (proxied by being inspected by

[23] This paper uses the first inspection (the sharpest new event) as the beginning of the treatment period. Continuing inspections should increase audit quality further (if the elasticity of improvement is not zero) and accumulate the learning experience (if there are serial correlations), strengthening my results.

the PCAOB) are 0.4 percentage points less likely to have data breaches (Column 1 of Table 4); this result is statistically significant at the 10% level. The magnitude of the coefficient translates into about a 20% reduction in the likelihood of data breaches (relative to the mean value). Results remain stable after including control variables (Column 2 of Table 4); firm×auditor fixed effects absorb much of the variation in adjusted R-squared so I examine and find that the within R-squared (excluding fixed effects) doubles (untabulated), which both suggest that (to the extent that the observable characteristics are representative of unobservables) an omitted variable bias is less of a concern (Altonji, Elder, and Taber 2005; Oster 2019). For the control variables, as in the determinant table, size, poor performance, and asset intangibility are positively and statistically correlated with the likelihood of data breaches.

One concern about the PCAOB tests is whether the timing is endogenous. I interviewed with several PCAOB regulators who institutionally verified that the PCAOB does not explicitly consider data breaches when selecting auditors. However, it might be possible that the selection criteria could perfectly predict the risk of data breaches. To mitigate this concern, in addition to including firm×auditor fixed effects (and robust to industry×year fixed effects), I include big firms (who were already selected before the sample analyses) in my control group to control for the risk criteria in the PCAOB selection. Additionally, I include firm control variables that are highly correlated with the likelihood of data breaches in order to gauge how the change in these controls affects the variable of interest; my results remain robust in this specification.

There are some important limitations for the test of *PCAOB First-Time Inspection*. Because my breaches data start in 2005, I cannot fully utilize the PCAOB-inspection regime change. The variation is on the intensive margin, i.e., differences in the timing of the inspections, mainly from small auditors. There are several advantages to analyzing small auditors. For example, it shows

that the effect holds for a broad spectrum of audit quality (not just for big auditors); because the clients of small auditors are less likely to be targeted by hackers, their data breaches likely stem from the vulnerability in firms' information systems as opposed to targeted hacking. Lastly, small auditors (inspected later) may be less sophisticated than big auditors, which means that they may learn and improve more as a result of PCAOB inspections; prior literature does find that the inspections improve audit quality for small auditors (e.g., DeFond and Zhang 2014). While small auditors (and their clients) are a sensible analysis group, they could be a low power test because of a smaller sample size. This is consistent with a lower statistical significance level in Table 4.[24] The economic effect is also small, because small firms have less stringent internal control testing by auditors (SOX 404b). Two other potential concerns are that 1) that the number of treated firms is not balanced over years and 2) the sample size of firms and auditors in later years is small. To assess how these concerns affect my results, I analyze firms with small (non-Big Four) auditors with matched size, performance, and asset intangibility in the appendix, and find the economic magnitude remains stable.

To mitigate limitations in the test of *PCAOB First-Time Inspection*, I exploit a second learning-experience based shock: auditors learning from their mistakes (in the spirit of Murfin 2012). If an auditor's client has a restatement (as one example of audit failure), the auditor could learn from this incident and apply the knowledge to future engagements, thereby affecting its auditing.[25] Auditors are incentivized to strengthen their audits in order to avoid similar audit

[24] A large number of the observations in this analysis comes from the inclusion of big auditors (who are already selected and treated) in the sample. The variation comes from small auditors and the results are similar to when I only keep small auditors with the matching (as shown in the appendix). Including these big auditors can serve two roles: (1) as discussed in the endogenous timing section, I use them to control for general trends in data protection technologies and for the risk criteria used in the PCAOB selection; (2) having them in the control group mitigates the concern in Goodman-Bacon (2018) about the wrong signs in averaging the heterogeneous treatment effects because my results are similar with and without big auditors.

[25] If the cause of audit failure is not permanent (i.e., not inherent in an auditor), I should expect this incentive. However, if auditors do have innate characteristics that prevent them from strengthening the audit process for the other clients,

failures in the future. These strengthened incentives are consistent with prior work demonstrating that a high-profile audit failure affects an audit firm's ability to attract and retain clients (Skinner and Srinivasan 2012; Weber et al. 2008). The improvement in auditing services is relevant for financial data protection because the IT audit and the test of IT general controls are an integral part of the top-down approach of an integrated audit (e.g., AS5; Haislip et al. 2016). I examine, within the same auditor, whether the improvement in audit quality, stemming from auditors' learning, leads to fewer data breaches in future engagements. Different auditors learn from restatements for the first-time in different years after 2003, which allows me to use firms whose auditors have not yet been treated (by "learning from restatements") as the control group.[26] The treatment group is *other* clients of an auditor who learns from incidents. That is, the treatment group does not include the firms that induce auditors' learning. *Auditors Learning From Restatements (Data Breaches)*, the variable of interest, is an indicator coded as one after auditors learn from restatements (data breaches) with their other clients, zero otherwise. The variation comes from two sources: (1) the different timing of "auditors' learning" and (2) the change in a treated auditor's *other* clients' behavior. This research design uses the staggered introduction of auditors' learning and a within-group design, which both greatly alleviate the concern of concurrent regulatory and economic shocks. Panel A of Table 5 shows that other clients of treated auditors are 0.7 percentage points less likely to have data breaches in Column 1. To gauge the economic magnitude, I translate this number into about a 40% reduction in the likelihood of data breaches (relative to the mean value). When adding controls, the within R-squared doubles (untabulated) but the coefficient remains

this concern would bias against my findings. An alternative interpretation could be that auditors compensate for reputation loss.

[26] I choose 2003 as the starting period for *auditors learning from restatements* because the nature of restatements changes after SOX (e.g., Hranaiova and Byers 2007; Burks 2011) and auditors make a significantly higher number of audit adjustments before SOX (consistent with the interview evidence).

stable in Column 2, suggesting an omitted variable is less of a concern (Oster 2019). For the control variables, as in the determinant table, size, poor performance, and asset intangibility are positively and statistically correlated with the likelihood of data breaches. I also find that the effects from auditors' learning takes (at least) two years to materialize (Figure 2).

Restatement incidents are indirect as they are accounting-related issues; however, given the tension of being indirect, they can show that financial statement auditing has a spillover effect on preventing data breaches. Because IT auditors work within the context of financial statement auditing, both financial statement and IT auditors conduct more work and examine their clients more closely after a restatement. The indirectness of these incidents also makes client learning (e.g., other clients changing their data protection procedures after learning from another client's restatement) less likely.

To substantiate a direct learning experience (and similarly to the specification in auditors' learning from restatements), I examine whether auditors' learning from data breaches could lead to fewer data breaches in future engagements with other clients.[27] In this specification, the control group is firms whose auditors have not yet been treated by "learning from data breaches." Because this incident directly relates to the outcome variable, client learning could be a concern. That is, an auditors' other clients in the same state may be treated from the information spillover or from network effects.[28] To mitigate these concerns, the treatment group is defined differently than the group of auditors' learning from restatements. Thus, I examine the change in the likelihood of data breaches for auditors' *other clients in other geographic areas* (i.e., other states), excluding other

[27] Prior research shows that audit fees increase after data breaches, suggesting auditors are aware of these breaches (Li et al. 2017 and Haislip et al. 2019). Lawrence et al. (2018) and Smith et al. (2019) show that data breaches are a cue for financial reporting deficiencies, and that they demonstrate a weakness in internal controls.
[28] However, auditors could have a greater information advantage (DeFond and Zhang 2014) than peers in the same geographic area because of firms' vague disclosure (Kopp et al. 2017; Kashyap and Wetherilt, 2019).

clients in the same state as the breached firms. Panel B of Table 5, Columns (1) and (2) (other clients [OC] in other geographic areas [OGA] with the same auditor) show that other clients of auditors with a data breach are 0.7 percentage points less likely to have data breaches. For ease of interpretation, I translate this number into around a 40% reduction in the likelihood of data breaches (relative to the mean value) in order to interpret the economic magnitude. Because other clients in a same state could be treated (by the same auditor and/or by the breached company due to network effects), I examine whether the effect is larger if we include other clients (of the same auditor) in the same state. In Column (3), I test and find that the effect is larger (the magnitude more than doubles) after including other clients in the same geographic area.[29] Thus, the effect varies predictably. There are two ways to interpret the increased economic magnitude in Column (3): first, other clients in the same geographic area could learn from breached firms; second, auditors' learning could be diffused more effectively when learning is local. To further exclude the possibility that industry risks are correlated with the estimates, I define the treatment as firms in different states and industries (at SIC 1–digit, 2-digit, and 3-digit levels) and repeat the same analyses (untabulated), finding that the results are similar. Coefficients of the control variables are similar to other tables. Size, poor performance, and asset intangibility are positively and statistically correlated with the likelihood of data breaches, which is consistent with the result in determinant table.

Although auditors' direct and indirect learning experiences have similar average magnitudes overall, their dynamic effects are different. The effect materializes relatively faster for direct incidents (shown in Figure 2). In addition, when I explore variation within the same state and industry (i.e., when I include state×industry×year fixed effects), the magnitudes are larger for

[29] The difference between Column 2 and Column 3 is slightly short of conventional levels of significance (p-value 0.149).

the direct incident. Exploring this variation can further mitigate concerns about client learning because if two similar firms in the same state and industry (but who differ along their auditor treatment) can learn from breached firms, we should not see a difference in the reduced likelihood of data breaches between these two firms.

There are some important assumptions and limitations in the auditor learning test. First, it takes time for learning to materialize. Second, there is an underlying assumption that learning does not depreciate. Specifically, although the improvement in audit quality would translate to an improvement in the quality of the IT audit and the test of IT general controls based on auditing standards (e.g., AS5) and academic findings (e.g., Haislip et al. 2016), how much *auditors learning from restatements* can improve the IT audit and the test of IT general controls could still be a concern. While I define the treatment as auditors' other clients in other states (and industries) to tighten the identification strategy, potential confounding shocks might still affect my results for *auditors learning from data breaches*.

Lastly, to assess the validity of the parallel-trends assumption, in Figure 2, I provide graphical evidence that treated and non-treated firms have similar patterns in the likelihood of data breaches before the shocks but that treated firms are less likely to experience the breaches afterward. One exception is the shock of PCAOB inspection, in which the treatment effect starts one period before the inspection ("t-1"). One interpretation is a downward trend in the pre-period. This downward trend indicates that some companies may already have data risks and thus the incentives to strengthen their internal controls when facing the improved audit quality stemming from the PCAOB inspections.[30]

[30] Another interpretation is that auditors might make preemptive adjustments before the end of inspection fieldwork. However, Gipper et al. (2019) finds evidence of no anticipation when examining the outcome variable ERC, which is less anticipatory for investors. One potential solution to a pre-trend is to use a covariate that is affected by the confound but not by the PCAOB inspection to correct for it (Freyaldenhoven et al. 2019). Additionally, I acknowledge that even

In summary, I examine, within firm-auditor relationships, whether an audit failure updates auditors' beliefs and strengthens their audit, thus reducing data breaches for other clients in distinct geographic areas. If firms do self-select into "better" auditors, firms in distinct geographic areas should also be aware of data breaches and change auditors accordingly. Nevertheless, this concern is mitigated in two ways. First, I explore variation within firm-auditor relationships, so my specification excludes the variation of changing auditors. Second, if an auditor achieves maximum audit ability and "optimal" auditors are selected by each firm, we should not see a reduction in the likelihood of data breaches for other clients in different geographic areas.

I use two different shocks to provide robust and consistent results that alleviate the endogeneity concerns. These shocks mitigate the endogeneity problem to the extent that they are not endogenously driven by firm-specific conditions. One concern is that differential trends in firm characteristics during this period may still result in differential data breaches, even in the absence of auditor shocks. Although my identification does allow different firm characteristics within the same auditor, the key to my identification is whether or not the alternative explanations correlate on timing with my treatment variable. To the degree that pre-determined underlying differences do not vary with the shocks I exploit (i.e., the shock does not affect both the supply and the demand side of audit service), it cannot explain my results. Another related concern is whether an improvement in audit quality leads auditors to detect more data breaches (an actual breach). But it would bias against finding my treatment effect (i.e., a reduction in the likelihood of data breaches).

4.3 Mechanism Testing for the Effect of Auditing on Mitigating Data Breaches

Documenting precise mechanisms is challenging, and requires detailed within-firm data. To overcome these data challenges, I provide both institutional and archival evidence to test the

if pre-trends are not detected, it may be either there are no pre-trends or pre-trends are undetected due to limited statistical power.

mechanisms. Institutionally, I conduct one-on-one interviews with 36 industry professionals in order to obtain institutional insights and collect information on mechanisms beyond empirical analyses. I conduct 19 interviews by phone, 14 interviews in person, and three interviews over online messages; interviewees include 11 accounting firm partners, five (non-partners) external auditors, nine internal auditors, one audit committee member, five corporate legal counsels/experts, and five regulators. For in person and phone interviews, the average length was around 42 minutes. Two channels are summarized from the ample anecdotes provided by interviewees: information transfer and internal controls, both consistent with the mechanisms proposed in the paper. In the appendix, I provide a brief interview summary for the accounting firm partners. Although interviews provide evidence unobservable in data, they may suffer from some biases (e.g., the social desirability bias (Furnham 1986) or the anchoring bias (Sherif et al. 1958)), especially when the interactions are in person. To mitigate these concerns, I also conducted anonymous surveys. In order to maximize the unbiasedness and informativeness of the survey responses, I used a neutral tone, did not mention the exact goal of my paper, asked professional consultants to help design the survey, and pre-tested it with some academics and practitioners. In its current form, I collected 20 survey responses. Seventy-seven percent of auditors think that accounting information and IT systems are intertwined. Ninety percent of auditors think that IT audit and internal control tests can help protect firms' financial reporting data. Seventy-seven percent of auditors think that IT audit and internal control tests can also help protect firms' non-financial data (e.g., employee information, consumer information, and other non-financial data). Eighty-seven percent of auditors believe that financial and non-financial data are saved in the same repository. Eighty-five percent of auditors agree that IT general controls operate in combination with other data control systems. I also asked auditors some open questions about how their work

could help protect clients' data. Most of their answers mention internal control reviews. For example, they specifically discuss attack and penetration tests, access management controls, change management controls, and IT control. One survey participant also notes that it depends on whether the data protection is an element of enterprise risk assessment. In summation, while firms' data protection process is unobservable, interview and survey evidence support my findings.

To provide archival evidence, I exploit cross-sectional differences in intertwined systems, audit committees, and internal control weaknesses. I set up a stylized theoretical model in the appendix in order to formalize the intuition of the two channels in the paper and to illustrate two cross-sectional predictions: (1) in more integrated data systems, information transfer from financial data to other systems is more likely; (2) when firms are more receptive to auditors' ex post monitoring, the ex ante incentive for strengthening internal controls is stronger.[31] The empirical proxy variable for integrated systems is firms' explicit disclosure about whether they have integrated systems (e.g., "Enterprise Resource Planning" (ERP), "Integrated Information Systems," and "Integrated Database"). Specifically, *More (Less) Pre Intertwined Syste*ms is an indicator coded as one if a firm mentions the terms "Enterprise Resource Planning" (ERP), "Integrated Information Systems," and "Integrated Database" in financial statements (including the 10-K, 10-Q, and 8-K) and in press releases from the beginning of their public disclosure on SEC Edgar until 2008, and zero otherwise.[32] ERP systems have a common database to support all

[31] Because intangible assets (e.g., customer lists) have become more prevalent in the past few decades (e.g., Haskel and Westlake 2018) and because auditors verify the value of these data, it may not be surprising that auditing plays a role in securing intangible assets (e.g., securing customer data). To confirm this conjecture and to examine whether auditing can help prevent data breaches in a broader setting, I implement another cross-sectional split based on the intangible asset of customer lists. I define intangible assets as the portion of assets that comes from items other than PP&E (as in the determinant table). To build the customer list split, I search through firms' financial statements to see if they disclose either a "customer list" or a "consumer list." Next, I use this disclosed term as a proxy variable for firms' customer lists (although this proxy could be coarse and noisy), and conduct cross-sectional analyses. I examine and find that the effect also exists for firms with fewer intangible assets and customer lists, although the magnitude is smaller than for the results with more intangible assets and customer lists (untabulated).

[32] Companies disclose these terms under "General and Administrative Expenses," "Operating Expenses," or "Item 9A. Controls and Procedures." To increase the power of test, I include disclosed observations until 2008. One potential

applications and have an integrated system to automate business processes (e.g., Morris 2011; Lecic and Kupusinac 2013). However, because ERP systems include features and "built-in" controls to help firms comply with the internal control over financial reporting (Morris 2011), many firms that do not disclose those terms in their financial statements presumably also implement ERP systems. For this reason, while this cross-sectional analysis provides informative statistics, the difference between disclosed and non-disclosed-ERP firms could be small. The proxy variable for firms' ex ante incentives for adopting high-quality internal controls is firms' audit committees and their internal control strength. Specifically, *With (Without) Pre Audit Committees* is an indicator coded as one if the firm has a percentage of audit committee members above (below) the median value for all firms between 2001 and 2004, and zero otherwise. *No (With) Pre Internal Control Weakness* is an indicator coded as one if the firm has (does not have) internal control weaknesses between 2004 and 2005, and zero otherwise. Firms with audit committees are likely more cooperative with external auditors and more receptive to auditors' ex post monitoring; measuring firms' strong internal control environment is a way to validate firms' ex ante incentives for adopting high-quality internal controls.

In Panels A — C of Table 6, I find that the reductions in the likelihood of data breaches are more pronounced when firms have more integrated systems, audit committees, or strong internal controls across the shocks I exploit, which is consistent with the cross-sectional predictions in the three specifications with different shocks. In the more versus less integrated system partition, the difference is statistically significant in *Auditors Learning From Restatements* (p-value 0.078) but not in *PCAOB First-time Inspection* and *Auditors Learning From Data*

concern about integrated systems is that hackers may have access to more data if systems are integrated and centralized. However, this concern would bias against my result. In addition, this concern applies *after* the system is breached instead of impacting the *prevention* of breaches.

Breaches (p-value 0.536 and 0.426 respectively), consistent with the above discussion that firms that do not disclose ERP-related terms in their financial statements may still implement ERP systems because of the benefits of complying with the financial reporting regulation (Morris 2011). In the with versus without pre-audit committees partition, the difference is statistically significant in *PCAOB First-time Inspection* and *Auditors Learning From Restatements* (p-value 0.073 and 0.006 respectively) but falls short of conventional levels of significance in *Auditors Learning From Data Breaches* (p-value 0.173). In the with versus without pre-internal control weakness, the difference is statistically significant in *Auditors Learning From Restatements* and *Auditors Learning From Data Breaches* (p-value 0.001 and 0.000 respectively) but also falls short of conventional levels of significance in *PCAOB First-time Inspection* (p-value 0.213). The coefficients of the control variables are similar to the other tables. Because the proxy variables I use for cross-sectional partitioning are noisy and subject to alternative interpretations, these results should be interpreted with caution. I acknowledge that these tests are imperfect and subject to endogeneity concerns, but they provide descriptive evidence and the sample size is also large.

5. Conclusion

With manual processes and documentation moving into the digital world, auditing these risks is critical. This paper documents a lower likelihood of data breaches for firms disciplined by improved auditing and supports a disciplining effect of the auditing process on reducing a new kind of agency friction between firms and data providers. Additionally, my paper also provides implications for (1) how data technology shifts the landscape of the accounting labor market because of the need to assess technology performance and related risks and because the auditor's skill set and training has evolved, and (2) the knowledge spillover (Simunic 1984) from the potential synergy between cybersecurity risk consulting/IT specialty and auditing services within

an accounting firm.

However, there are also several limitations. First, the main threat to identification is a potential violation of the parallel-trends assumption. Although I use different settings in an attempt to rule out concerns about omitted variables and contemporaneous changes, it is still possible that there are other confounding factors. Second, the within-firm internal control procedures for data protection are not observed. These internal controls are consistent with the SEC's Section 21(a) investigative report, emphasizing the importance of internal accounting controls to prevent cyber fraud (SEC 2018). While I test mechanisms institutionally and empirically, their unobservable nature is an important caveat to my findings. Third, though auditing's impact in reducing the likelihood of data breaches is a benefit, it is not enough to justify an increase in mandatory investment in auditing. My study does not (nor does it intend to) examine all the potential costs and benefits of auditing, nor is auditing necessarily the best way to prevent data breaches. An interesting avenue for future research is examining the effectiveness of different monitoring policies on preventing data breaches – a "pecking order" theory for addressing the agency frictions between firms and data providers.

References

Altamuro, J., and Beatty, A. (2010). How does internal control regulation affect financial reporting?. Journal of Accounting and Economics, 49(1-2), 58-74.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of Political Economy, 113(1), 151-184.

Alves, M. D. C. G. (2010). Information technology roles in accounting tasks-A multiple-case study. International Journal of Trade, Economics and Finance, 1(1), 103.

American Institute of Certified Public Accountants (AICPA). (1984). The effects of computer processing on the audit of financial statements. Statement on Auditing Standards No. 48.

American Institute of Certified Public Accountants (AICPA). (2006). Planning and supervision. Statement of Auditing Standards No. 108.

American Institute of Certified Public Accountants (AICPA). (2006). Understanding the entity and its environment and assessing the risks of material misstatement. Statement of Auditing Standards No. 109.

American Institute of Certified Public Accountants (AICPA). (2006). Performing audit procedures in response to assessed risks and evaluating the audit evidence obtained. Statement of Auditing Standards No. 110.

Angrist, J. D., and Pischke, J. S.(2009). Mostly harmless econometrics: An empiricist's companion, 1st ed. Princeton, NJ: Princeton University Press.

Aobdia, D. (2018). The impact of the PCAOB individual engagement inspection process—Preliminary evidence. The Accounting Review, 93(4), 53-80.

Aobdia, D., and Shroff, N. (2017). Regulatory oversight and auditor market share. Journal of Accounting and Economics, 63(2-3), 262-287.

Barrios, J., Lisowsky, P., and Minnis, M. (2018). Measurement matters: Financial reporting and productivity. Working paper, University of Chicago.

Ball, R. (1980). Discussion of accounting for research and development costs: The impact on research and development expenditures. Journal of Accounting Research, 27-37.

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). CEM: Coarsened exact matching in Stata. The Stata Journal, 9(4), 524-546.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? Evidence from India. Quarterly Journal of Economics, 128(1), 1-51.

Brands, K., and Smith, P. (2016). Ready or not, here comes accounting automation. Strategic Finance, 97(9), 70.

Burks, J. J. (2011). Are investors confused by restatements after Sarbanes-Oxley? The Accounting Review, 86(2), 507-539.

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

Center for Audit Quality. (2016). Understanding cybersecurity and the external audit.

Center for Audit Quality. (2017). The CPA's role in addressing cybersecurity risk.

Chen, S., Sun, S. Y., and Wu, D. (2010). Client importance, institutional improvements, and audit quality in China: An office and individual auditor level analysis. The Accounting Review, 85(1), 127-158.

Cheng, M., Dhaliwal, D., and Zhang, Y. (2013). Does investment efficiency improve after the disclosure of material weaknesses in internal control over financial reporting?. Journal of Accounting and Economics, 56(1), 1-18.

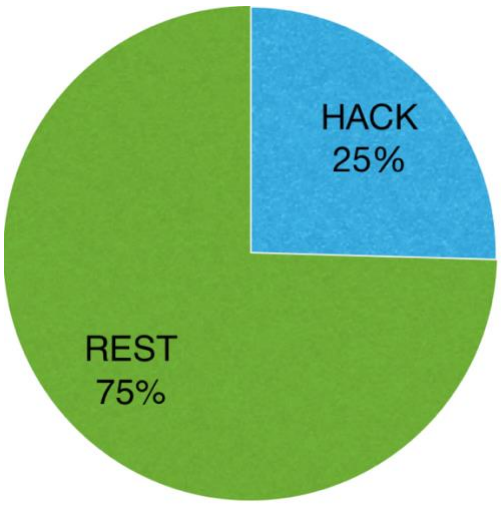
Christensen, H. B., Floyd, E., Liu, L. Y., and Maffett, M. (2017). The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records. Journal of Accounting and Economics, 64(2-3), 284-304.

Conley, T., Gonçalves, S., and Hansen, C. (2018). Inference with dependent data in accounting and finance applications. Journal of Accounting Research, 56(4), 1139-1203.
DeFond, M. L., and Lennox, C. S. (2017). Do PCAOB inspections improve the quality of internal control audits?. Journal of Accounting Research, 55(3), 591-627.
DeFond, M., and Zhang, J. (2014). A review of archival auditing research. Journal of Accounting and Economics, 58(2-3), 275-326.
De Ridder, M. (2019). Market power and innovation in the intangible economy. Working Paper, University of Cambridge.
Duguay, R. (2019). The economic consequences of financial audit regulation in the charitable sector. Working Paper, Yale University.
Duffie, D., and Younger, J. (2019). Cyber runs. Working Paper, Stanford University.
Efendi, J., Mulig, E. V., and Smith, L. M. (2006). Information technology and systems research published in major accounting academic and professional journals. Journal of Emerging Technologies in Accounting, 3(1), 117-128.
Feng, M., Li, C., and McVay, S. (2009). Internal control and management guidance. Journal of Accounting and Economics, 48(2-3), 190-209.
Feng, M., Li, C., McVay, S. E., and Skaife, H. (2015). Does ineffective internal control over financial reporting affect a firm's operations? Evidence from firms' inventory management. The Accounting Review, 90(2), 529-557.
Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. American Economic Review, 109(9), 3307-38.
Furnham, A. (1986). Response bias, social desirability and dissimulation. Personality and Individual Differences, 7(3), 385-400.
Gipper, B., Leuz, C., and Maffett, M. (2019) Public audit oversight and reporting credibility: Evidence from the PCAOB inspection regime. Review of Financial Studies, forthcoming.
Goldfarb, A., and Tucker, C. (2012). Privacy and innovation. Innovation Policy and the Economy, 12(1), 65-90.
Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Working Paper, National Bureau of Economic Research.
Gordon, L. A., and Loeb, M. P. (2002). The economics of information security investment. ACM Transactions on Information and System Security (TISSEC), 5(4), 438-457.
Gwebu, K. L., Wang, J., and Wang, L. (2018). The role of corporate reputation and crisis response strategies in data breach management. Journal of Management Information Systems, 35(2), 683-714.
Haislip, J., Kolev, K., Pinsker, R., and Steffen, T. (2019). The economic cost of cybersecurity breaches: A broad-based analysis. Working Paper, Yale University.
Haislip, J. Z., Peters, G. F., and Richardson, V. J. (2016). The effect of auditor IT expertise on internal controls. International Journal of Accounting Information Systems, 20, 1-15.
Hanlon, M., and Shroff, N. (2019). Insights into auditor public oversight boards: Whether, how, and why they "work." Working Paper, MIT.
Haskel, J., and Westlake, S. (2018). Capitalism without capital: The rise of the intangible economy. Princeton University Press.
Heckman, J. J. (1981). Heterogeneity and state dependence. In Studies in Labor Markets (pp. 91-140). University of Chicago Press.
Hoffman, B. W., Sellers, R. D., and Skomra, J. (2018). The impact of client information technology capability on audit pricing. International Journal of Accounting Information Systems, 29, 59-75.
Hogan, C. E., and 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.
Hranaiova, J., and Byers, S. L. (2007). Changes in market responses to financial statement restatement announcements in the Sarbanes-Oxley era. Working Paper, SSRN 1319354.
Jin, G. Z. (2018). Artificial intelligence and consumer privacy. Working Paper, National Bureau of Economic Research.

Kamiya, S., Kang, J. K., Kim, J., Milidonis, A., and Stulz, R. M. (2018). What is the impact of successful cyberattacks on target firms? Working Paper, National Bureau of Economic Research.
Kannan, K., Rees, J., and Sridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. International Journal of Electronic Commerce, 12(1), 69-91.
Kashyap, A. K., and Wetherilt, A. (2019). Some principles for regulating cyber risk. AEA Papers and Proceedings 109: 482-87.
Kopp, E., Kaffenberger, L., and Jenkinson, N. (2017). Cyber risk, market failures, and financial stability. Working Paper, International Monetary Fund.
Krishnan, J., Krishnan, J., and Song, H. (2016). PCAOB international inspections and audit quality. The Accounting Review, 92(5), 143-166.
Lamoreaux, P. T. (2016). Does PCAOB inspection access improve audit quality? An examination of foreign firms listed in the United States. Journal of Accounting and Economics, 61(2-3), 313-337.
Lawrence, A., Minutti-Meza, M., and Vyas, D. (2018). Is operational control risk informative of financial reporting deficiencies? Auditing: A Journal of Practice and Theory 37 (1): 139–165.
Lecic, D., and Kupusinac, A. (2013). The impact of ERP systems on business decision-making. TEM Journal, 2(4), 323.
Leuz, C., and Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. Journal of Accounting Research, 54(2), 525-622.
Li, H., W. G. No, and J. E. Boritz. (2017). Are external auditors concerned about cyber incidents? Evidence from audit fees. Working paper, University of Waterloo.
Li, C., Peters, G. F., Richardson, V. J., and Weidenmier Watson, M. (2012). The consequences of information technology control weaknesses on management information systems: The case of Sarbanes-Oxley internal control reports. MIS Quarterly 36(1): 179-203.
Liu, L. Y. (2019). How do data breaches affect consumer spending? Working Paper, University of Chicago.
Mansi, S. A., Maxwell, W. F., and Miller, D. P. (2004). Does auditor quality and tenure matter to investors? Evidence from the bond market. Journal of Accounting Research, 42(4), 755-793.
Minnis, M. (2011). The value of financial statement verification in debt financing: Evidence from private US firms. Journal of Accounting Research, 49(2), 457-506.
Morris, J. J. (2011). The impact of enterprise resource planning (ERP) systems on the effectiveness of internal controls over financial reporting. Journal of Information Systems, 25(1), 129-157.
Murfin, J. (2012). The supply-side determinants of loan contract strictness. Journal of Finance, 67(5), 1565-1601.
Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187-204.
Public Company Accounting Oversight Board (PCAOB). (2003). The personnel management element of a firm's system of quality control-competencies required by a practitioner-in-charge of an attest engagement. Quality Control Standards Section No. 40.
Public Company Accounting Oversight Board (PCAOB). (2006). An audit of internal control over financial reporting that is integrated with an audit of financial statements and related other proposals. Auditing Standard No. 5.
Public Company Accounting Oversight Board (PCAOB). (2007). An audit of internal control over reporting that is integrated with audit of financial statements and related independence rule and conforming amendments. Auditing Standard No. 5.
Public Company Accounting Oversight Board (PCAOB). (2010). Identifying and assessing risks of material misstatement. Auditing Standard No. 12. Appendix B – Consideration of Manual and Automated Systems and Controls.
Public Company Accounting Oversight Board (PCAOB). (2010). The auditor's responses to the risks of material misstatement. Auditing Standard No. 13.
Public Company Accounting Oversight Board (PCAOB). (2013). Staff Audit Practice Alert No. 11: Considerations for audits of internal control over financial reporting.
Rice, S. C., and Weber, D. P. (2012). How effective is internal control reporting under SOX 404?

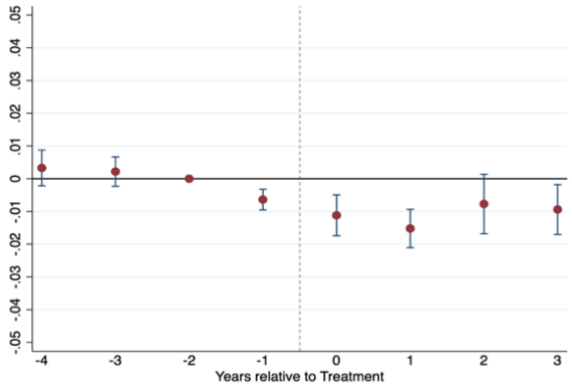
Determinants of the (non-) disclosure of existing material weaknesses. Journal of Accounting Research, 50(3), 811-843.
Romanosky, S., Telang, R., and Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? Journal of Policy Analysis and Management, 30(2), 256-286.
Schoenfeld, J. (2020). Auditing in the era of big data. Working Paper, Dartmouth College.
Schroeder, J. H., and Shepardson, M. L. (2015). Do SOX 404 control audits and management assessments improve overall internal control system quality? The Accounting Review, 91(5), 1513-1541.
Securities and Exchange Commission (SEC). (2018). Report of investigation pursuant to Section 21(a) of the Securities Exchange Act of 1934 regarding certain cyber-related frauds perpetrated against public companies and related internal accounting controls requirements.
Sheneman, A. G. (2018). Cybersecurity risk and the cost of debt. Working Paper, Ohio State University.
Sherif, M., Taub, D., and Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. Journal of Experimental Psychology, 55(2), 150.
Simunic, D. A. (1984). Auditing, consulting, and auditor independence. Journal of Accounting Research, 679-702.
Skinner, D. J., and Srinivasan, S. (2012). Audit quality and auditor reputation: Evidence from Japan. The Accounting Review, 87(5), 1737-1765.
Smith, T., J. L. Higgs, and R. Pinsker. (2019). Do auditors price breach risk in their audit fees? Journal of Information Systems, forthcoming.
Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. The Journal of Strategic Information Systems, 28(3), 257-274.
Weber, J., Willenborg, M., and Zhang, J. (2008). Does auditor reputation matter? The case of KPMG Germany and ComROAD AG. Journal of Accounting Research, 46(4), 941-972.
Westermann, K. D., Cohen, J., and Trompeter, G. (2019). PCAOB inspections: Public accounting firms on "trial." Contemporary Accounting Research, 36(2), 694-731.
Wooldridge, J. 2010. Econometric analysis of cross section and panel data, 2nd ed. Cambridge, MA: MIT Press.
Yang, D. C., and Guan, L. (2004). The evolution of IT auditing and internal control standards in financial statement audits: The case of the United States. Managerial Auditing Journal, 19(4), 544-555.
Yoon, K., Hoogduin, L., and Zhang, L. (2015). Big Data as complementary audit evidence. Accounting Horizons, 29(2), 431-438.

Figure 1: Public Companies' Data Breaches by Type

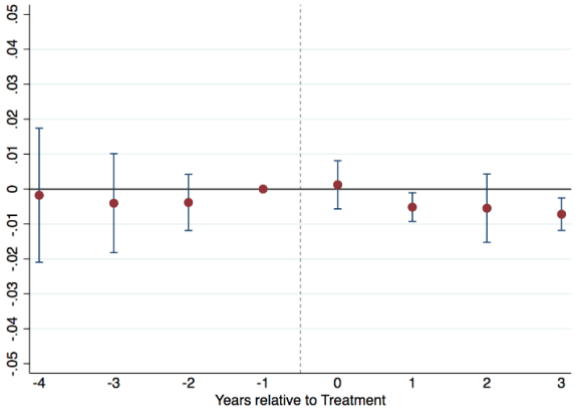


Notes: This figure presents total data breaches by type. After manually matching breaches with public companies, there are 1,214 observations from 2005 to 2017. *HACK* (25%): Hacked by an outside party or infected by malware; *Rest* (75%): Insiders' mishandled data, including lost laptops without encryption, sensitive information posted publicly, etc.

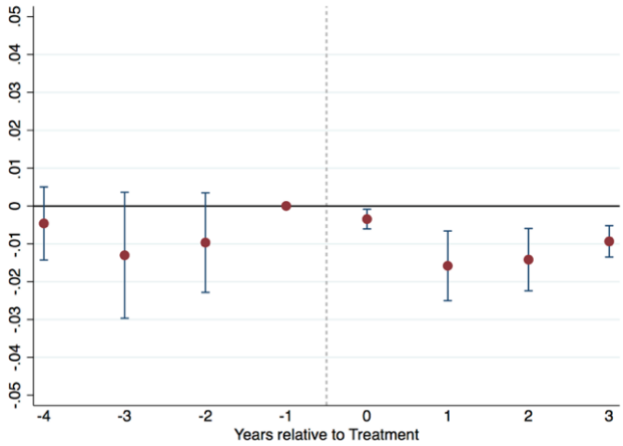
Figure 2: Trend of Counterfactual Treatment Effects on the Likelihood of Data Breaches
Panel A: PCAOB:



Panel B: Learning from Restatements:



Panel C: Learning from Data Breaches:



Notes: This figure displays OLS regression coefficient estimates and two-tailed 95% confidence intervals with a Bonferroni correction based on standard errors clustered at the auditor level for analyses of the change in the likelihood of data breaches (as in Tables 4 and 5). I map out estimated counterfactual treatment effects in event time. In one regression, I include indicators for all years relative to the treatment except the benchmark period (i.e., the coefficient is constrained to equal zero).

Table 1: Descriptive Statistics

Variable	*N*	*Mean*	*Std. Dev.*	*Median*	*P10*	*P90*
Breached Public Firms (Firm-Year Level)						
Size	1,211	9.767	2.313	9.803	6.715	12.671
Loss	1,211	0.142	0.349	0.000	0.000	1.000
Asset Intangibility	1,211	0.834	0.191	0.917	0.541	0.993
Total Public Firms (Firm-Year Level)						
Size	55,827	6.577	2.203	6.552	3.709	9.461
Loss	55,827	0.295	0.456	0.000	0.000	1.000
Asset Intangibility	55,827	0.783	0.246	0.887	0.355	0.992
PCAOB First-Time Inspection	55,827	0.911	0.285	1.000	1.000	1.000
Auditors Learning From Restatements	55,827	0.735	0.441	1.000	0.000	1.000
Auditors Learning From Data Breaches	55,827	0.168	0.374	0.000	0.000	1.000
Determinants of Data Breaches (Firm-Year Level)						
Breach	33,677	0.015	0.122	0.000	0.000	0.000
% of Audit Committee Members	33,677	0.306	0.100	0.300	0.188	0.429
Big Auditors (Indicator)	33,677	0.745	0.436	1.000	0.000	1.000
Loss	33,677	0.245	0.430	0.000	0.000	1.000
Asset Intangibility	33,677	0.794	0.233	0.888	0.393	0.990
Size	33,677	6.712	1.980	6.683	4.107	9.287
No Internal Control Weakness	33,677	0.744	0.437	1.000	0.000	1.000
Public Firms and Firms Submit X-17A-5 Filings (Firm-Year Level)						
Big Auditors (Indicator)	111,028	0.577	0.494	1.000	0.000	1.000
Size	111,028	4.263	4.076	5.296	-1.997	8.892
Log (Liability)	111,028	3.331	4.667	4.370	-3.866	8.492
Log (Revenue)	111,028	3.954	3.604	4.508	-1.161	8.266
Treatment*Post	111,028	0.052	0.222	0.000	0.000	0.000
BLS (State-Occupation-Year Level)						
Log (Total Employment)	491,508	6.788	1.820	6.709	4.382	9.222
Log (Mean Annual Wage)	491,508	11.899	1.031	11.889	10.528	13.267

Notes: This table presents summary statistics for the firm-year and state-occupation-year level data sets used in the analysis. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. *Breach* is an indicator coded as one if the firm has a data breach in a given year, and zero otherwise. *% of Audit Committee Members* is the number of audit committee members divided by total number of board members. *Big Auditors (Indicator)* is an indicator coded as one if a firm hires a big auditor, and zero otherwise. Big auditors are defined as the Big Four. *No Internal Control Weakness* is an indicator coded as one for a firm with no internal control weakness, and zero otherwise. *Log(Liability)* is the natural log of total liabilities. *Log(Revenue)* is the natural log of revenue. *Log(Employment)* is the natural log of employment at the state-occupation-year level. See Appendix A1 for further details on variable definitions.

Table 2: Determinants of Data Breaches

Dependent Variable: Breach	*(1)*	*(2)*
% of Audit Committee Members	-0.026***	-0.025**
	(-2.67)	(-2.60)
Loss	0.003*	0.003*
	(1.77)	(1.66)
Asset Intangibility	0.027***	0.026***
	(4.64)	(4.62)
Size	0.013***	0.013***
	(6.94)	(6.90)
No Internal Control Weakness	-0.009***	-0.010***
	(-4.76)	(-5.43)
Big Auditors (Indicator)	-0.008**	-0.006**
	(-2.46)	(-2.02)
Fixed Effects:		
Industry (2-Digit)	Yes	Yes
Year	No	Yes
Observations (Firm-Year)	33,677	33,677
Adjusted R-squared	0.042	0.044
Cluster	Industry	Industry

Notes: This table presents the determinants of data breaches. *Breach* is an indicator coded as one if a firm has a data breach in a given year, and zero otherwise. *% of Audit Committee Members* is the number of audit committee members divided by total number of board members. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Big Auditors (Indicator)* is an indicator coded as one if a firm hires a big auditor, and zero otherwise. Big auditors are defined as the Big Four. *No Internal Control Weakness* is an indicator coded as one for a firm with no internal control weakness, and zero otherwise. See Appendix A1 for further details on variable definitions. I include industry (two-digit) fixed effects in Column (1). I include industry (two-digit) and year fixed effects in Column (2). I cluster standard errors by industry (two digit) and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Table 3: Descriptive Evidence of Auditing on Data Breaches

Dependent Variable: Breach	*(1)*	*(2)*
Auditor Change (Non-Big → Big)	-0.007*	
	(-1.96)	
Auditor Change (Big → Non-Big)		0.004*
		(1.77)
Size	0.003***	0.003***
	(2.99)	(2.88)
Loss	0.003**	0.003**
	(2.59)	(2.56)
Asset Intangibility	0.013*	0.014*
	(1.74)	(1.80)
Fixed Effects		
Firm	Yes	Yes
Year	Yes	Yes
Firm Controls	Yes	Yes
Observations (Firm-Year)	56,589	56,589
Adjusted R-squared	0.124	0.124
Cluster	Industry	Industry

Notes: This table presents descriptive evidence on the relationship between auditing and data breaches. *Breach* is an indicator coded as one if a firm has a data breach in a given year, and zero otherwise. *Auditor Change (Non-Big → Big)* is an indicator coded as one after a firm changes from a non-big to a big auditor, and zero otherwise. *Auditor Change (Big → Non-Big)* is an indicator coded as one after a firm changes from a big auditor to a non-big auditor, and zero otherwise. *Breach* is an indicator coded as one if a firm has a data breach in a given year, and zero otherwise. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if a firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. See Appendix A1 for further details on variable definitions. I include firm and year fixed effects. I cluster standard errors by industry (two digit) and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Table 4: Effect of Auditing on Data Breaches with PCAOB

Dependent Variable: Breach	*(1)*	*(2)*
PCAOB First-Time Inspection	-0.004*	-0.004*
	(-1.93)	(-1.93)
Size		0.003***
		(6.53)
Loss		0.003**
		(2.49)
Asset Intangibility		0.013**
		(2.58)
Fixed Effects		
Firm×Auditor	Yes	Yes
Year	Yes	Yes
Firm Controls	No	Yes
Observations (Firm-Year)	55,827	55,827
Adjusted R-squared	0.105	0.105
Cluster	Auditor	Auditor

Notes: This table reports results on the effect of overseeing accounting information systems on data breaches, using the shock of PCAOB first-time inspection. *Breach* is an indicator coded as one if a firm has a data breach in a given year, and zero otherwise. *PCAOB First-Time Inspection*, the variable of interest, is an indicator coded as one after PCAOB first-time inspection for firms audited by an inspected auditor, zero otherwise. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. See Appendix A1 for further details on variable definitions. I include firm×auditor and year fixed effects. I cluster standard errors by auditor and report t-statistics in parentheses. Variations in identification strategy are illustrated in Appendix 1. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Table 5: Effect of Auditing on Data Breaches with Auditor Learning

Panel A: Auditor Learning from Restatements

Dependent Variable: Breach	*(1)*	*(2)*
Auditors Learning From Restatements	-0.007***	-0.007***
	(-6.87)	(-6.72)
Size		0.003***
		(6.49)
Loss		0.003**
		(2.50)
Asset Intangibility		0.013***
		(2.60)
Fixed Effects		
Firm×Auditor	Yes	Yes
Year	Yes	Yes
Firm Controls	No	Yes
Observations (Firm-Year)	55,827	55,827
Adjusted R-squared	0.116	0.116
Cluster	Auditor	Auditor

Panel B: Auditor Learning from Data Breaches

Dependent Variable: Breach	*OC in OGA* *(1)*	*OC in OGA* *(2)*	*OC* *(3)*
Auditors Learning From Data Breaches	-0.007***	-0.007***	-0.015*
	(-3.09)	(-3.16)	(-1.76)
Size		0.003***	0.003***
		(6.38)	(5.99)
Loss		0.003**	0.003**
		(2.48)	(2.48)
Asset Intangibility		0.013**	0.013**
		(2.56)	(2.51)
Fixed Effects			
Firm×Auditor	Yes	Yes	Yes
Year	Yes	Yes	Yes
Firm Controls	No	Yes	Yes
Observations (Firm-Year)	55,827	55,827	55,827
Adjusted R-squared	0.105	0.105	0.105
Cluster	Auditor	Auditor	Auditor

Notes: This table reports results on the effect of auditing on data breaches with auditor learning. Panel A presents results using the shock of auditor "learning from restatements" and Panel B presents results using the shock of auditor "learning from data breaches." *Breach* is an indicator coded as one if a firm has a data breach in a given year, and zero otherwise. *Auditors Learning From Restatements (Data Breaches)*, the variable of interest, is an indicator coded as one after auditors learn from restatements (data breaches) with their other clients, zero otherwise. *OC in OGA,* in Columns (1) — (2), defines treatment firms as other clients (*OC*) in other geographic areas (*OGA*) with the same auditor. *OC,* in Column (3), defines treatment firms as other clients (*OC*) with the same auditor. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. See Appendix A1 for further details on variable definitions. I include firm×auditor and year fixed effects. I cluster standard errors by auditor and report t-statistics in parentheses. Variations in the identification strategy are illustrated in Appendix 1. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Table 6: Cross-Sectional Analyses on the Effect of Auditing

Panel A: PCAOB First-Time Inspection

	Pre Intertwined Systems		*Pre Audit Committees*		*Pre Internal Control Weakness*	
Dependent Variable: Breach	*More*	*Less*	*With*	*Without*	*No*	*With*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
PCAOB First-Time Inspection	-0.005**	-0.003	-0.005**	-0.001	-0.004	-0.001
	(-2.30)	(-1.62)	(-2.03)	(-0.52)	(-1.48)	(-0.44)
Difference (p-value)	0.536		0.073		0.213	
Size	0.007**	0.003***	0.006***	0.002	0.004***	0.003***
	(2.59)	(4.77)	(3.76)	(1.48)	(9.63)	(2.98)
Loss	0.005	0.002	0.002*	0.004*	0.003**	0.003**
	(1.43)	(1.47)	(1.88)	(1.81)	(2.17)	(2.37)
Asset Intangibility	-0.005	0.016***	0.018***	0.010	0.023***	0.008
	(-0.25)	(4.57)	(3.39)	(1.01)	(4.02)	(0.83)
Fixed Effects						
Firm×Auditor	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations (Firm-Year)	9,813	46,014	30,743	25,084	25,777	30,050
Adjusted R-squared	0.085	0.111	0.143	0.033	0.144	0.053
Cluster	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor

Panel B: Auditors Learning From Restatements

	Pre Intertwined Systems		*Pre Audit Committees*		*Pre Internal Control Weakness*	
Dependent Variable: Breach	*More*	*Less*	*With*	*Without*	*No*	*With*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Auditors Learning From	-0.011***	-0.007***	-0.009***	-0.004***	-0.010***	-0.003***
Restatements	(-4.71)	(-6.08)	(-5.66)	(-5.34)	(-6.45)	(-2.94)
Difference (p-value)	0.078		0.006		0.001	
Size	0.007**	0.003***	0.006***	0.001	0.004***	0.003***
	(2.61)	(4.69)	(3.77)	(1.46)	(9.58)	(2.98)
Loss	0.005	0.002	0.002*	0.004*	0.003**	0.003**
	(1.44)	(1.47)	(1.89)	(1.81)	(2.15)	(2.38)
Asset Intangibility	-0.005	0.016***	0.018***	0.010	0.023***	0.008
	(-0.25)	(4.62)	(3.35)	(1.02)	(3.96)	(0.84)
Fixed Effects						
Firm×Auditor	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations (Firm-Year)	9,813	46,014	30,743	25,084	25,777	30,050
Adjusted R-squared	0.085	0.111	0.143	0.033	0.144	0.053
Cluster	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor

Table 6 Continued

Panel C: Auditors Learning From Data Breaches

	Pre Intertwined Systems		*Pre Audit Committees*		*Pre Internal Control Weakness*	
Dependent Variable: Breach	*More*	*Less*	*With*	*Without*	*No*	*With*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Auditor Learning From Data	-0.009**	-0.007***	-0.007***	-0.002	-0.010***	0.001
Breaches	(-2.42)	(-2.84)	(-2.82)	(-0.70)	(-4.35)	(0.54)
Difference (p-value)	0.426		0.173		0.000	
Size	0.007***	0.003***	0.006***	0.002	0.004***	0.003***
	(2.64)	(4.73)	(3.82)	(1.48)	(9.18)	(2.99)
Loss	0.005	0.002	0.002*	0.004*	0.003**	0.003**
	(1.43)	(1.45)	(1.85)	(1.81)	(2.13)	(2.38)
Asset Intangibility	-0.005	0.016***	0.018***	0.010	0.023***	0.008
	(-0.24)	(4.58)	(3.42)	(1.01)	(4.06)	(0.83)
Fixed Effects						
Firm×Auditor	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations (Firm-Year)	9,813	46,014	30,743	25,084	25,777	30,050
Adjusted R-squared	0.085	0.111	0.143	0.033	0.144	0.053
Cluster	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor

Notes: This table reports cross-sectional results on the effect of auditing on data breaches. Panel A reports the results using the PCAOB shock. Panel B reports results using "auditors learning from restatements." Panel C reports results using "auditors learning from data breaches." *More (Less) Pre Intertwined Syste*ms is an indicator coded as one if the firm mentioned the terms "Enterprise Resource Planning" (ERP), "Integrated Information Systems," and "Integrated Database" in their financial statements (including the 10-K, 10-Q, and 8-K) and in press releases until 2008, and zero otherwise. *With (Without) Pre Audit Committees* is an indicator coded as one if the firm has a percentage of audit committee members above (below) the median value for all firms between 2001 and 2004, and zero otherwise. *No (With) Pre Internal Control Weakness* is an indicator coded as one if the firm has (does not have) internal control weaknesses between 2004 and 2005, and zero otherwise. The dependent variable, *Breach*, is an indicator coded as one if the firm has a data breach in a given year, and zero otherwise. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. See Appendix A1 for further details on variable definitions. I include firm×auditor and year fixed effects. I cluster standard errors by industry (two-digit) and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

Appendix: Supporting Documents

1. Variable Definitions
2. Brief Summary of Interviews and Surveys
3. Validation of Underlying Empirical Assumptions
4. Public Companies' Data Breaches by SIC Industry (2-digit)
5. Number of Data Breaches by Year
6. Repeat Analyses Incorporating Audit Analytics Data
7. Results with Coarsened Exact Matching and Small Auditors
8. Effective Date of State Security Breach Notification Laws
9. Examples of Firms' Disclosure and Practitioners' Discussions
10. Simple Stylized Model

A1: Variable Definitions

Breach	An indicator coded as one if the firm has a data breach in a given year, and zero otherwise.
% of Audit Committee Members	The number of audit committee members divided by the total number of board members.
Asset Intangibility	One minus the proportion of PPE in total assets.
Loss	An indicator coded as one if the firm has a negative income, and zero otherwise.
No Internal Control Weakness	An indicator coded as one for a firm with no internal control weakness, and zero otherwise.
Big Auditors (Indicator)	An indicator coded as one if a firm hires a big auditor, and zero otherwise. Big auditors are defined as the Big Four.
Auditor Change (Non-Big → Big)	An indicator coded as one after the firm changes from non-big to big auditors, and zero otherwise.
Auditor Change (Big → Non-Big)	An indicator coded as one after the firm changes from big to non-big auditors, and zero otherwise.
PCAOB First-Time Inspection	An indicator coded as one after PCAOB first-time inspection for firms audited by an inspected auditor, zero otherwise.
Auditors Learning From Restatements (Data Breaches)	An indicator coded as one after auditors learn from restatements (data breaches) with their other clients, zero otherwise. The initial period for auditors' learning from restatements (data breaches) is 2003 (2005).
Size	The natural log of total assets.
More (Less) Pre Intertwined Systems	An indicator coded as one if the firm mentioned the terms "Enterprise Resource Planning" (ERP), "Integrated Information Systems," and "Integrated Database," in their financial statements (including the 10-K, 10-Q, and 8-K) and in press releases until 2008, and zero otherwise.
With (Without) Pre Audit Committees	An indicator coded as one if the firm has a percentage of audit committee members above (below) the median value for all firms between 2001 and 2004, and zero otherwise, and zero otherwise.
No (With) Pre Internal Control Weakness	An indicator coded as one if the firm has (does not have) internal control weaknesses between 2004 and 2005, and zero otherwise.
Treatment	An indicator coded as one if a company submits X-17A-5 filings, and zero otherwise.
Post	*Post* is an indicator coded as one if the year is after 2013, and zero otherwise.
Log(Employment)	The natural log of employment for an occupation in a given state and year.
DBState	An indicator coded as one if the state passes state security breach notification laws, and zero otherwise.
Auditor	An indicator coded as one if accountants and auditor occupations, and zero otherwise.

A2: Brief Summary of Interviews and Surveys

I conducted one-on-one interviews with 36 industry professionals to obtain institutional insights and to collect information on mechanisms beyond empirical analyses. I conducted 19 interviews by phone, 14 interviews in person, and three interviews via online messages; the interviewees included 11 accounting firm partners, five (non-partner) external auditors, nine internal auditors, one audit committee member, five corporate legal counsels/experts, and five regulators. In-person and phone interviews were around 42 minutes on average. I did not make any audio recordings for privacy reasons, but took notes in all interviews. This section summarizes the mechanisms suggested by the interviews and survey findings.

To keep from leading my interviewees, I usually started by asking about the relationship between external auditors and data breaches. Their responses helped provide rich mechanisms and anecdotes about how auditors could potentially help protect their clients' data. Another goal of these interviews was to validate the premise of variations used in the empirical analyses. For example, I asked the interviewees if the PCAOB inspections increased their audit quality and whether or not they cared about clients' data breaches and restatements. If they did care, I followed up by asking if they would apply relevant knowledge to other clients' future engagement.

All interviewees believe that it is not an external auditor's main job to detect data breaches as it is not prescribed in auditing standards. However, the interviewees care about data breaches and provide anecdotes on how their work could potentially affect their clients' data protection procedures. Two channels can be summarized from the ample anecdotes provided in the interviews: information transfer and internal control. For information transfer, auditors can inform firms of relevant and useful information related to data protection. For example, audit partners said that auditors could raise awareness, provide relevant information about systems, and could nudge managers about relevant knowledge and practice. Even though auditors only test some controls, the issues raised could spread due to correlations between different controls and data systems. External and internal auditors could also share their findings about firms' control environment with each other. For instance, external auditors use internal auditors' work and share information with them.

Another important channel is internal controls. For example, IT auditors' tests on IT general controls would include system and process controls, including logical access controls, change controls, change management controls, computer operations, data transfer, and system and process controls. One auditing partner stated that companies build not just financial data but also the whole environment, including controls over data access, data migration, and system change. When companies implement internal controls, it is hard to implement in one part (financial related controls) but not the other (non-financial related controls).

More than 90% of interviewees care about data breaches for several different reasons. One reason is that data breaches are related to failures in internal controls and could be an indication for bigger problems. For example, when data breaches happen, auditors need to understand the root cause in order to assess whether clients have adequate controls in place. They would also examine whether the data breach is isolated or whether it is an indication of systemic problem in a firm's control environment, something which could

incur financial consequences. They would check clients' IT control environment and discuss with the IT department more generally. For instance, one audit partner said that IT directors participated in audit committee meetings. Another reason is public perception, which is indirectly related to auditors' reputational risks. Although auditing standards do not specifically prescribe auditors' role in detecting and preventing data breaches, the public holds a (somewhat) strong belief in this role for auditors. This perception could indirectly affect auditors' behavior. The last reason is that auditors care about business risk and entity-level controls and data breaches may impact firms' business risks. Securing assets, virtual and physical, is important and data are a significant client asset.

Another important goal for these interviews was to validate the premise in my empirical strategies. When asked about the PCAOB, more than 90% of interviewees responded that the PCAOB could improve audit quality and procedures.[33] For example, one partner mentioned that a recent PCAOB inspection helped their future audit process because of the new information that they learned. This partner also believed that the PCAOB has relevant expertise but does not necessarily have firm-specific knowledge.[34] Another audit partner mentioned that external auditors have become more expensive and that clients have higher expectations of the audit process after the PCAOB inspection.

When asked about the application of knowledge learned from previous incidents, auditing partners talked about changing questions and procedures, and adjusting thought processes. When assessing risks for their other clients, auditors also raise skepticism and awareness, although they try to make sure engagement processes are consistent. Because external auditors reflect on why these incidents happen, they are able to take knowledge and experience to other firms. For example, they use their professional skillsets, look for commonalities, and ask deeper questions in order to discover patterns.

Although interviews provide evidence unobservable in data, they may suffer from some biases (e.g., the social desirability bias (Furnham 1986) or the anchoring bias (Sherif et al. 1958)), especially when the interactions are in person. To mitigate these concerns, I also conducted anonymous surveys. In order to maximize the unbiasedness and informativeness of the survey responses, I used a neutral tone, did not mention the exact goal of my paper, asked professional consultants to help design the survey, and pre-tested it with some academics and practitioners. I currently have 20 survey responses, which can be summarized as follows. Seventy-seven percent of auditors think accounting information and IT systems are intertwined. Ninety percent of auditors think IT audit and internal control tests can help protect firms' financial reporting data. Seventy-seven percent of auditors believe IT audit and internal control tests can also help protect firms' non-financial data (e.g., employee information, consumer information, and any other non-financial data). Eighty-seven percent of auditors indicate that financial and non-financial data are stored in the same data repository. Eighty-five percent of auditors reveal that IT general controls operate in combination with other data control systems. I also asked an open question about how auditors could help protect their clients' data; most of their answers mention internal control reviews. For example, they discussed attack and

[33] This finding is consistent with prior survey findings (Hanlon and Shroff 2019; Westermann et al. 2019).
[34] This is also consistent with the findings in Hanlon and Shroff (2019), who find that almost all their surveyed inspectors have a degree in accounting and rich work experience at a Big-N firm; no inspectors went to the public oversight boards directly after school.

penetration tests, access management controls, change management controls, and IT control. One survey participant also noted that it depends on whether the data protection is an element of enterprise risk assessment.

A3. Validation of Underlying Empirical Assumptions

An assumption maintained throughout this paper is that auditors have the relevant skills to test data protection controls. To further validate this assumption, I provide institutional discussions and empirical evidence below.

Institutionally, auditors learn relevant skills from their education and certification. Many professional accounting programs (e.g., graduate level) have courses on data analytics and AICPA has consistently issued auditing IT controls (see the appendix). Specifically, AICPA has issued the Statement on Standards for Attestation Engagements (SSAE) No. 16 and No. 18 (AT-C 105 and 205 Examinations), which provides detailed guidance on SOC 1 (internal control over financial reporting) and SOC 2 (non-financial data protection) audits. Both SOC 1 and SOC2 audits evaluate internal controls, policies, and procedures. The SOC 1 audit reports on user entities' internal control over financial reporting, while the SOC 2 audit examines firms' non-financial data control policies and procedures in order to help them to achieve five "trust services principles" (security, availability, processing integrity, confidentiality, and privacy). SOC 1 and SOC 2 audits are not mutually exclusive as the same vendor could process and store both non-financial (e.g., user information) and financial information (e.g., user-related transactions that affect firms' financial reporting). Big auditing firms are also equipped with resources such as personnel, training, and experience. In the payment card industry, auditors conduct a PCI Compliance Audit to ensure customers' data are protected. If a client's IT system is overly complex, external auditors would invite their IT specialists to assist with the auditing process.

Empirically, within the specific empirical setting of this paper, it is reassuring that data breaches are negatively related to firms' internal control strength in my determinant table. I also provide two empirical assessments. First, I examine how internal control weaknesses (ICW) change with the shocks I exploit. Conceptually, examining this outcome is sensible. However, the variation may be small and the empirical proxies may be imprecise. For example, prior research (e.g., Rice and Weber 2011) finds that only a minority of firms acknowledge their existing control weaknesses during misstatement periods. Rice and Weber (2011) also find that auditor effort is positively related to firms' acknowledgement of control weakness, so an increase in auditor effort may lead to more disclosures about these weakness. Thus, audit quality improvements could potentially predict either more or less ICW: when there is improved internal control, there would be less ICW; when the detection and report of ICW is more likely, reported ICW will increase. Empirically, I do not find significant changes from the shocks of *PCAOB first-time inspections* and *auditors learning from restatements*. However, I do find more ICW reports after *auditors' learning from data breaches*. That is, companies are more likely to report ICW after their auditors learn from data breaches, further reinforcing the idea that data breaches are a form of internal control failure. My second empirical assessment examines the change in restatements. I find that the magnitude is

two to three times larger in the shock of *PCAOB First-Time Inspection* than in reducing the likelihood of data breaches. A similar effect only occurs for other industries in the shock of *Auditors Learning From Restatements*. Results are almost the same (though with weak statistical power) in the shock of *Auditors Learning From Data Breaches*.

Next, I explore two regulatory shocks outside of my paper's setting to help me identify how audit services change with the rising cost of data breaches. The first regulatory shock is the "Regulation S-ID: Identity Theft Red Flags Rule" jointly issued by the SEC and the Commodity Futures Trading Commission (CFTC) in 2013. On April 19, 2013, the two agencies published their joint final rules and guidelines and included a compliance date of November 20, 2013. The Red Flags Rule requires financial institutions to implement a robust written program that can identify, detect, prevent, and mitigate identity theft. The rules help firms comply with the SEC's enforcement authority. Companies covered by this rule include most registered brokers, dealers, and investment companies, as well as some registered investment advisers.[35] Because firms' incentives to mitigate the identity theft of consumers are stronger after Regulation S-ID, a higher likelihood that firms hire high quality auditors in the post-Regulation S-ID period (relative to the control group) suggests that auditors play a role in detecting consumer information theft. Because big auditors have the incentive and capabilities to reduce potential business risks in order to maintain their reputation and reduce legal liabilities (e.g., DeFond and Zhang 2014), and because this variable is available in both the treatment and control group, I examine the change in the likelihood of hiring big auditors between the treatment and control group after Regulation S-ID. My baseline regression, suppressing time and firm subscripts, is

$$Big\ Auditors = \alpha_1 Treatment \times Post + \sum \alpha_i Fixed\ Effects + \gamma_i Controls + \epsilon \quad (1)$$

The dependent variable, *Big Auditors*, is an indicator variable equal to one if a firm hires a big auditor, and zero otherwise. Companies that submit X-17A-5 filings to the SEC are subject to Regulation S-ID and are the treatment group (*Treatment*). Companies that submit 10K (but not X-17A-5) filings are the control. The year after the 2013 compliance date is the post-period (*Post*). I include firm fixed effects to account for time-invariant firm differences in auditor hiring and include year fixed effects to flexibly account for changes over time in firms' auditor hiring that are common to both the treatment and control groups. In Column 1 of Table A1 Panel A, the treatment group is 2% more likely to hire big auditors than is the control group in the post Regulation S-ID period. From descriptive statistics in Table 1, we see that a typical treatment firm is much smaller than the typical control firm (e.g., the mean and median of the combined treatment and control is smaller than the mean and median of the control alone), suggesting that there are differences in the underlying firm characteristics of the treatment and control groups. To assess how these

[35] See https://www.sec.gov/rules/final/2013/34-69359.pdf

differences affect my estimates, I include control variables in Column 2. I include *Size, Log (Liability),* and *Log (Revenue)*, which are all reported by both the treatment and control firms (firms submitting X-17A-5 filings only report a limited set of financial numbers). These control variables play an economic role in firms' decision to hire big auditors (e.g., Mansi et al. 2004; Chen et al. 2011) and including them in our specification helps us gauge how they affect the variable of interest (Altonji, Elder, and Taber 2005; Oster 2019). After including the control variables in Column 2, the effect holds with a lightly larger magnitude (2.7%). For the control variables, larger firms and firms with higher revenue are more likely to hire big auditors. Liability is statistically insignificant in the regression; this could be due to collinearity among control variables.

The second regulatory shock is "State Security Breach Notification Laws," which has a staggered implementation across states.[36] The law requires companies to notify their consumers in a timely manner if their personal information was breached. Romanosky, Telang, and Acquisti (2011) argue that by increasing the costs of breaches, data breach disclosure laws could incentivize firms to strengthen their data protection. I explore the staggered implementation of this law and examine the subsequent change in auditor employment (relative to other occupations) in order to further corroborate the role of auditors in firms' data protection. My baseline regression, suppressing time, state, and occupation subscripts, is

$$Log(Employment) = \alpha_1 DBState * Auditor + \alpha_2 DBState + \sum \alpha_i Fixed\ Effects + Controls + \epsilon \quad (2)$$

The dependent variable, *Log(Employment)*, is the natural log of employment at the state-occupation-year level. *DBState* is an indicator variable equal to one for states that pass breach notification laws, and zero otherwise. *Auditor* is an indicator variable equal to one if the occupation is auditors, and zero otherwise. I use a different fixed effects structure than those in previous regressions because I also have different observation units (at state, year, and occupation levels). The fixed effect structure also varies depending on the specification. In Column (1), I include state fixed effects to account for static state differences in occupation and include year fixed effects to control for changes in occupation over time that are common across states. I also include occupation fixed effects to control for time-invariant occupation characteristics. In Column (2), in addition to year fixed effects, I include state×occupation fixed effects to control for average state-level differences in occupation. In Column (3), in addition to occupation fixed effects, I include state×year fixed effects to control for the time-varying economic changes in states that could

[36] See the appendix for the effective dates of the state security breach notification laws. Although consumer residency determines notification requirements, firms in states with data breach notification laws are aware of such incidents and are motivated to strengthen their data protection to avoid the potential costs of notification. This is consistent with the findings in Romanosky, Telang, and Acquisti (2011).

differentially affect my outcome variables across treatment and control states. In Column (4), I include state×year and state×occupation fixed effects in order to control for time-varying economic changes in states and for average state-level differences in occupation employment, respectively.[37] Across these four specifications (shown in Table A2 Panel B), I find consistent results that auditor employment increases (relative to other occupations) from 8% (Columns 3 and 4) to 10% (Columns 1 and 2) in states that passed data breach notification laws. Although the evidence is indirect, it suggests that auditors have a role in mitigating the risk of data breaches. Note that state security notification laws affect firms that internalize benefits of audit services and auditors that supply audit services, so the results should be interpreted as an estimate of how state laws influence the equilibrium outcomes of auditor employment.

In this analysis, the key identifying assumption is that the timing of the regulatory shock is not correlated with other factors that led to a change in auditor supply. One potential concern is that a string of high-profile data breaches led to the regulatory shock, thus affecting the audit market (Ball 1980). However, this interpretation reinforces auditors' role in mitigating the risk of data breaches, which is consistent with my findings. Additionally, after the passage of Regulation S-ID, data breach scandals would have a similar effect on the control group. Moreover, Romanosky, Telang, and Acquisti (2011) conduct empirical analyses and find no systematic evidence for endogenous timing when states pass breach notification laws.

In Table A1 Panels A and B, I exploit two regulatory shocks and find an increase in audit services when the cost of data breaches rises. One interpretation is that firms' awareness of auditing's potential role in preventing data breaches increases with the cost of data breaches. Firms are also more likely to internalize the benefit of external auditors when they are willing to increase data protection and when the deficiencies discovered in financial data systems are more likely to transfer to other data systems (financial data are the majority in Regulation S-ID). Due to data limitations, however, my results are subject to other interpretations. Because I do not have much granular information about firms' demand for audit services (e.g., audit fees or specific audit services) other than the classification of big auditors in the setting of Regulation S-ID, I cannot rule out that the choice to hire big auditors may serve other purposes. In the setting of the State Security Breach Notification Laws, I have auditor occupation data at the state-year level but do not have a detailed breakdown by type (i.e., internal auditors, government auditors, and other specialized auditors). These other auditors have responsibilities and skillsets directly related to data breaches (e.g., internal auditors are responsible for monitoring firms' data breaches). It is possible that my results capture this variation. If these alternative interpretations do not perfectly explain my results, however, these two analyses help further validate the assumption that auditors

[37] Including occupation×year fixed effects leaves insufficient variation to estimate reliable treatment effects because it excludes a large number of treatment and control observations.

have the relevant skillset.

Table A1: Changes in Audit Services When the Cost of Data Breaches Increases

Panel A: Regulation S-ID

Dependent Variable: Big Auditors	*(1)*	*(2)*
Treatment*Post	0.020***	0.027***
	(4.95)	(6.40)
Size		0.023***
		(9.56)
Log (Liability)		-0.004
		(-0.67)
Log (Revenue)		0.003**
		(2.01)
Fixed Effects		
Firm	Yes	Yes
Year	Yes	Yes
Firm Controls	No	Yes
Observations (Firm-Year)	111,028	111,028
Adjusted R-squared	0.856	0.857
Cluster	Firm	Firm

Panel B: State Security Breach Notification Laws

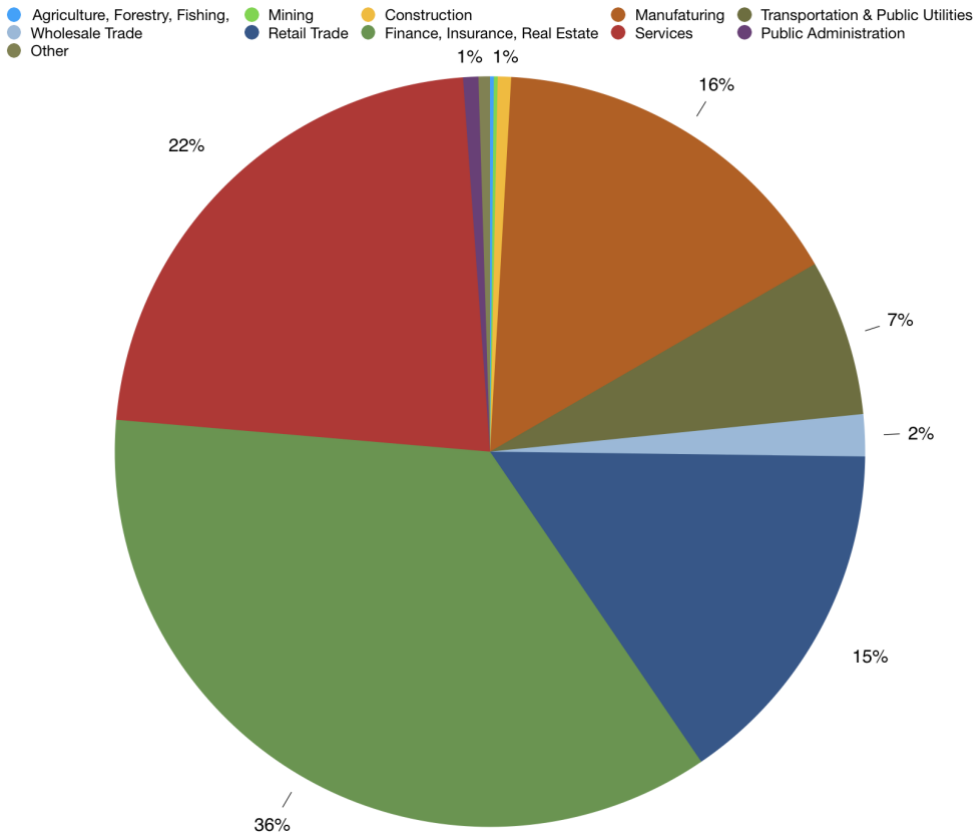
Dependent Variable: Log(Employment)	*(1)*	*(2)*	*(3)*	*(4)*
DBState*Auditor	0.105***	0.098***	0.084***	0.076***
	(3.51)	(5.00)	(2.86)	(4.41)
DBState	-0.006	-0.005		
	(-0.20)	(-0.15)		
Fixed Effects				
State	Yes	No	No	No
Year	Yes	Yes	No	No
Occupation	Yes	No	Yes	No
State×Year	No	No	Yes	Yes
State×Occupation	No	Yes	No	Yes
Observations (State-Occupation-Year)	491,508	489,879	491,508	491,508
Adjusted R-squared	0.835	0.931	0.839	0.935
Cluster	State	State	State	State

Notes: This table reports results on the auditor's role using two data-protection regulatory settings. Panel A reports results in the setting of "Regulation S-ID: Identity Theft Red Flags Rule." Panel B reports results in the setting of "State Security Breach Notification Laws." *Big Auditors* is an indicator coded as one if the firm hires a big auditor (Big Four), and zero otherwise. *Treatment* is an indicator coded as one if companies submit X-17A-5 filings, and zero otherwise. *Post* is an indicator coded as one if the year is after 2013, and zero otherwise. *Log(Employment)* is the natural log of employment at the state-occupation-year level. *DBState* is an indicator coded as one if the state passes the state security breach notification laws, and zero otherwise. *Auditor* is an indicator coded as one for accountant and auditor occupations, and zero otherwise. *Firm Controls* includes *Size, Log(Liability),* and *Log(Revenue)*, which are reported by both the treatment and control firms. *Size* is the natural log of total assets. *Log(Liability)* is the natural log of total liabilities. *Log(Revenue)* is the natural log of revenue. I cluster standard errors by firm in Panel A and by state in Panel B. I report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

A4: Public Companies' Data Breaches by SIC Industry (2-digit)

Figure A1 shows that 36% of matched public companies are from the finance, insurance, and real estate sectors, and that almost 40% are from the service, wholesale trade, and retail trade industries.

Figure A1: Public Companies' Data Breaches by SIC Industry (2-digit)

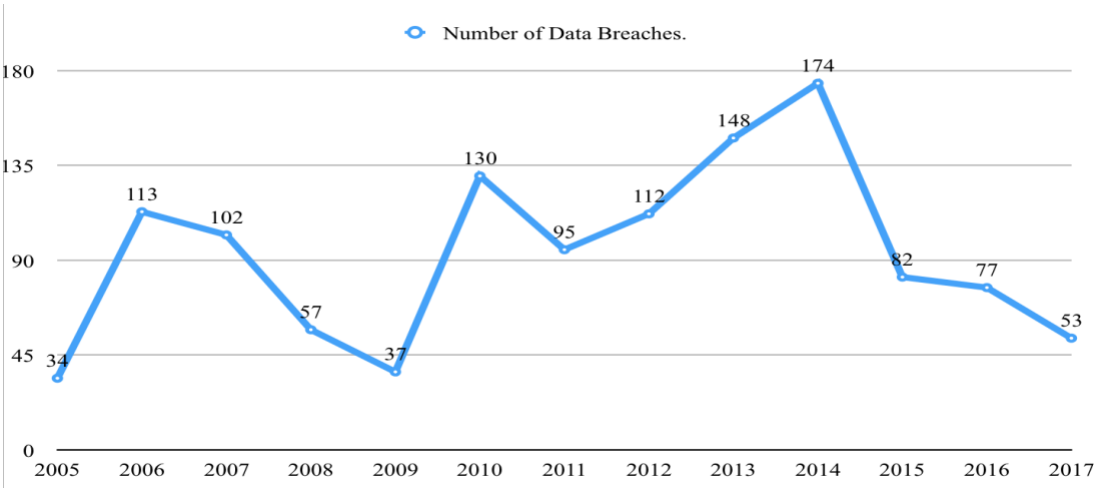


Notes: This figure presents total data breaches by type. After manually matching with public companies, there are 1,214 observations (524 unique firms). Specifically, the industries are Agriculture, Forestry, Fishing (0.16%); Mining (0.16%); Construction (0.57%); Manufacturing (15.76%); Transportation & Public Utilities (6.73%); Wholesale Trade (1.81%); Retail Trade (15.27%); Finance, Insurance, Real Estate (35.88%); Services (22.50%); Public Administration (0.66%); and Other (0.49%).

A5: Number of Data Breaches by Year

Figure A2 presents the number of data breaches by year. 2005 has the lowest number of data breaches, but this could stem from measurement issues in early data collection. The number of data breaches varies widely over years; 2014 has the highest number of data breaches. Although it is possible that the data collection becomes more accurate and that more data breaches were discovered in later periods, this should bias against my results.

Figure A2: Number of Data Breaches by Year



Notes: This figure presents the number of data breaches by year. After manually matching with public companies, there are 1,214 observations (524 unique firms).

A6. Repeat Analyses Incorporating Audit Analytics Data

In this section, I repeat the main analyses in the table while incorporating the Audit Analytics Data. Although Audit Analytics data are well-organized and do not require manual collection and checking, I choose not to use it as the main data source for two reasons: (1) Audit Analytics data starts from 2011, which limits my identification strategies and variations. (2) Audit Analytics data only includes cyberattacks, while the PRC dataset also includes non-cyberattack breaches, which closely relates to firms' internal controls. In Table A2, I find similar but stronger results across different shocks as compared to the results documented in Tables 4 and 5.

Table A2: Repeat Analyses Incorporating Audit Analytics Data

Dependent Variable: Breach	*PCAOB First-Time Inspection*		*Auditors Learning From Restatements*		*Auditors Learning From Data Breaches*		
					OC in OGA	*OC in OGA*	*OC*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Different Shocks	-0.006**	-0.006**	-0.010***	-0.010***	-0.005***	-0.005***	-0.018*
	(-2.42)	(-2.39)	(-8.34)	(-8.13)	(-2.66)	(-2.79)	(-1.89)
Size		0.003***		0.003***		0.003***	0.003***
		(5.05)		(5.02)		(5.00)	(4.69)
Loss		0.002		0.002		0.002	0.002
		(1.33)		(1.34)		(1.31)	(1.29)
Asset Intangibility		0.019***		0.019***		0.019***	0.019***
		(2.97)		(3.00)		(2.98)	(2.96)
Fixed Effects							
Firm×Auditor	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	No	Yes	No	Yes	No	Yes	No
Observations (Firm-Year)	55,827	55,827	55,827	55,827	55,827	55,827	55,827
Adjusted R-squared	0.114	0.114	0.114	0.114	0.114	0.114	0.115
Cluster	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor

Notes: This table repeats the analyses in Tables 4 and 5 while including the cyberattack data from Audit Analytics. *Breach* is an indicator coded as one if the firm has a data breach in a given year, and zero otherwise. Columns (1) — (2) repeat analyses in Table 4. Columns (3) — (6) repeat analyses in Table 5. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has a negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. I include firm×auditor and year fixed effects. I cluster standard errors by auditor and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

A7. Results with Coarsened Exact Matching and Small Auditors

To facilitate direct comparisons of firms with similar size, performance, and asset intangibility, in this section, I match firms based on size, performance, and asset intangibility using Coarsened Exact Matching ("CEM") (see Blackwell et al., 2009). I coarsen the sample into 20 CEM bins for size and asset intangibility with a tradeoff between preserving observations and the ex-post similarity of the distributions of matching variables across the treatment and control groups. I let firms have an exact match for loss indicators.

In addition to the CEM matching, I also consider the influence of big auditors on my empirical analyses. The variation in the result of *PCAOB First-time Inspection* comes solely from differences in the timing of inspections. Variation from big auditors' clients does not play an important role because of the firm×auditor fixed effects and keeping big auditors' clients in the analyses help us compare coefficients across different shocks. However, to empirically test the variations from small auditors' clients, I present results with only smaller auditors' clients in Column (1) using matched control variables (to increase the statistical power, e..g, Freyaldenhoven et al. 2019). The variations in *Auditors Learning from Restatements* and *Auditors Learning from Data Breaches* come from two sources: (1) the different timing of "auditors learning" and (2) the change in a treated auditor's *other* clients' behavior over time. Thus, it is important to keep big auditors in the sample because their reputational incentives are stronger than are small auditors' after an audit failure (e.g., Weber et al. 2008; DeFond and Zhang 2014). For this reason, my matching results for *Auditors Learning from Restatements* and *Auditors Learning from Data Breaches* include both big and small auditors' clients.

In Table A3, I show that CEM results are similar to those presented in Tables 4 and 5. While the matching can help ensure similarity on some observables, it assumes that firms with similar observables react similarly to concurrent events. Furthermore, it alters the sample composition instead of estimating treatment effects for the population of firms. Lastly, the evidence regarding the parallel trends assumption is reassuring.

Table A3: Results with Coarsened Exact Matching

Dependent Variable: Breach	*PCAOB First-Time Inspection*		*Auditors Learning From Restatements*		*Auditors Learning From Data Breaches*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Different Shocks	-0.004*	-0.004*	-0.008***	-0.008***	-0.014***	-0.015***
	(-1.64)	(-1.64)	(-6.16)	(-6.03)	(-7.55)	(-7.62)
Size		0.001		0.003***		0.005***
		(0.69)		(5.41)		(9.19)
Loss		-0.000		0.004**		0.004*
		(-0.02)		(2.37)		(1.79)
Asset Intangibility		-0.000		0.012***		0.017*
		(-0.17)		(2.60)		(1.85)
Fixed Effects						
Firm×Auditor	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	No	Yes	No	Yes	No	Yes
Observations (Firm-Year)	13,796	13,796	52,869	52,869	54,763	54,763
Adjusted R-squared	0.035	0.035	0.106	0.107	0.141	0.141
Cluster	Auditor	Auditor	Auditor	Auditor	Auditor	Auditor

Notes: This table repeats the analyses in Tables 4 and 5 with the CEM based on size, performance, and asset intangibility. I coarsen my sample into 20 bins, reflecting a tradeoff between maintaining observations and the ex post similarity of the distributions. *Breach* is an indicator coded as one if the firm has a data breach in a given year, and zero otherwise. Columns (1) — (2) repeat analyses in Table 4 with only small auditors' clients. Columns (3) — (6) repeat analyses in Table 5. *Firm Controls* include *Size, Loss,* and *Asset Intangibility*. *Size* is the natural log of total assets. *Loss* is an indicator coded as one if the firm has negative income, and zero otherwise. *Asset Intangibility* is defined as one minus the proportion of PPE in total assets. I include Firm×Auditor and year fixed effects. I cluster standard errors by auditor and report t-statistics in parentheses. *, **, and *** indicate statistical significance (two-sided) at the 10%, 5%, and 1% levels, respectively.

A8. Effective Dates of State Security Breach Notification Laws

Table A4: Effective Date of State Security Breach Notification Law

State	*Effective Date*	*Statute*
Alabama	1-Jun-18	Ala. Code § 8-38-1 et seq
Alaska	1-Jul-09	Alaska Stat. § 45.48.010 et seq
Arizona	31-Dec-06	Ariz. Rev. Stat. § 18-551 et seq
Arkansas	12-Aug-05	Ark. Code § 4-110-101 et seq
California	1-Jul-03	Cal. Civ. Code § 1798.80 et seq; Cal. Health & Safety Code § 1280.15
Colorado	1-Sep-06	Colo. Rev. Stat. § 6-1-716
Connecticut	1-Jan-06	Conn. Gen. Stat. § 36a-701b
Delaware	28-Jun-05	Del. Code Ann. tit. 6 § 12B-101 et seq
District of Columbia	1-Jul-07	D.C. Code § 28-3851 et seq
Florida	1-Jul-14	Fla. Stat. § 501.171
Georgia	5-May-05	Ga. Code § 10-1-910 et seq
Hawaii	1-Jan-07	Haw. Rev. Stat. § 487N-1 et seq
Idaho	1-Jul-06	Idaho Code § 28-51-104 et seq
Illinois	27-Jun-06	815 Ill. Comp. Stat. 530/5 et seq
Indiana	1-Jul-06	Ind. Code § 24-4.9-1-1 et. seq
Iowa	1-Jul-08	Iowa Code § 715C.1 et seq.
Kansas	1-Jan-07	Kan. Stat. § 50-7a01 et seq
Kentucky	15-Jul-14	Ky. Rev. Stat. § 365.732
Louisiana	1-Jan-06	La. Rev. Stat. § 51:3071 et seq La. Admin. Code tit. 16, § 701
Maine	31-Jan-06	10 Me. Rev. Stat. § 1346 et seq.
Maryland	1-Jan-08	Md. Code Com. Law § 14-3501 et seq.
Massachusetts	31-Oct-07	Mass. Gen. Laws 93H § 1 et seq
Michigan	2-Jul-07	Mich. Comp. Laws §§ 445.63, .72
Minnesota	1-Jan-06	Minn. Stat. § 325E.61.
Mississippi	1-Jul-11	Miss. Code § 75-24-29
Missouri	28-Aug-09	Mo. Rev. Stat. § 407.1500
Montana	1-Mar-06	Mont. Code §§ 30-14-1701 - 1702, 1704
Nebraska	14-Jul-06	Neb. Rev. Stat. § 87-801 et seq.
Nevada	1-Jan-06	Nev. Rev. Stat. 603A.010 et seq.
New Hampshire	1-Jan-07	N.H. Rev. Stat. §§ 359-C:19 - C:21; N.H. Rev. Stat. § 332-I:5
New Jersey	1-Jan-06	N.J. Stat. §§ 56:8-161, 163, 165 - 166
New Mexico	16-Jun-17	N.M. Stat. §§ 57-12C-1 - 57-12C-12
New York	7-Dec-05	N.Y. Gen. Bus. Law § 899-aa
North Carolina	1-Dec-05	N.C. Gen. Stat. §§ 75-61, 75-65
North Dakota	1-Jun-05	N.D. Cent. Code §§ 51-30-01 - 07
Ohio	17-Feb-06	Ohio Rev. Code §§ 1349.19 - 192
Oklahoma	1-Nov-08	Ok. Stat., Tit. 24, §§ 161 - 166
Oregon	1-Oct-07	Or. Rev. Stat. §§ 646A.600 - 646A.628
Pennsylvania	20-Jun-06	73 Pa. Stat. § 2301 et seq
Rhode Island	1-Mar-06	R.I. Gen. Laws §§ 11-49.3-1 - 11-49.3-6
South Carolina	1-Jul-09	S.C. Code Ann. § 39-1-90
South Dakota	1-Jul-18	SDCL §§ 22-40-19 - 22-40-26
Tennessee	1-Jul-05	Tenn. Code Ann. §§ 47-18-2105-2107
Texas	1-Apr-09	Tex. Bus. & Com. Code §§ 521.002, 521.053, 521.151-152
Utah	1-Jan-07	Utah Code §§ 13-44-101 et seq.
Vermont	1-Jan-07	9 V.S.A. §§ 2430, 2435
Virginia	1-Jul-08	Va. Code § 18.2-186.6; Va. Code § 32.1-127.1:05; Va. Code § 58.1-341.2
Washington	24-Jul-05	Wash. Rev. Code § 19.255.010 et seq.
West Virginia	6-Jun-08	W.V. Code § 46A-2A-101 et seq
Wisconsin	31-Mar-06	Wis. Stat. § 134.98
Wyoming	1-Jul-07	Wyo. Stat. §§ 40-12-501, 40-12-502

A9. Examples of Firms' Disclosure and Practitioners' Discussions

a. Target 2013 10K Disclosure

"The Data Breach we experienced involved the theft of certain payment card and guest information through unauthorized access to our network. Our investigation of the matter is ongoing, and it is possible that we will identify additional information that was accessed or stolen, which could materially worsen the losses and reputational damage we have experienced. For example, when the intrusion was initially identified, we thought the information stolen was limited to payment card information, but later discovered that other guest information was also stolen."

b. Assure Professional Discussion on the Relationship between Audit and Target Data Breach (April 29, 2014)

"The massive data breach that Target incurred this winter was a textbook example of why audits are so important, especially when it comes to financial data."

c. The SEC's Cease-and-Desist Order on Yahoo! for Failing to Disclose Data Breaches

2. Despite its knowledge of the 2014 data breach, Yahoo did not disclose the data breach in its public filings for nearly two years. To the contrary, Yahoo's risk factor disclosures in its annual and quarterly reports from 2014 through 2016 were materially misleading in that they claimed the company only faced the risk of potential future data breaches that might expose the company to loss of its users' personal information stored in its information systems, as well as potential future litigation, remediation, increased costs for security measures, loss of revenue, damage to its reputation, and liability, without disclosing that a massive data breach had in fact already occurred. Yahoo management's discussion and analysis of financial condition and results of operations ("MD&A") in those reports was also misleading to the extent it omitted known trends or uncertainties with regard to liquidity or net revenue presented by the 2014 data breach.

d. Examples of Discussions from Practitioners and the Media

FIRM MGMT

How Accountants Can Help Clients Avoid Data Breaches

CHRIS NOVAK, GLOBAL DIRECTOR, RISK TEAM WITH VERIZON ENTERPRISE SOLUTIONS ON APR 14, 2017

Ia INTERNAL AUDITOR

What Can Auditors Do About Data Breaches?

In the wake of the Target incident, internal auditors should provide assurance that basic security measures are in place in their organization's commerce system.

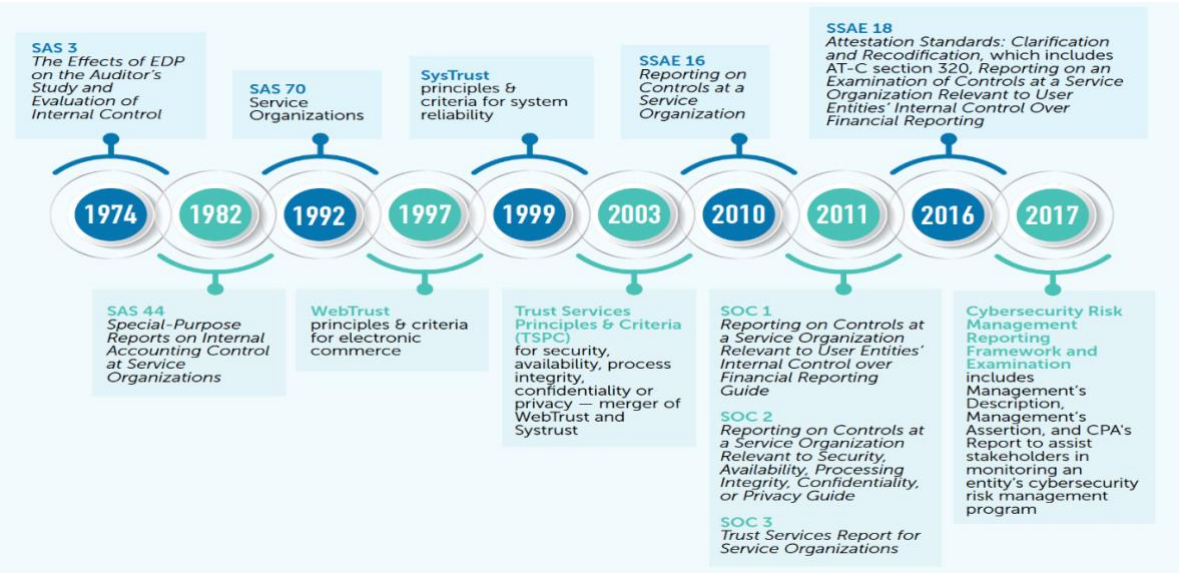
MarketWatch

Equifax auditors are on the hook for data security risk controls

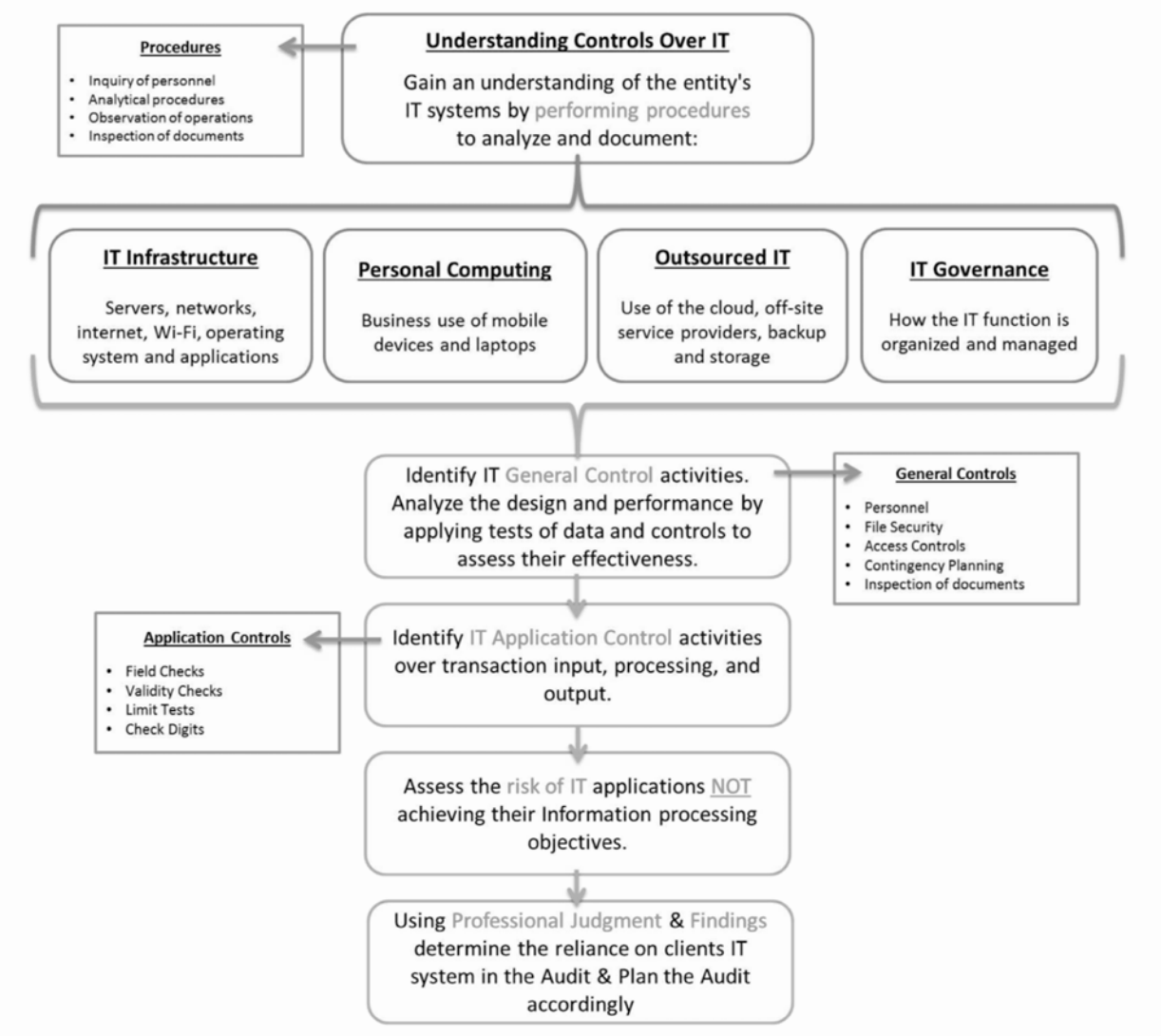
By Francine McKenna
Published: Oct 3, 2017 9:50 a.m. ET

Before an auditor reviews numbers, it must make sure that execs set the right "tone at the top" on controls, including of IT systems

Auditing IT Controls *(Source: AICPA and Center for Audit Quality):*



e. One Example of Auditing Procedures on Controls over IT:



A10. Simple Stylized Model

To formalize the intuition of the two channels in the paper, I provide a simple stylized model to illustrate my cross-sectional predictions. Let θ be the true vulnerability/weakness in a firm's overall data system. Let θ_i be the vulnerability/weakness in a firm's audited financial data systems. Let θ_j be the vulnerability/weakness in this same firm's other data system. The firm does not know the true values of θ_i and θ_j, but knows the distributions. $\mathbb{E}(\theta_i) = \bar{\theta}$, $Var(\theta_i) = \sigma_\theta^2$, and $cov(\theta_i, \theta_j) = \zeta\sigma_\theta^2$, where ζ measures the degree to which firms' different data systems are interrelated.

Let s_i be the signal related to the vulnerability/weakness in firms' audited financial data systems. The signal structure is as follows: $s_i = \theta_i + \epsilon_i$, where $\theta_i \sim N(\bar{\theta}, \sigma_\theta^2)$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Also, $cov(\epsilon_i, \epsilon_j) = 0, \forall j \neq i$.

Let $\tau_\epsilon = \frac{1}{\sigma_\epsilon^2}$ be the precision of signal s_i, and let $\tau_\theta = \frac{1}{\sigma_\theta^2}$ be the precision of θ_i. Then, $\mathbb{E}(\theta_i|s_i) = \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon} s_i + \left(1 - \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}\right)\bar{\theta}$

The reason an informative signal from the audited data system could transfer to other data systems is that s_i could help make inferences about θ_j.

When $\zeta \in (0,1)$, $\mathbb{E}(\theta_j|s_i) = \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}\zeta s_i + \left(1 - \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}\zeta\right)\bar{\theta}$.

When $\zeta = 1$, vulnerabilities in different data systems are perfectly correlated, so $\mathbb{E}(\theta_j|s_i) = \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon} s_i + \left(1 - \frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}\right)\bar{\theta}$.

When $\zeta = 0$, vulnerabilities in different data systems are completely independent, so $\mathbb{E}(\theta_j|s_i) = \bar{\theta}$.

Thus, information transfers only work when $\zeta \in (0,1]$. When the two systems are economically related, the signal of one leads to inferences about the other.

The comparative statics are determined below:

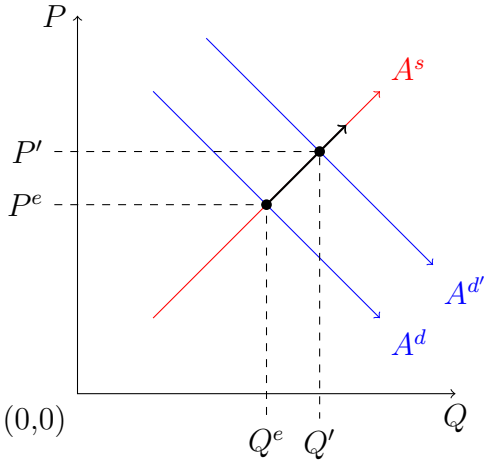
1. information transfers increase in ζ (i.e., correlations between the vulnerabilities in different data systems).

2. information transfers increase in $\frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}$ (i.e., informativeness in the signal regarding vulnerability in the audited data system).

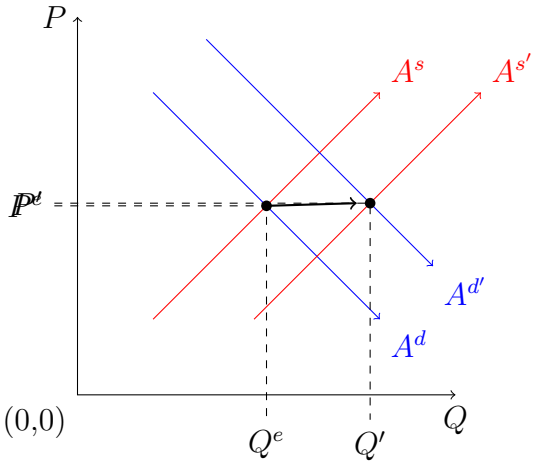
An important factor of $\frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}$ is audit quality. In my empirical specifications, I explore the variation on the supply side of audit services in order to construct the shock on improved

audit quality. I explore scenarios in which audit processes are strengthened: a shift in the supply curve and movement along the supply curve. Prior research finds that audit fees increase after the PCAOB inspections and after data breaches, which I also verify in my setting. Thus, both the quantity and the price increase in the shock of the PCAOB and of data breaches. This could be viewed as a movement along the supply curve. I also find that audit fees do not increase after restatements. This could be a shift in the supply curve. Simple graphical illustrations are below (the x axis is audit quality and the y axis is the price of audit quality).

A Movement along the Supply Curve



A Shift in the Supply Curve



Because firms are ultimately responsible for the protection of data, any informative signals recommended by auditors should be implemented on the demand side. To reach the new equilibrium point, a shock on the supply side should affect (indirectly) the demand side. Furthermore, in equilibrium, the demand curve shifts in order to rationalize the empirical finding that prices either increase or stay the same.[1] Thus the shift in the demand curve is (indirectly) affected by the plausible exogenous variation I exploit on the supply side.[2]

This simple theoretical setup offers one cross-sectional prediction: the implementation elasticity on the demand side (i.e., the change in the implementation of data protection in response to changes in audit services). This elasticity could determine the new equilibrium point. When firms are more receptive to auditors' ex post monitoring, the ex ante incentive for strengthening internal controls is stronger. There is another cross-sectional prediction obtained from the first part's signal correlation structure: in more integrated data systems, information transfer from audited data to other systems is more likely. Based on these ideas, I use several proxy variables to provide cross-sectional predictions:

- Proxy variables for ζ: firms' explicit disclosure of "Enterprise Resource Planning" (ERP), "Integrated Information Systems," and "Integrated Database" in their financial statements (including the 10-K, 10-Q, and 8-K) and in press releases.
- Proxy variables for $\frac{\tau_\epsilon}{\tau_\theta+\tau_\epsilon}$: shocks on audit quality.
- Proxy variables for implementation elasticity in the setting:
  - Audit committees
  - Internal control weaknesses

[1] The pricing outcome in equilibrium is also consistent with the interviews with partners.

[2] I acknowledge that the model has no explicit decision problem (i.e., firms do not have choice variables). The optimal level of data protection could vary (satisfying both first-order and second-order conditions) depending on a firm's prior conditions (e.g., information set, budget constraints, risk aversion); Gordon and Loeb (2002) presents an economic model to explicitly model firms' decision problem and optimal level of data protection. In this appendix, I use the supply and demand curve to graphically illustrate how the equilibrium point changes, i.e., the shift in the demand curve incorporates a firm's optimal data protection decisions.