Trade Secrets and Cybersecurity Breaches

Michael Ettredge, Distinguished Professor
Accounting Area
School of Business
University of Kansas
mettredge@ku.edu

Feng Guo, Assistant Professor
Debbie and Jerry Ivy College of Business
Iowa State University
fengguo@iastate.edu

Yijun Li, Doctoral Student
Accounting Area
School of Business
University of Kansas
yijunli@ku.edu

June 2018

We acknowledge helpful comments from Tom Kubick, Adi Masli and Mike Wilkins. We appreciate helpful suggestions from the editors and a reviewer. We also thank SeekEdgar LLC for support regarding the 10-K data. All errors are our own.

Trade Secrets and Cybersecurity Breaches

Abstract: We study the association between firms' disclosures in Forms 10-K of the existence of trade secrets, and cyber theft of corporate data (which we refer to as "Breaches"). Prior academic research explaining occurrence of Breaches is scarce, and no prior study has focused specifically on Breaches that likely target trade secrets. We provide such evidence, and our use of Form 10-K contents related to trade secrets is a first step toward determining whether corporations actually attract Breach activity through their public disclosures. We find that firms mentioning the existence of trade secrets have a significantly higher subsequent probability of being Breached relative to firms that do not do so. Our results are stronger among younger firms, firms with fewer employees, and firms operating in less concentrated industries. By conducting a battery of additional tests, we attempt to go beyond merely establishing correlations to provide evidence whether such proprietary information can actually attract cyberattacks. Specifically, our results are robust to additional control variables, an instrumental variable approach, firm fixed effects, and a propensity score matching technique.

Key words: Cyber attack, hackers, IT security, trade secrets, disclosure, breaches

Trade Secrets and Cybersecurity Breaches

"*The threat is incredibly serious—and growing. Cyber intrusions are becoming more commonplace, more dangerous, and more sophisticated. Our nation's critical infrastructure, including both private and public sector networks, are targeted by adversaries. American companies are targeted for trade secrets and other sensitive corporate data ...*" U. S. Federal Bureau of Investigation[1]

1. Introduction

We study the association between firms' disclosures in Form 10-K of the existence of corporate proprietary information, namely trade secrets, and cyber theft (which we refer to as cybersecurity breaches, or "Breaches" for brevity). Trade secrets consist of a vast variety of information, including "all forms and types of financial, business, scientific, technical, economic, or engineering information, including patterns, plans, analyses, program devices, formulas, designs, prototypes, methods, techniques, processes, procedures, programs, or codes, whether tangible or intangible …"[2] Prior academic research explaining occurrence of Breaches is scarce, and no prior study has focused specifically on Breaches that likely target trade secrets. We provide such evidence, and our use of Form 10-K contents related to trade secrets is a first step toward determining whether corporations actually attract Breach activity through their public disclosures.

Trade secrets are one of the primary means through which firms create and maintain value. A firm's ability to prevent trade secrets from being stolen, copied or eroded is one of the key factors ensuring its longevity. Even so, trade secret theft has become a serious threat to the U.S. economy, so much so that "State sponsored trade-secret theft . . . embattles [the United

[1] Federal Bureau of Investigation web site https://www.fbi.gov/investigate/cyber as of May, 2018.
[2] See Section 1839 of Public Law 104-294, the Economic Espionage Act of 1996, available at https://www.gpo.gov/fdsys/pkg/STATUTE-110/pdf/STATUTE-110-Pg3488.pdf.

States'] status as world leader in innovation…".[3] A recent study by PricewaterhouseCoopers claims that damage from trade secret theft in the U.S. is in the range of one to three percent of its GDP.[4] A study by Ocean Tomo LLC (2015) estimates that intellectual property assets, including trade secrets, can constitute more than 80 percent of an S&P 500 company's value.[5] Such sources of wealth constitute tempting targets for thieves. Traditionally, most legal cases involving trade secret thefts attribute the thefts to firms' former employees (Almeling et al., 2010; Klasa et al., 2018). However, thefts of trade secrets via cyber-attacks against U.S. firms have proliferated over the past few years and have drawn increasing attention from U.S. government officials.[6] Due to the very nature of trade secrets (i.e., highly valuable and carefully protected), information about the existence of trade secrets is likely to attract cyber-attacks by hackers who are well-funded and organized, and who are motivated primarily by financial gain.[7] Trade secrets are relatively more vulnerable to cyber-attacks, compared to other types of intellectual property such as patents, trademarks, and copyrights (Villasenor, 2015). We therefore posit that disclosure of the existence of trade secrets in Form 10-K can trigger subsequent Breaches.

We employ textual analyses of Form 10-K contents to identify information used to test our hypothesis. First, we rely on a novel and parsimonious disclosure-based measure to gauge whether firms rely upon trade secrets. Specifically, following Glaeser (2018), we search firms' 10-K filings for the phrases "trade secret" and "trade secrecy." Next, we manually identify and collect data on Breach incidents (cybersecurity attacks) from various public data sources. We

[3] Available at: http://www.washingtonpost.com/world/national-security/us-launches-effort-to-stem-trade-secret-theft/2013/02/20/26b6fbce-7ba8-11e2-a044-676856536b40_story.html
[4] Available at: https://create.org/wp-content/uploads/2014/07/CREATe.org-PwC-Trade-Secret-Theft-FINAL-Feb-2014_01.pdf
[5] Available at: http://www.oceantomo.com/blog/2015/03-05-ocean-tomo-2015-intangible-asset-market-value/.
[6] For example, the White House issued an Executive Order on December 29, 2016 to take "Additional Steps to Address the National Emergency with Respect to Significant Malicious Cyber-Enabled Activities."
[7] Anecdotal evidence indicates that hackers target technological innovations, manufacturing metrics, design specifications, business strategy, and litigation plans among other information. See for example: http://time.com/106319/heres-what-chinese-hackers-actually-stole-from-u-s-companies/.

first document that firms mentioning the existence of trade secrets have a significantly higher probability of subsequently being Breached relative to firms that do not disclose such phrases. Additionally, our results are more pronounced among younger firms, firms with fewer employees, and firms operating in more competitive industries. These cross-sectional results are consistent with the notion that firms' trade secrets are more likely to be hacked when the trade secrets are more valuable or when alternative ways to obtain firms' trade secrets, such as hiring away firms' employees, are not available.

Our main results suggest a positive association between disclosure of the existence of trade secrets and subsequent data breaches. However, existence of trade secrets and data breaches could be jointly determined by a firm's observable or unobservable characteristics. To better establish an inference of causality between proprietary information and cyberattacks, we conduct a battery of additional tests. First, we identify additional control variables based on Gordon et al. (2010) to control for endogenous disclosure decisions related to cyber-defense and cyber-vulnerability. Our main inferences remain similar after we control for these additional control variables. Second, we make use of the staggered adoption, by most states in the U.S., of trade secret protections across our sample period, as an instrumental variable for firms' reliance on trade secrets. We obtain a similar result. To rule out that our results are driven by time-invariant and un-observable industry or firm characteristics, we add industry and firm fixed effects. Our results are robust to additional fixed effect estimators. Last, we employ a propensity score matching technique. Our results do not appear to be driven by selection bias due to observables. Collectively, our results suggest a robust positive relation between firms' public disclosure of reliance on trade secrets and subsequent cyberattacks, after accounting for omitted variable bias caused by observable and un-observable factors.

We make three important contributions to the literature. First, our study adds to the small but growing literature that studies the determinants of occurrences of Cyber Breaches. Constantly changing technologies and increasingly sophisticated tools in the past 30 years pose substantial danger to U.S. companies' cyber security. Prior studies focus on the role of IT governance (Higgs et al., 2016) and cyber security risk factors (Wang et al., 2013). We add to this literature by examining whether disclosure of the existence of trade secrets in 10-Ks is positively associated with subsequent Breaches.

Second, we contribute to the literature on corporate disclosure policies. From a manager's perspective, our results suggest that it is useful to understand the unintended consequences of mentioning the existence of trade secrets in Form 10-K disclosures. While Glaeser (2018) identifies potential benefits from such disclosure, such as providing information to investors and legal benefits, our paper provides managers with an additional consideration when setting disclosure policy, that is, mentioning the existence of trade secrets can attract cyberattacks.

Finally, we also contribute to the nascent literature on the consequences of trade secrecy. Despite the economic significance of trade secrets to U.S. firms, trade secrecy has received little attention in the literature empirically due to data limitations (i.e. information about trade secrets is carefully protected). We use a new empirical measure of trade secrecy based on Glaeser (2018) to document a potential cost of relying on trade secrecy, rather than patenting, to protect firms' intellectual properties. We also highlight the importance of an additional step that managers should consider to protect firms' trade secrets from cyber threats: non-disclosure in Forms 10-K of the existence of trade secrets.

In Section 2 we provide background on trade secrets, cybersecurity, hackers, and Breaches. We also place our study in the context of prior literature and state our hypothesis. In

Section 3 we describe our methods for identifying Breach incidents, and for identifying firms' disclosures in Form 10-Ks of the existence of trade secrets. We then discuss our sample, variables, and models. Section 4 presents our main results. Section 5 provides additional analyses bearing upon the robustness of our results, and Section 6 concludes.

2. Background and Hypothesis Development

2.1 Trade Secrets

Firms use various methods to protect their intellectual property. Prior studies often focus on the importance of patenting (Hall et al., 2014). The patent system provides a patentee with exclusive rights to use an invention for a limited duration of time. However, the patent system requires patentees to disclose the technological details of their inventions. In addition, the inventions must meet the relevant patentability requirements, such as novelty, usefulness, and non-obviousness. Trade secrets provide an alternative to patents as a means of protecting proprietary information. A prefatory note to the Uniform Trade Secrets Act (UTSA), promulgated by the National Conference of Commissioners on Uniform State Laws (1985, 1), states:

> A valid patent provides a legal monopoly for seventeen years in exchange for public disclosure of an invention. If, however, the courts ultimately decide that the Patent Office improperly issued a patent, an invention has been disclosed to competitors with no corresponding benefit. In view of the substantial number of patents that the courts invalidate, many businesses now elect to protect commercially valuable information by relying on the state trade secret protection law.

The UTSA (1985, 5) defines a trade secret as information "including a formula, pattern, compilation, program, device, method, technique, or process, that:

> (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and

> (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy."

Trade secrets lack the legal protection provided by patent status, but trade secrets have the advantage that technological details need not be revealed to the public, as is the case with patents. Some evidence suggests that trade secrets can be a more effective means to protect proprietary knowledge than patents (Cohen et al., 2000; Arundel, 2001; Jensen and Webster, 2009). Trade secrets comprise a large portion of firms' intellectual property assets (Ocean Tomo LLC, 2015).[8] Famous trade secrets include Google's search algorithms, Coca-Cola's ingredients, Big Mac's special sauce, the manufacturing process for producing the lubricant WD-40, etc.[9]

Despite the economic importance of trade secrets, few empirical studies have investigated trade secret theft, presumably due to the difficulty in obtaining data on theft incidents (Glaeser, 2018; Png, 2017 a b). Trade secret theft is a serious threat to the U.S. economy and the link between trade secret theft and cybersecurity has drawn attention from U.S. government officials, some of whom have stated that "cyberspace is an increasingly important avenue for espionage."[10] We turn next to a discussion of cybersecurity, hackers, and Breaches.

2.2 Cybersecurity, Hackers and Breaches

Cybersecurity issues have generated significant attention among academics, the media, practitioners, and regulators. Accounting practitioners have acknowledged the importance of cybersecurity risks. The AICPA (2016) issued an exposure draft of "Proposed criteria for management's descriptions of an entity's cybersecurity risk management program." The Center for Audit Quality (2017, 10) describes the elements of the proposed voluntary cybersecurity reporting framework, which include a CPA's report that "contains an opinion on the description

[8] Available at http://www.oceantomo.com/blog/2015/03-05-ocean-tomo-2015-intangible-asset-market-value/.
[9] Available at http://info.vethanlaw.com/blog/trade-secrets-10-of-the-most-famous-examples
[10] Available at https://www.washingtonpost.com/world/national-security/us-launches-effort-to-stem-trade-secret-theft/2013/02/20/26b6fbce-7ba8-11e2-a044-676856536b40_story.html?utm_term=.c39c57d4ce17

of the entity's cybersecurity risk management program and the effectiveness of the controls within the program to achieve the entity's cybersecurity objectives." Following a number of high-profile cybersecurity incidents (e.g. Equifax Breach), including one that targeted the SEC itself, the SEC updated its guidance for public companies to disclosure cybersecurity risks and incidents in February 2018.[11]

Based on a 2015 study from the Ponemon Institute, the number of cyber-attacks against U.S. companies continues to grow not only in frequency but also in severity.[12] The National Computer Security Survey (NCSS), co-sponsored by the Bureau of Justice Statistics and the National Cyber Security Division (NCSD) of the U.S. Department of Homeland Security, acknowledges three general types of cybercrime:[13]

- Cyber attacks are crimes in which the computer system is the target. Cyber attacks consist of computer viruses (including worms and Trojan horses), denial of service attacks, and electronic vandalism or sabotage.
- Cyber theft comprises crimes in which a computer is used to steal money or other things of value. Cyber theft includes embezzlement, fraud, *theft of intellectual property*, and theft of personal or financial data. [*Emphasis added*.]
- Other computer security incidents encompass spyware, adware, hacking, phishing, spoofing, pinging, port scanning, and theft of other information, regardless of whether the breach was successful.

We restrict our attention to "cyber theft" in which intruders (commonly referred to as "hackers") obtain access to data held in corporate information systems. These are distinct from "cyber attacks" and from "other computer security incidents" that do not have the goal of stealing data.

[11] The SEC's prior guidance was CF Disclosure Guidance: Topic No. 2 – Cybersecurity, which was issued in October, 2011, and that expressed the Division of Corporation Finance's view on registrants' obligations to disclose cybersecurity risks and incidents. The new guidance is generally consistent with the 2011 guidance, but adds two topics: disclosure controls and procedures, and insider trading. For more information, see https://www.pwc.com/us/en/cfodirect/publications/in-brief/sec-cybersecurity-risk-disclosures.html?elq_mid=10426&elq_cid=828593.

[12] Available at: http://www.prnewswire.com/news-releases/ponemon-institutes-2015-global-cost-of-data-breach-study-reveals-average-cost-of-data-breach-reaches-record-levels-300089057.html.

[13] Bureau of Justice Statistics web site https://www.bjs.gov/index.cfm?ty=tp&tid=41 as of June, 2018. The web site states that: "The goal of NCSS is to produce reliable national and industry-level estimates of the prevalence of computer security incidents (such as denial of service attacks, fraud, or theft of information) against businesses and the resulting losses incurred by businesses."

Hackers that attack corporate information systems have a variety of motives and goals. Holt and Kilger (2012) assert that these include the following. Hackers have an intense desire to understand and master information technology. They can gratify these desires by hacking systems for the entertainment value, without consideration of any financial reward. Some hackers are motivated primarily by ego. Such hackers are status-conscious, and they gain status in their sub-culture through their ability to hack challenging systems in novel ways. These hackers might steal sensitive information primarily as evidence of hacking accomplishments rather than for financial reward. Some hackers are motivated by the desire to belong to a social group. Entrance into and opportunities within a variety of hacker groups are facilitated by attention-getting exploits. Successful hackers may be invited to join closed forums in which members exchange tools and information not otherwise available. Still other hackers are motivated primarily by a cause. Such hackers use the internet to advocate for their political, nationalist, religious, environmental, or other beliefs. These hackers might hijack web sites that express opinions or sell products that they oppose.

We are most interested in those hackers who are motivated by money. The volume and potential value of information available via the internet have exploded in recent years. Holt and Kilger (2012, 10) state: “money has become a particularly important motivation for malicious and criminal hackers over the last two decades.” Hackers sell certain types of personal data they have stolen from retailers and financial institutions, such as credit card information, for profit. However, trade secrets obtained via hacking arguably are more difficult to monetize compared to credit card information or social security numbers. Trade secrets are most likely to be stolen not by amateur hackers or informal hacker groups, but by well-trained and well-supported hackers on behalf of companies that can use such information. For example, Malawer (2014, 1) states:

"Chinese economic cyber espionage, government hacking into computer networks of companies to gain commercial advantage for Chinese firms, is one of the most complex issues confronting U.S. national security and foreign policy today."

2.3 Prior Literature

Prior empirical, archival studies of cybersecurity fall into several major categories. One category consists of studies that investigate cyber security investment decisions. For example, Gordon and Loeb (2002), Gordon et al. (2003), Gordon et al. (2015a), Gordon et al. (2015b), and Gordon et al. (2015c) utilize economic-based models to provide guidance on the optimal level of cybersecurity investments and to assess to what extent government regulations could induce firms to increase their investments in cybersecurity. Another stream of literature analyzes the determinants and consequences of corporate disclosures of information security activities. Gordon et al. (2006) document that firms provide significantly more disclosures of information about security activities in their annual filings after passage of the Sarbanes-Oxley Act (SOX) in 2002. One study finds that firms that voluntarily disclose items concerning information security have higher market valuations by investors (Gordon et al., 2010). A third major category consists of studies of the consequences of cyber security incidents, especially stock market reactions. Spanos and Angelis (2016) identify 37 such papers. Lawrence et al. (2018) report that Breaches explain (are positively associated with) subsequent instances of financial reporting control weaknesses, restatements, receipt of SEC comment letters, and audit fees.

A limited but growing body of studies attempt to explain occurrences of Breaches. Higgs et al. (2016) report, in their strongest results (Table 4), that the probability that a Breach is reported in a year is positively associated with a firm's contemporaneous size, R&D expenditures, leverage, and disclosure of Breaches in prior years. It is negatively associated with the occurrence of a loss. The focus of their study is on the role of governance. They find that Breaches are positively associated with the existence of corporate risk committees, compliance committees, and technology committees.

The prior paper most closely related to ours is Wang et al. (2013). Those authors investigate whether companies' disclosures in 10-Ks of cyber security risk factors are associated with subsequent announcements of a variety of cyber security incidents. The authors argue that managers truthfully disclose their firms' cybersecurity conditions to avoid future litigation risk. They employ textual analyses of 10-Ks filed immediately prior to these incidents to identify security risk factors disclosed by managers. They find that disclosures of risk-mitigation factors are negatively associated with subsequent cyber security incidents. Our paper differs from theirs in a number of ways, but the most important is that we examine whether disclosure of the existence of trade secrets in 10-Ks is positively associated with subsequent Breaches.

2.4 Hypothesis

In this paper, we study the association between the disclosure in Form 10-K of the existence of corporate proprietary information, namely trade secrets, and Breaches. A positive association is plausible because (1) managers have incentives to mention the existence of trade secrets in Form 10-K, and because (2) hackers have incentives to attack firms that make such disclosures. With respect to the first incentive, Glaeser (2018, 2) argues that: "Firms use these disclosures to protect themselves in the case of subsequent patent or trade secret litigation and as part of discussions of business risk factors." Managers arguably have few or no incentives to mention existence of trade secrets in other venues, due to the secrecy inherent in the very definition of a trade secret. In one survey of corporate executives, nearly 60 percent stated that the possibility of "giving away company secrets" is an important barrier to providing more disclosure (Graham et al., 2005, 62). With respect to the second incentive, Glaeser (2018, 2) states: “The existence of a trade secret is often public knowledge, while the nature of the trade secret is not. As a result, firms are willing to disclose the existence of a trade secret in their 10-K, without revealing how the trade secret works.” Disclosure of “the existence of a trade secret,” however, could attract hackers whose goal is to determine “how the trade secret works” by stealing data from corporate information systems.

On the other hand, one of the defining characteristics of trade secrets is that the information they contain must not be “readily ascertainable by proper means.” In other words, a firm’s proprietary information is not a trade secret unless the firm actively attempts to protect it. Law practitioners suggest that “For a technology company, limiting access to research files, distributing files only on a need-to-know basis, erecting network firewalls, and requiring passwords for computer access are appropriate.”[14] If firms having trade secrets employ extra care in protecting those secrets against cyberattacks, it is possible that disclosures of the existence of trade secrets are not associated with increased propensity for hacker attacks.[15] Although this “deterrence” argument is plausible, we believe that the argument for a positive association is stronger. Accordingly, we state our main hypothesis in its alternative form:

H1a: Disclosures of the existence of trade secrets are positively associated with subsequent Breaches.

3. Sample, Data and Empirical Methods

3.1 Identifying Cyber Attacks

[14] Available at: http://www.slindenelson.com/Articles/What-Is-a-Trade-Secret.shtml

[15] A negative association is conceivable, but seems unlikely.

Our study requires a test sample of companies that are successfully cyber-attacked, and that disclose the Breaches. We also require a control sample of companies that are not successfully attacked. The SEC requires firms to disclose material Breaches, i.e., those likely to affect investor decisions, which facilitates our identification of a test sample.[16] However, the SEC does not provide any specific guidance on what it considers material. Federal law, and many state laws, require companies to notify the public of Breaches that compromise personal information such as credit card data and medical records. There are no comparable notification laws for Breaches of corporate proprietary information.[17] Furthermore, it is doubtful that managers have adequate incentives to comply with required SEC disclosures of material Breaches. As one observer states, "the harm of the disclosure, both through publicizing internal vulnerabilities and reputational damage, can be worse than the initial attack."[18] In summary, it is possible that our sample of control firms not disclosing Breaches includes companies that have been successfully attacked, but that did not report the Breaches. Such misclassified control firms should bias against our ability to observe a positive association between disclosure of trade secret existence and subsequent Breaches.

No existing database of Breaches is perfectly suitable for our purposes. We therefore attempt to build a comprehensive dataset of Breaches from 2007 to 2015. We start with the ITRC/CyberScout Annual Data Breach Reports used in Lawrence et al. (2018). Identify Theft Resource Center (ITRC) is a non-profit organization that aims to provide nationwide information on data breaches and cybersecurity. It maintains a Breach database and publishes an annual data breach incident report with financial support from CyberScout, an Arizona-based cybersecurity

[16] The pros and cons of such disclosure are discussed in Wall Street Journal (2016) available at https://www.wsj.com/articles/should-companies-be-required-to-share-information-about-cyberattacks-1463968801.

[17] Appendix C of Lawrence et al., (2018) provides a summary of data breach disclosure laws.

[18] Available at https://blogs.wsj.com/riskandcompliance/2014/03/27/when-to-disclose-a-data-breach-how-about-never/.

firm. According to ITRC/CyberScout, all reported Breaches are confirmed by various media sources or governmental agencies. The data breach identification sample begins in 2007, since that is the earliest year for which we have access to ITRC/CyberScout reports.

Next, we use several additional sources to identify Breaches that might be missed by ITRC/CyberScout reports. The Heritage Foundation publishes reports on cyber-attacks that target U.S. companies.[19] However, those reports only cover incidents since 2013. We searched online for additional data depositories, and found a web site that is maintained by members of the information security community, and that has compiled a list of cyber-attacks world-wide since 2011.[20] We add Breaches that are not already included in the ITRC/CyberScout reports if we can verify the incidents against a media source.

Last, we perform an independent search on cyber-attack news in LexisNexis using its SmartIndexing Technology, an automated textual classification system. LexisNexis maintains a list of index terms to identify the subjects of relevant news, which enables researchers to quickly identify all news items that relate to a specific topic.[21] With this comprehensive sample of data breaches, we manually perform a name match with the *Compustat* to obtain attacked firms' identifiers to use in extracting SEC filings and financial statement information. Our data breach sample consists of 591 cyber incidents, representing 511 attacked firm-years observations from 318 attacked firms. We compare this number with several studies using alternative data sources, and the number of cyber-attacks we identify in our final sample is reasonably close to prior literature.[22]

[19] Available at: http://www.heritage.org/cybersecurity

[20] Available at: www.hackmageddon.com.

[21] LexisNexis constantly updates its index term list based on customer feedback and news trend. As of the time of our study, three index terms are available for identifying cyber related incidents – data security, cybercrime, cyberterrorism.

[22] Higgs et al. (2016) use a dataset from Privacy Rights Clearinghouse (privacyrights.org), and report a sample of 634 reported breaches (361 firms) between 2005 and 2014. Our sample is comparable to theirs.

One limitation of our sample of Breaches is that we cannot observe the underlying motives of all attackers. Further, the full scale of the Breaches and of the information stolen might not be available for all attacks. In fact, anecdotal evidence suggests companies often choose not to disclose the full scale of breach incidents, especially when such incidents involve intellectual properties.[23] As a result, although our test sample firms were all attacked, it is not possible to verify that in every attack a corporate trade secret was stolen. If our Breach sample contains attacks that are not specifically targeting trade secrets, this should tend to bias against the results specified in our hypothesis as this introduces noise in our dependent variable.[24]

3.2 Identifying Disclosure of Trade Secret Existence

Due to their inherent secrecy, it is impossible to observe whether any given firm actually has trade secrets (Glaeser, 2018; Png, 2017 a, b). However, we can identify which firms *disclose* that they have trade secrets. We employ SeekEdgar software to search the contents of firms' Form 10-Ks to determine the disclosure of the existence of trade secrets.[25] Following Glaeser (2018), we search for the key words "trade secret" and "trade secrecy" located anywhere in firms' 10-K disclosures. We assume that a firm relies on trade secrets in a fiscal year if the firm mentions one of these two key words in its 10-K for that year. Some examples of successful matches (in 10-K context) are as follows:

[23] http://www.foxbusiness.com/features/2011/10/28/disclosure-debate-when-should-companies-reveal-cyber-attacks.html

[24] Even if attackers do not successfully steal a specific trade secret in a Breach, they might obtain intermediate information, such as employee passwords, that will enable them to steal trade secrets in the future. Nonetheless, we perform an un-tabulated exploratory analysis by hand-collecting descriptions for all *Breaches* that happened in 2015. We use the definition of trade secret in Section 1839 of Public Law 104-294, the Economic Espionage Act of 1996, to classify all breaches. We limit this exercise to one year due to its cost, as it involves reading media coverage, notifications to government agencies, and firm disclosures. The most recent sample year, 2015, best enables us to classify all *Breaches* into *Trade Secret Specific Breaches* and *Non-Trade Secret Specific Breaches*. We find our results are concentrated in the *Trade Secret Specific Breaches* subsample. The estimated coefficient on *Trade Secret Specific Breaches* is 0.540 with a one-tailed p-value of 0.076; the estimated coefficient on *Non-Trade Secret Specific Breaches* is 0.206 with a one-tailed p-value of 0.332. However, the difference between the two coefficients is not significant at conventional levels.

[25] We employ SeekEdgar software and database (Available at: https://www.seekinf.com:8443/search.jsp). SeekEdgar is an extraction engine that facilitates text-based searches of all SEC Edgar filings.

> "We intend to continue our policy of taking all measures we deem necessary to protect our patent, copyright, *trade secret* and trademark rights. We regard our internally-developed software embedded in our products as proprietary, and we utilize a combination of patent, copyright, *trade secret* laws, internal security practices and employee invention assignment and non-disclosure agreements for intellectual property protection" [*emphasis added*].[26]

> "Our source code is protected both as a *trade secret* and as an unpublished copyrighted work. However, third parties may develop similar technology independently. In addition, effective copyright and *trade secret* protection may be unavailable or limited in some foreign countries. While protecting our proprietary technology is important to our success, our business as a whole is not significantly dependent upon any single patent, copyright, trademark or license"[27] [*emphasis added*].

> "NTIC's ZERUST® rust and corrosion inhibiting products are manufactured according to NTIC's specifications primarily by selected sub-contractors and joint ventures under *trade secrecy* agreements and/or license agreements"[28] [*emphasis added*].

Glaeser (2018) validates the use of the key words "trade secret" and "trade secrecy" by showing that use of these phrases in 10-Ks is associated with less proprietary disclosure, greater information asymmetry, and less patenting activities. However, this measure is not without error. For example, a firm might employ trade secrets, but might choose not to mention their existence in its 10-K disclosures. We argue that this may not be a serious problem. When firms mention the existence of a trade secret in 10-K disclosures, they don't reveal details about the content of the trade secret. This is evident in the examples provided above. As a result, mentioning the existence of trade secrets in a firm's mandatory disclosures does not impose any direct proprietary costs on the firm. Furthermore, firms sometimes benefit from such disclosure. First, firms could signal greater value of their stocks by discussing the existence of trade secrets, and can provide assurance to their investors by discussing how the firms take the appropriate steps to protect the trade secrets from misappropriations (Glaeser, 2018). In addition, firms could also receive legal benefits by disclosing the existence of trade secrets (Glaeser, 2018). For example,

[26] Available at: https://www.sec.gov/Archives/edgar/data/1122051/000156459015011515/xcom-10k_20150930.htm
[27] Available at: https://www.sec.gov/Archives/edgar/data/883241/000088324115000014/snps10311510-k.htm
[28] Available at: https://www.sec.gov/Archives/edgar/data/875582/000117184311003555/f10k_111811.htm

when Valspar alleged that Van Kuren committed trade secret misappropriation, the court cited Valspar's 10-K disclosure as direct evidence of Valspar's reliance on trade secrets:

> "In its Form 10-K, Valspar provided that its "knowledge and trade secret information regarding [its] manufacturing processes and materials have . . . been important in maintaining[its] competitive position" and that it "require[s] employees to sign confidentiality agreements relating to proprietary information." (Id. at 5.)"[29]

Nevertheless, in Section 5, we perform a number of robustness checks to address the voluntary nature of trade secret disclosure.

3.3 Sample and Methodology

Our *Breach* sample period covers 2007 through 2015. We start (end) with fiscal year 2007 (2015) because this is the first (last) year that we have access to the ITRC/CyberScout Data Breach Reports. Since we use information available at year t to predict breach incidents in year t+1, our *Trade Secret* and financial data period covers 2006 through 2014. We start from all 51,616 firm-year observations in *Compustat* with non-missing relevant variables. Next, we match this dataset with *Audit Analytics* and *SeekEdgar* to get information on auditors and SEC filings. The final sample consists 39,992 firm-year observations (7,462 unique firms) that allow us to construct our test variables and control variables.

Our hypothesis H1a states that disclosures of the existence of trade secrets are positively associated with subsequent Breaches. To test this, we regress whether a firm has a data breach event in year t+1 on whether the firm discloses its reliance on trade secrets in year t, along with a set of control variables that have been shown to be determinants of cyber-attacks (Kwon et al., 2013; Higgs et al., 2016). Specifically, we estimate the following logistic regression for our full sample:

[29] Please see https://www.duanemorris.com/site/static/valspar_v_vankuren.pdf for the details of the court document.

$$\begin{aligned} Pr(Data\ Breach_{t+1}) = {} & b_0 + b_1 Trade\ Secret_t + b_2 Cyber\ Defense_t + b_3 Cyber\ Vulnerability_t \\ & + b_4 SIZE_{\ t} + b_5 BTM_{\ t} + b_6 AGE_{\ t} + b_7 ROA_{\ t} + b_8 LOSS_{\ t} \\ & + b_9 R\&D_{\ t} + b_{10} Advertising_{\ t} + b_{11} BIG4_t + b_{12}\ Log(Audit\ Fee)_t \\ & + b_{13} IT\ Deficiency_t + b_{14} 10\text{-}K\ Length_t + b_{15}\ Log(FOG\ INDEX)_t \\ & + b_{16} RETAIL_t + b_{17} FINANCIAL_t + year\ dummies + \varepsilon \qquad (1) \end{aligned}$$

Data Breach is an indicator variable taking the value of one if there is at least one breach incident during year t+1. Our main test variable, *Trade Secret*, is an indicator variable for the disclosure of the existence of trade secrets at year t. The next two control variables, *Cyber Defense* and *Cyber Vulnerability*, are two broad measures capturing firms' disclosures related to cyber defense capability and vulnerability. To construct these two variables, we follow Gordon et al. (2006) and Lawrence et al. (2018) to create a list of words that could potentially capture the cyber defense mechanisms and cyber weaknesses based on firms' SEC Form 10-K filings.[30, 31] For both variables, we take the natural logarithm of the counts for use in the logistic regression. We include a number of additional control variables that represent firm size, complexity and risk in audit research. For example, we control for firm *SIZE* as the natural logarithm of total assets (AT), and *BTM* as the ratio of book value of equity (CEQ) to its market value (PRCC_F×CSHO) at fiscal year-end. We also control for firm *AGE* as the natural logarithm of the number of years since the firm's first appearance in *Compustat*. To control for profitability, we include both a continuous measure *(ROA)*, and an indicator *(LOSS)* for negative net income. *R&D* and

[30] The words (phrases) used for Cyber Defense include: risk governance, risk model, risk control, risk policy, risk framework, risk document, risk system, risk technology, risk training, risk committee, risk management, risk board, risk review, risk oversight, risk governance, chief risk officer, CRO, enterprise risk management, ERM, risk compensation, risk incentive, risk method, risk compliance, risk system, risk data integration, risk limit, risk control.

[31] The words (phrases) used for Cyber Vulnerability include: data integrity, data risk, risk report, risk dashboard, operation failure, operational failure, operation risk, operational risk, IT risk, information technology risk, privacy breach, identity theft, computer virus, security breach, hacker, cyber-attack, cyber risk, cyber security, security incident, computer breach, computer intrusion.

Advertising expenses are scaled by total assets. Another set of control variables represent audit-related inputs and outputs. *BIG4* is an indicator for the four largest international auditors (Deloitte, EY, KPMG, and PwC). *Log(Audit Fee)* is computed as the natural logarithm of total audit fees reported in Audit Analytics (audit_fees). *IT Deficiency* is an indicator variable taking a value of one if there is any internal control deficiency in information technology, software, security and access issues.[32] Given that the *Cyber Defense* and *Cyber Vulnerability* metrics are based on key words found in Form 10-K filings, we also control for the natural logarithm of the total number of words in the filings *(10-K Length)*, as well as the readibility of 10-K text *(Fog-Index).* To account for the fact that data breaches are more common in certain industries, we include indicator variables for the retail industry (SIC: 5000-5999) and the financial industry (SIC: 6000-6999). We control for retail industry and the financial industry membership, but not other industries, because these two industries have the highest proportion of cyber attacks. In our sample, the retail industry reports 2.3% (=77/3,292) of years observed exhibit data breaches, and the financial industry reports 1.8% (=170/9,343) of years observed exhibit data breaches. Although the manufacturing industries have a larger number of total data breaches, the proprotion is very small. Only 0.6% of the manufacturing firm-year observations have reported data breaches. We include year dummies to account for macro-economic changes and we cluster standard errors by firm (Petersen, 2009).[33]

4. Empirical Results

4.1 Descriptive Evidence

[32] We also replaced *IT Deficiency* with an indicator variable that captures all types of internal control weaknesses. The inferences remain the same.

[33] In additional analysis, we add industry fixed effects and firm fixed effects. Our inference holds to additional fixed-effect estimators.

Figure 1 depicts the time trend of breach incidents and the percentage of firms disclosing existence of trade secrets. Since we use disclosures of trade secret existence in year t to predict Breaches in year t+1, Figure 1a starts in 2007 and ends in 2015, while Figure 1b starts in 2006 and ends in 2014. Both the number of Breach firms and the percent of firms disclosing trade secrets exhibit long-term increasing trends. The highest number of *Breaches* (76 unique firms) occurs in 2013, coincident with the 2013 Target data breach, which involved a massive data hacking of customer list and credit card information.[34] The percent of firms disclosing the existence of trade secrets also grows steadily, with the highest percent occurring in 2014. In Figure 2, we plot data Breach incidents based on attacked firms' headquarters locations (U.S. firms only). This map suggests that companies headquartered in California and New York experience the highest number of data breach incidents during our sample period. This is consistent with the concentration of technology companies in California and of financial firms in New York. Table 1 Panel A shows the industry distribution of data breach incidents and trade secret firms. The most vulnerable industry to data breaches is financial (33.27%), followed by services (20.54%), and manufacturing (19.76%). Manufacturing (56.47%) and service industries (25.90%) also have the highest concentrations of trade secret firms. Table 1 Panel B shows the number of *Breaches* reported and the number of unique firms experiencing *Breaches* in each year.[35]

Table 2 Panel A presents summary statistics for the main sample. About 1.3% of the firm-year observations (318 unique firms) exhibit breach incidents, while about 31.4% of the

[34] Contrary to the belief that trade secrets only include intellectual properties, such as methods and formulas, the legal community considers customer information (i.e., customer name, address, telephone number, purchasing behavior) as one type of trade secret (for example: https://www.natlawreview.com/article/customer-lists-trade-secrets; and https://www.tradesecretslaw.com/2017/04/articles/trade-secrets/are-my-customer-lists-a-trade-secret/). Several states also explicitly include customer lists as trade secrets in state laws (e.g., Conn Gen. Stat. § 35-51(d); O.C.G.A. § 10-1-761(4); Or. Rev. Stat. § 646.461(4); 12 Pa. Cons. Stat. Ann. § 5302.)

[35] Note a firm can experience several separate breaches in the same year, so the total number of *Breaches* reported is greater than the number of *Breaches* firm-years.

firm-year observations (3,333 unique firms) disclose trade secret existence in Form 10-K. Among firms that discuss cyber defense and cyber vulnerability in their 10-Ks, there are 7.3 mentions of cyber defense related words/phrases on average, and 4.36 mentions of cyber vulnerability related words/phrases on average (un-tabulated). Statistics for other variables are generally consistent with prior research. We observe less than 1% of the observations (269 unique firms) have an IT related internal control material weakness, which is anecdotally consistent with the low number of IT specific weakness numbers reported in Haislip et al., (2016). Further, about 8.2% of the observations are from the retail industry, while about 23.4% of the observations are from the financial industry.[36]

In Panel B of Table 2, we split the sample by *Trade Secret*, and test the difference in means and medians between the two split samples. It is notable that means and medians of all variables are significantly different between firms that do versus do not disclose the existence of trade secrets. The univariate tests reveal some interesting patterns. For example, trade secret firms on average have more Breach incidents, worse cyber defense mechanisms, and greater cyber vulnerability, as captured by the two 10-K-based disclosure measures. Further, trade secret firms have higher R&D and advertising expenditures, and more IT specific internal control deficiencies.

In Panel C of Table 2, we also split the sample by *Breach*, and test the difference in means and medians between the two split samples. Interestingly, we also observe a higher probability of disclosing *Trade Secret* at year t, if there is a *Breach* incident in year t+1.

[36] We retain financial firms in the full sample for our main analysis because the financial industry is important in the cybersecurity setting. The Department of Homeland Security classifies the financial services sector as one of the 16 critical U.S. infrastructure sectors whose assets, systems, and networks must be protected against threats from cyberspace. If hacked, the impact of security breaches on firms in the financial services sector is severe and has a long-lasting effect (for example, the 2017 Equifax breach). Nonetheless, we confirm that our results hold for both non-financial firms and financial firms.

Furthermore, *Breach* firms on average have more cyber defense mechanisms and higher cyber vulnerability. Table 3 shows the Pearson and Spearman correlations among the test and control variables. We observe a positive correlation between *Breach* and *Trade Secret*, which is consistent with the main hypothesis (Pearson Correlation = 0.011; p-value = 0.021). Interestingly, both *Cyber Defense* and *Cyber Vulnerability* are positively associated with *Breach.* However, *Cyber Defense* has a negative correlation with *Trade Secret*, while *Cyber Vulnerability* has a positive association with *Trade Secret*. The correlations suggest that multivariate models are necessary to obtain reliable results.

4.2 Multivariate Analysis

To test the main hypothesis that the disclosure of the existence of trade secrets is positively associated with subsequent Breaches, we conduct a multivariate logistic regression using model (1). The results are presented in Table 4. All p-values are based on two-sided tests, and are calculated with standard errors clustered by firm (Petersen, 2009). The first column shows results for the full sample, while the second (third) column provides results for a reduced sample excluding financial firms (including only financial firms). Overall, our logistic models perform reasonably well. Taking the full sample model as an example, the area under the ROC curve is 0.90, and the Pesudo-R squared is 0.285. Collectively, these statistics suggest our multivariate models are reasonable. Next, we study the relations between *Breach* and *Trade Secret*, and with the control variables.

We find that the disclosure of the existence of trade secrets has a significant positive association with subsequent Breaches (Estimate = 0.324, p-value = 0.009). The effect is also economically significant. On average, mentioning a trade secret increases the probability of a

Breach incident by over 30%.[37] We also find that better *Cyber Defense* is associated with reduced likelihood of future Breaches (Estimate = -0.127, p-value = 0.050), while greater *Cyber Vulnerability* is associated with higher likelihood of future Breaches (Estimate = 0.515, p-value < 0.01). The estimated coefficients of control variables are generally consistent with prior research. For example, the probability of having a future Breach incident is positively correlated with *SIZE* (Estimate = 0.612, p-value < 0.01), but has an insignificant correlation with *LOSS*, similar to the findings in Higgs et al. (2016). Consistent with Kwon et al. (2013), *BTM* has a negative but insignificant association with *Breach.* Profitability (*ROA*) is negatively associated with *Breach*, but has an insignificant coefficient. Further, firms with greater expenditures on *R&D* (Estimate = 3.102, p-value = 0.003) and *Advertising* (Estimate = 10.709, p-value < 0.01) have greater probability of encountering a future Breach incident. We also find that audit fee is positively associated with the probability of future Breaches (Estimate = 0.434, p-value < 0.01).

Firms in the retail industry have a greater probability of Breach incidents (Estimate = 0.994, p-value < 0.01).[38] Given that the financial industry has a different regulatory and reporting environment, we estimate the logistic model using a reduced sample that excludes all financial industry firms (Column 2), as well as a sub-sample for financial firms only (Column 3). We find for both non-financial firms and financial firms, the disclosure of *Trade Secret* is positively associated with subsequent *Breach* (Column 2: Estimate = 0.273; p-value = 0.041; Column 3: Estimate = 0.391; p-value = 0.068). The signs and significance levels of control variables remain similar, except for *Cyber Defense*, *R&D*, and *BIG4.* We find *Cyber Defense* and *R&D* are not significantly related to *Breach* for financial firms, while having a *BIG 4* auditor is negatively

[37] The average marginal effect of *Trade Secret* is 0.004. Since the ex-ante probability of *Breach* is only 0.013, this marginal effect translates to a 30.77% (0.004/0.013) increase in the probability of *Breach.*

[38] Our results hold if we drop firms in the retail industry, the financial industry, or both.

associated with future Breach incidents (Estimate = -0.626, p-value = 0.013) for non-financial firms.

4.3 Cross Sectional Variation

In this section, we explore some cross-sectional variations in the association between trade secrets and Breach incidents. Specifically, we interact *Trade Secret* with the company age (*AGE*), the total number of employees (*EMP*), and the Herfindahl-Hirschman index (*HHI*). We choose these three variables because they broadly represent the life-stage, the labor intensity, and the competitive environment a company faces. Prior literature shows these three factors have profound implications for firms (e.g., Lev and Schwartz, 1971; Mueller, 1972; Ali et al., 2014). We acknowledge that these three dimensions are unlikely to capture all potential moderating factors, and that we do not have strong priors regarding expected signs of association. However, studying these three factors provides a starting point for understanding cross-sectional differences in the association between trade secrets and Breach incidents.

Results in Table 5 reveal some interesting patterns. First, we find the coefficient for *Trade Secret* × *Age* is significantly negative (Estimate = -0.324, p-value = 0.037), indicating the effects of mentioning trade secrets on future Breach incidents are stronger among younger firms. This is reasonable as firms in early stages (i.e., start-ups, growth firms) often possess novel technology or products that attract hackers. Second, the coefficient for *Trade Secret* × *EMP* is significantly negative (Estimate = -0.011, p-value < 0.01), which suggests the effects of mentioning trade secrets on future Breach incidents are stronger for firms with less numerous employees. A possible explanation for this phenomenon is that younger firms tend to have fewer employees, and that both firm age and employee number have incremental explanatory power. An alternative explanation is that firms using less intensive labor inputs rely more on other inputs to generate value for customers. Trade secrets could be one such alternative input that attracts hackers. Finally, we observe that the coefficient for *Trade Secret* × *HHI* is significantly negative (Estimate = -1.462, p-value = 0.019). Since the Herfindahl-Hirschman index is a reverse measure of the equality of supplier market shares, this suggests that in less concentrated industries, i.e., ones with more equal market shares, the effects of disclosing trade secret existence on future Breach incidents are stronger. A possible explanation is that gaining knowledge of a firm's trade secrets is more valuable in markets that are more competitive (less concentrated). An alternative explanation, that is consistent with our preceding discussion, is that younger firms tend to operate in less concentrated markets, and to have trade secrets that attract attackers. In summary, an explanation that is consistent with all the Table 5 results is as follows. Younger firms tend to have fewer employees and to operate in less concentrated markets. They also tend to have and disclose the existence of trade secrets.[39] If firm age, number of employees, and market concentration each has incremental explanatory power for Breaches, the association of Breaches with trade secret disclosure could vary in cross section with these three factors. We emphasize that our results and this explanation are tentative and exploratory in nature.

5. Additional Analysis

5.1. The Initiation and Cessation of Trade Secret Disclosure

Prior literature suggests some firms' textual disclosures in SEC filings are sticky and do not change significantly year-over-year (Brown and Tucker, 2011; Cohen et al., 2018). Since the disclosure-based *Trade Secret* measure is constructed from SEC 10-K filings, it is a natural concern that a firm disclosing *Trade Secrets* could mention 'trade secret' or 'trade secrecy' in their 10-K disclosures for several consecutive years. To address this issue, we explore the impact

[39] Furthermore, younger firms might have less developed cybersecurity systems.

on breaches of those years in which firms initiate or cease disclosure of the existence of trade secrets. Specifically, we set a new variable, *Trade Secret Initiation*, equal to one if *Trade Secret* changes from 0 to 1 between two fiscal years. Then we replace the *Trade Secret* in model (1) with this newly created indicator variable. The result is presented in Table 6 Column (1). Overall, we observe a significant positive coefficient estimate for *Trade Secret Initiation* in predicting future data breaches (Estimate = 0.476, p-value = 0.004).

We also explore whether ceasing to mention trade secret existence in SEC Form 10-K is associated with reduced likelihood of having breaches. To do so, we set a new variable, *Trade Secret Cessation*, equal to one if *Trade Secret* changes from 1 to 0 between two fiscal years. The result is presented in Table 6 Column (2). Although we observe an attenuated likelihood of having future breaches, the coefficient estimate is not significant at conventional levels (Estimate = -0.131, p-value = 0.595). This is plausible, because the knowledge that trade secrets exist may not quickly decay after firms stop mentioning them in filings. In Table 6 Column 3, we enter both *Trade Secret Initiation* and *Trade Secret Cessation* in the same regression, and find consistent inferences (*Trade Secret Initiation* Estimate = 0.471, p-value = 0.006; *Trade Secret Cessation* Estimate = -0.093, p-value = 0.712). This analysis suggests our main result is not driven by firms that repeatedly mention *Trade Secret* in SEC filings.

5.2. Addressing Endogeneity of Voluntary Disclosure

5.2.1 Additional Control Variables

Our main test variable (*Trade Secret*) and two control variables (i.e., *Cyber Defense, Cyber Vulnerability*) are constructed by textual analysis based on firms' SEC 10-K filings. During our study period the SEC did not require companies to disclose the existence of trade secrets (*Trade Secret*) or information underlying our two cybersecurity-related control variables

(*Cyber Defense*, *Cyber Vulnerability*).[40] We utilize several methods to address this concern. First, we rely on prior studies and look for additional determinants of cyber-related disclosure. Gordon et al. (2010) find firms' voluntary disclosure on information security depends on a set of factors, including firm size, operating performance, industry, long-term assets, stock turnover, stock return volatility, analyst following, and institutional ownership. While we have controlled for the first three in our main analysis, in Table 7 we explore whether including the remaining controls alters our inference.[41]

In Column 1 of Table 7, we control for the proportion of long-term assets to total assets (1-ACT/AT) in our main model, and find it is not significantly correlated with *Breach* (Estimate = 0.287; p-value = 0.536). Our variable of interest, *Trade Secret*, still has a significant positive coefficient estimate (p-value = 0.019). In the next two columns, we consider additional capital market pressure to disclose. Specifically, in Columns 2 and 3 of Table 7, we include stock turnover (*TURNOVER)* and stock return volatility (*Return Volatility*), respectively. We find our main inference still holds and both stock turnover and stock return volatility have a significant positive effect on future *Breach* occurrence (*TURNOVER* Estimate = 0.123; p-value = <0.01; *Return Volatility*: Estimate = 3.138; p-value = 0.002). This suggests that firms with higher stock liquidity and return volatility are more likely to have *Breaches*. Next, we consider some effects of external monitoring. Specifically, in Columns 4 and 5 of Table 7 we add the number of analysts following (*# of Analysts*) and the percentage of institutional ownership (*Institution Own*). Consistent with our main analysis, we still observe a significant and positive relation between *Trade Secret* and *Breach*. Further, the results suggest firms with more analysts

[40] The SEC issued a Concept Release (No. 33-10064) in 2016 to seek public comments on a proposal to expand the disclosure requirement under item 101(c)(1)(iv) of Regulation S-K to other types of intellectual property (e.g., trade secret). For public comments on this, see https://www.sec.gov/comments/s7-06-16/s70616-352.pdf

[41] We do not include all these controls in the main analysis, as some of them require additional data coverage (i.e., IBES, CRSP, Thomson Reuters 13f), which would adversely affect sample size.

following (Estimate = 0.354; p-value < 0.01) and higher institutional ownership (Estimate = 0.501; p-value = 0.065) are more likely to experience future *Breaches*.

Because firms' 10-K disclosures are subject to SEC review at least every three years, we use SEC staff comment letters regarding inadequate disclosure of cyber issues as additional controls to provide expert post-filing opinions on the adequacy of firms' disclosures on cyber-defense and cyber-vulnerability. To do so, we use the same set of keywords that we used to identify our *Cyber Defense* and *Cyber Vulnerability* variables to search for SEC comment letters on these topics. Next, we create two indicators *SEC Letters on Cyber Defense* and *SEC Letters on Cyber Vulnerability* for the existence of such letters and include them in model (1) as additional control variables. The result is presented in Column 6 of Table 7. We find our main inference holds after controlling for SEC comment letters (Estimate = 0.330; p-value = 0.007). Interestingly, the coefficient estimate on *SEC Letters on Cyber Defense* suggests that the inadequate disclosure of cyber defense attracts *Breaches* (Estimate = 0.511; p-value = 0.001). On the other hand, the insignificant coefficient on *SEC Letters on Cyber Vulnerability* suggests that inadequate disclosure on vulnerability does not attract *Breaches*. Collectively, we find that our main inference is robust to controlling for either firm's own disclosure determinants or the existence of SEC comment letters that demand more cyber-related disclosure.

We include all seven variables in the last results column. We find the coefficient on the *Trade Secret* variable in Column 7 of Table 7 (with additional control variables) remains qualitatively similar to the coefficient on the *Trade Secret* variable in Column 1 of Table 4 (without additional control variables). This indicates that the relation between *Breach* and *Trade Secret* is not affected by the additional cyber disclosure determinants. On the other hand, only *# of Analysts* among the seven additional control variables is significant in Column 7, potentially due to intercorrelations or to reduction in sample size caused by the additional variables.[42] Taken together, the implications of firms' disclosures in Form 10-K of the existence of corporate proprietary information on future data breaches are distinct from and incremental to those of cyber security disclosures. [43]

5.2.2 Instrumental Variable Approach

In this section, we examine whether the relation between *Trade Secret* and *Breach* is robust to unobservable omitted characteristics. Because we are unable to control for unobservable omitted variables we rely on the instrumental variable approach. Following Glaeser (2018) and Png (2017 a b) we use the *Trade Secret Index*, which measures the staggered adoption by most states in the U.S. of trade secret protections as an instrument for the mentioning of trade secrets. Historically, trade secret protections were governed by state common law. However, such common law contains many uncertainties and ambiguities. In 1979, the National Conference of Commissioners on Uniform State Laws published the Uniform Trade Secrets Act (UTSA) as a model for state laws to improve trade secret protections in three key aspects: substantive law, procedures, and remedies (Png, 2017b). Because each state had a different level of trade secret protection prior to UTSA and differed in the extent to which they adopted UTSA codifications,[44] Png (2017 a b) construct an index specifying six items that characterize the three major aspects of trade secrets protection under the UTSA. Specifically, Png (2017 b, 169) considers the following six items: "1) *Substantive law*: (a) whether a trade secret must be in continuous business use; (b) whether the owner must take reasonable efforts to

[42] *IT Deficiency* is dropped from Column 7 of Table 7 because *IT Deficiency* only have value zero in this subsample.
[43] We acknowledge that adding additional controls only address one type of endogeneity: correlated omitted variables (Glaeser and Guay 2017). As our focus is on the relation between trade secret disclosure and cybersecurity breaches, analyzing companies' choices in cyber-related disclosure is beyond the scope of this study. We refer interested readers to Gordon et al. (2006) as well as the SEC's new interpretive guidance on cybersecurity disclosures, and encourage future research in this area.
[44] The UTSA has been adopted by 48 states, and by the District of Columbia, as of 2016. For details, see http://www.beckreedriden.com/trade-secrets-laws-and-the-utsa-a-50-state-and-federal-law-survey-chart/.

protect the secret; (c) whether mere acquisition of the secret is misappropriation; 2) *Civil procedure*: the limitation on the time for the owner to take legal action for misappropriation; and 3) *Remedies*: (a) whether an injunction is limited to eliminating the advantage from misappropriation; (b) the multiple of actual damages available in punitive damages." If a state's trade secret law contains "n" components of the six items, the trade secret index for that state will be "n/6." The *Trade Secret Index* values range between 0 and 1.

This is a relevant and valid instrument. The *Trade Secret Index* is a known determinant of firms' disclosure of trade secret usage and is associated with firms' reliance on trade secrets (Glaeser, 2018; Png, 2017 a b). For example, Png (2017 b) documents a positive association between R&D expenditures (that can generate trade secrets) and the promulgation of UTSA. Glaeser (2018) finds that the promulgation of UTSA is negatively associated with firms' patenting activities, consistent with increased reliance on trade secrets. We confirm that in our sample *Trade Secret Index* and *Trade Secret* have a Pearson Correlation of 0.032 (p-value < 0.01). On the other hand, the *Trade Secret Index* instrument satisfies the exclusion restriction for an instrumental variable. It is unlikely to be correlated with the incidence of cyber-attacks directly, because the index does not capture cross-sectional differences in attackers' motives and opportunities to commit *Breaches*. We also confirm that in our sample the Pearson Correlation between the *Trade Secret Index* and *Breach* is 0.005 (p-value = 0.343).

The first two columns of Table 8 show the first stage results. *Trade Secret Index* has a positive and significant association with *Trade Secret* (Estimate = 0.047, p-value = 0.008). Therefore, the only plausible influence of trade secret law on breach incidents is through *Trade Secret.* The next two columns of Table 8 present the second stage results. We observe the coefficient for *Predicted(Trade Secret)* is positive and significant (Estimate = 2.151, p-value <

0.01), which is consistent with the findings in Table 4. This suggests that the positive associations between disclosure of trade secret existence and future *Breach* incidents are not driven by unobservable factors.

It is also possible that managers make strategic disclosure choices related to cyber defense capability and vulnerability, which means that our control variables, *Cyber Defense* and *Cyber Vulnerability*, can also be endogenous. However, based on prior literature (Verrecchia 1983; Dye 1985; Skinner 1994), we argue that managers have little incentive to manipulate the disclosure of internal information about cyber defense mechanisms and cyber vulnerability. Firms are likely to subject themselves to future litigation costs if they overstate (understate) their cyber defense mechanisms (cyber vulnerability) in their 10-K disclosure. Similarly, the stock market cannot properly evaluate firms' future uncertainty regarding their information security if they understate (overstate) their cyber defense mechanisms (cyber vulnerability) in their 10-K disclosure (Wang et al. 2013). In fact, Wang et al. (2013) find that the disclosure of cyber defense mechanism (cyber vulnerability) is negatively (positively) associated with the realization of information security risk factors.

5.2.3 Industry and Firm Fixed Effects

In this section, we re-estimate the main model specifications using the full sample with industry and firm fixed effects. Although we attempt to include a comprehensive list of control variables in the main analyses, we might have omitted some time-invariant industry or firm characteristics that influence our results. Fixed effect regressions can address any time-invariant industry/firm-specific characteristics that affect *Trade Secret*, *Breach*, or both. Table 9 presents the results. Column 1 reproduces the main results from Table 4 Column 1 for comparison purposes. We include the industry fixed effects (2-digit SIC) in the second column and firm fixed effects in the third column. We omit *RETAIL* and *FINANCIAL* indicators in Column 2 since their effects are captured by fixed effects estimators. The variable of interest, *Trade Secret*, still has a positive and significant coefficient in both columns. The number of observations in the second column is slightly lower due to lack of variations in some industry and year pairs. Most of the control variables retain their sign and significance, except for *Cyber Defense*. We also observe the pseudo R-squared increases by over 15 percent (4.5 percentage points) after including industry fixed effects. In the third column, instead of running a logistic regression with firm fixed effects, we run an OLS regression with the same set of control variables.[45] Consequently, the adjusted R-squared is not directly comparable to the pseudo R-squared reported in Column 2.[46] Collectively, this exercise suggests our results hold even after we control for any industry- or firm- specific and time-invariant characteristics.

5.2.4 Propensity Score Matching Techniques

Table 2 Panel B reveals that trade secret firms are different from non-trade secret firms on a number of dimensions. To address the concern that trade secret mention is a firm choice that induces a selection bias, we employ a propensity score matching method to account for observable factors that determine the decision to mention trade secrets in 10-K disclosures. Specifically, in the first stage, we explain trade secret mention by regressing *Trade Secret* on all independent variables available in model (1). Un-tabulated results show that this model performs reasonably well (Area under ROC = 0.84). Next, we use a nearest neighborhood match with a maximum distance of 0.0008.[47] We also require that the matched control firm-years are from the

[45] Running a logistic regression with firm fixed effects may cause a complete separation issue if the outcome is a rare event (see Katz (2001) and Greene (2004) for a detailed discussion). Nonetheless, in an untabulated robustness check, we run a conditional logit model with fixed effects (Stata: xtlogit) and confirm our main inference holds.

[46] We find the adjusted R-squared increases from 0.048 to 0.179 (an increase over three-fold), if we run OLS regressions in both Column 2 and Column 3.

[47] The inference is unchanged for a range of maximum distance between 0.0001 and 0.001.

same fiscal year and industry (2-digit SIC code) as the firms that mention trade secrets. This results in a sample of 2,540 matched pairs (firm-year observations). Un-tabulated results show that the mean differences between the treatment group and the matched control group are balanced for most covariates, except for *Cyber Vulnerability, LOSS, R&D,* and *Advertising.* Following the recommendation in Shipman et al. (2017), we employ these variables as controls in second stage models to account for any remaining unmatched differences.

Table 10 shows the second stage model results using the propensity score matching method. We still observe a positive and significant coefficient for *Trade Secret* (Estimate = 0.707, p=value = 0.021).[48] The propensity score matching method results in test firms that are as similar as possible to control firms, with the exception that the test firms mention trade secrets in their 10-Ks and the control firms do not. This method enhances confidence that the positive association between *Trade Secret* and *Breach* is due to revelation of trade secret existence, and is not due to other, uncontrolled differences between the two samples.

6. Conclusion

We study the association between firms' disclosures in Form 10-K of the existence of corporate proprietary information, namely trade secrets, and cyber theft (which we refer to as cybersecurity breaches, or "Breaches" for brevity). Prior academic research explaining occurrence of Breaches is scarce, and no prior study has focused specifically on Breaches that likely target trade secrets. We provide such evidence, and our use of Form 10-K contents related to trade secrets is a first step toward determining whether corporations actually attract Breach activity through their public disclosures. We find that firms mentioning the existence of trade

[48] We drop *IT Deficiency* from the model because it does not have a coefficient estimate due to its lack of variation. Results are unchanged if we include this variable in the regression.

secrets have a significantly higher probability of subsequently being breached relative to firms that do not disclose such phrases. Our results are more pronounced among younger firms, firms with fewer employees, and firms operating in less competitive industries. By conducting a battery of additional tests, we attempt to go beyond merely establishing correlations to provide evidence whether such proprietary information can actually attract cyberattacks. Specifically, our results are robust to additional control variables that are shown to determine the cyber disclosures, an instrumental variable approach, firm fixed effects, and a propensity score matching technique.

The results of our study should be evaluated in light of limitations. First, the SEC Commissioner Robert J. Jackson Jr. states that "the lack of a representative data set for cybersecurity incidents poses a number of challenges to firms and policy makers". [49] Due to these data limitations, we cannot verify that corporate trade secrets were stolen in every breach. However, we rely on publicly available information to anecdotally classify all breach incidents in 2015 into trade secret specific breaches versus others. In this small sample, we are able to show that our results are concentrated in the trade secret specific breaches. Second, although we try to build a comprehensive dataset of cyber-attacks as possible, we acknowledge that many firms might choose not to disclose their cyber-attack incidences due to ambiguous SEC guidance (Wang et al. 2013; Stein 2018). Therefore, we might under-estimate the relation between trade secret disclosures and cybersecurity breaches. The SEC updated its guidance for public companies to disclose cybersecurity risks and incidents in February 2018, so future studies can build a more refined cybersecurity breach dataset. Third, our results may be affected by omitted variable bias or selection bias even though we have attempted to address these issues through

[49] https://www.sec.gov/news/public-statement/statement-jackson-2018-02-21.

econometric means. Future studies might employ more advanced econometric techniques and/or better identification strategies to study the determinants of cybersecurity breaches.

References

Ali, A., S. Klasa, and E. Yeung. 2014. Industry concentration and corporate disclosure policy. *Journal of Accounting and Economics* 58 (2): 240-264.

Almeling, D. S., Snyder, D. W., Sapoznikow, M., McCollum, W. E., and J. Weader. 2010. A statistical analysis of trade secret litigation in federal courts. *Gonzaga Law Review* 45, 291- 334.

American Institute of Certified Public Accountants (AICPA). 2016. Exposure Draft: Proposed description criteria for management's description of an entity's cybersecurity risk program (September 15, 2016).

Arundel, A. 2001. The relative effectiveness of patents and secrecy for appropriation. *Research Policy* 30 (4): 611-624.

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

Center for Audit Quality. 2017. The CPA's role in addressing cybersecurity risk (May 2017). Available at: http://thecaq.org/cpas-role-addressing-cybersecurity-risk.

Cohen, L., C. J. Malloy, and Q. H. Nguyen. 2018. Lazy Prices. Available at SSRN: https://ssrn.com/abstract=1658471.

Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not): National Bureau of Economic Research. Available at: http://www.nber.org/papers/w7552.

Dye, R. A. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research* 23 (1): 123-145.

Glaeser, S. 2018. The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics (Forthcoming).*

Glaeser, S, and W. R. Guay. 2017. Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017). *Journal of Accounting and Economics* 64 (2-3): 305-312.

Gordon, L. A., and M. P. Loeb. 2002. The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)* 5 (4): 438-457.

Gordon, L. A., M. P. Loeb, and W. Lucyshyn. 2003. Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy* 22 (6): 461-485.

Gordon, L. A., M. P. Loeb, W. Lucyshyn, and T. Sohail. 2006. The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities. *Journal of Accounting and Public Policy* 25 (5): 503-530.

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

Gordon, L. A., M. P. Loeb, W. Lucyshyn, and L. Zhou. 2015a. Externalities and the magnitude of cyber security underinvestment by private sector firms: a modification of the Gordon-Loeb model. *Journal of Information Security* 6 (1): 24-30.

Gordon, L. A., M. P. Loeb, W. Lucyshyn, and L. Zhou. 2015b. The impact of information sharing on cybersecurity underinvestment: a real options perspective. *Journal of Accounting and Public Policy* 34 (5):509-519.

Gordon, L. A., M. P. Loeb, W. Lucyshyn, and L. Zhou. 2015c. Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity* 1 (1): 3-17.

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1): 3-73.

Greene, W. 2004. The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects. *Econometrics Journal* 7: 98-119.

Haislip, J. Z., A. Masli, V. J. Richardson, and J. M. Sanchez. 2016. Repairing organizational legitimacy following information technology (IT) material weaknesses: executive turnover, IT expertise, and IT system upgrades. *Journal of Information Systems* 30 (1): 41-70.

Hall, B., C. Helmers, M. Rogers, and V. Sena. 2014. The choice between formal and informal intellectual property: a review. *Journal of Economic Literature* 52 (2): 375-423.

Higgs, J. L., R. E. Pinsker, T. J. Smith, and G. R. Young. 2016. The relationship between board-level technology committees and reported security breaches. *Journal of Information Systems* 30 (3): 79-98.

Holt, T., and Kilger, M. 2012. Know your enemy: The social dynamics of hacking. Working paper available at http://www.honeynet.org.

Jensen, P. H., and E. Webster. 2009. Knowledge management: does capture impede creation? *Industrial and Corporate Change* 18 (4): 701-727.

Katz, Ethan. 2001. Bias in Conditional and Unconditional Fixed Effects Logit Estimation. *Political Analysis* 9:379-384.

Klasa, S., Ortiz-Molina, H., Serfling, M., and S. Srinivasan. 2018. Protection of trade secrets and capital structure decisions. *Journal of Financial Economics* 128(2): 266-286.

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

Lawrence, A., Minutti-Meza, M., and D. Vyas. 2018. Is Operational Control Risk Informative of Financial Reporting Deficiencies? *Auditing: A Journal of Practice & Theory* 37 (1): 139-165.

Lev, B., and A. Schwartz. 1971. On the use of the economic concept of human capital in financial statements. *The Accounting Review* 46 (1): 103-112.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2-3): 221-247.

Malawer, S. 2014. Confronting Chinese economic cyber espionage with WTO litigation. New York Law Journal (December 23, 2014). Available at: https://www.law.com/newyorklawjournal/almID/1202712784205.

Mueller, D. C. 1972. A life cycle theory of the firm. *The Journal of Industrial Economics* 20 (3): 199-219.

National Conference of Commissioners on Uniform State Laws. 1985. Uniform trade secrets act with 1985 amendments. Available at: http://www.uniformlaws.org/shared/docs/trade%20secrets/utsa_final_85.pdf.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies* 22 (1): 435-480.

Png, I. P. 2017a. Secrecy and patents: Theory and evidence from the Uniform Trade Secrets Act. *Strategy Science* 2 (3): 176-193.

Png, I. P. 2017b. Law and innovation: evidence from state trade secrets laws. *Review of Economics and Statistics* 99 (1): 167-179.

Public Law 104-294, The Economic Espionage Act of 1996. Available at: https://www.gpo.gov/fdsys/pkg/STATUTE-110/pdf/STATUTE-110-Pg3488.pdf.

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

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

Spanos, G., and L. Angelis. 2016. The impact of information security events to the stock market: A systematic literature review. *Computers & Security* 58: 216-229.

Stein, K. M. 2018. Statement on Commission Statement and Guidance on Public Company Cybersecurity Disclosures. Available at https://www.sec.gov/news/public-statement/statement-stein-2018-02-21.

Verrecchia, R. E. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5: 179-194.

Villasenor, J. 2015. Corporate Cybersecurity Realism: Managing Trade Secrets in a World Where Breaches Occur. *AIPLA QJ*, *43*, 329.

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

VARIABLE APPENDIX

Variable	Definition
Breach	An indicator variable taking the value of one if there is at least one breach incident during year t+1
Trade Secret	An indicator variable for the disclosure of the existence of trade secrets at year t
Trade Secret Index	Instrument variable used for *Trade Secret*. Constructed following Png (2017 b,169), it measures the state-level strength of the legal protection of trade secrets, including six dimensions on substantive law, civil procedure, and remedies.
Cyber Defense	A disclosure based measure on cyber defense mechanisms. It is the natural logarithm of the number of words related to firms' cyber defense mechanisms in SEC Form 10-K filings. The keywords are derived from Gordon et al. (2006) and Lawrence et al. (2018). Refer to footnote 30 for details.
Cyber Vulnerability	A disclosure based measure on cyber defense weaknesses. It is the natural logarithm of the number of words related to firms' cyber weaknesses in SEC Form 10-K filings. The keywords are derived from Gordon et al. (2006) and Lawrence et al. (2018). Refer to footnote 31 for details.
SIZE	The natural logarithm of total assets (AT)
BTM	The ratio of book value of equity (CEQ) to its market value (PRCC_F×CSHO) at fiscal year-end
AGE	The natural logarithm of the number of years of a firm since its first observation in Compustat
ROA	Compustat net income (NI) divided by total assets (AT)
LOSS	An indicator variable for loss year (NI<0)
R&D	Compustat research and development expense (XRD) divided by total assets (AT)
Advertising	Compustat advertising expense (XAD) divided by total assets (AT)
BIG 4	An indicator for four largest international auditors (Deloitte, EY, KPMG, and PwC)
Log(Audit Fee)	The natural logarithm of total audit fees reported in Audit Analytics (audit_fees)
IT Deficiency	An indicator variable taking the value of one if there is any internal control deficiency in information technology, software, security and access issues following Haislip et al. (2016)
10-K Length	The natural logarithm of the total number of words in the SEC Form 10-K filings
Log(FOG-INDEX)	The natural logarithm of Fog readability measure for the SEC Form 10-K filings, refer to Li (2008) for details
RETAIL	An indicator variable for the retail industry (SIC: 5000-5999)
FINANCIAL	An indicator variable for the financial industry (SIC: 6000-6999)

VARIABLE APPENDIX (Continued)

Variable	Definition
EMP	The total number of employees reported in Compustat (EMP)
HHI	The Herfindahl-Hirschman index constructed using market share based on sales and 4-digit SIC.
Trade Secret Initiation	An indicator for the first year when firm start to disclose the existence of trade secrets
Trade Secret Cessation	An indicator for the first year when firm stop to disclose the existence of trade secrets
Long-term Assets	1 minus the ratio of current assets (ACT) to total assets (AT)
Stock Turnover	Trading volume accumulated over a 12-month period ending three-month after the fiscal year-end scaled by shares outstanding, calculated from CRSP monthly file
Stock Return	Standard deviation of the firm's market adjusted returns over a 12-month period ending three-month after the fiscal year-end, calculated

Volatility	from CRSP monthly file. Stock return is adjusted using equal weighted return
# of Analysts	Natural log of the total number of analysts following, using I/B/E/S summary file
Institutional Own	The percentage of shares owned by institutional investors, reported in Thomson Reuters 13f data
SEC Letter on Cyber Defense	An indicator for the existence of SEC comment letters on the lack of adequate disclosure on cyber defense. See footnote 30 for terms used in the text searches of SEC letters.
SEC Letter on Cyber Vulnerability	An indicator for the existence of SEC comment letters on the lack of adequate disclosure on cyber vulnerability. See footnote 31 for terms used in the text searches of SEC letters.

Figure 1a

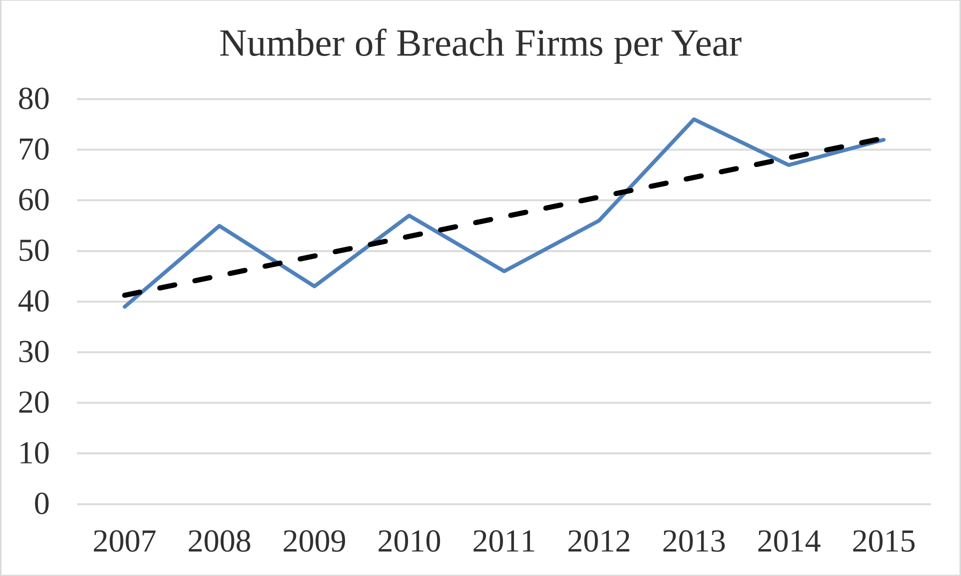

Figure 1b

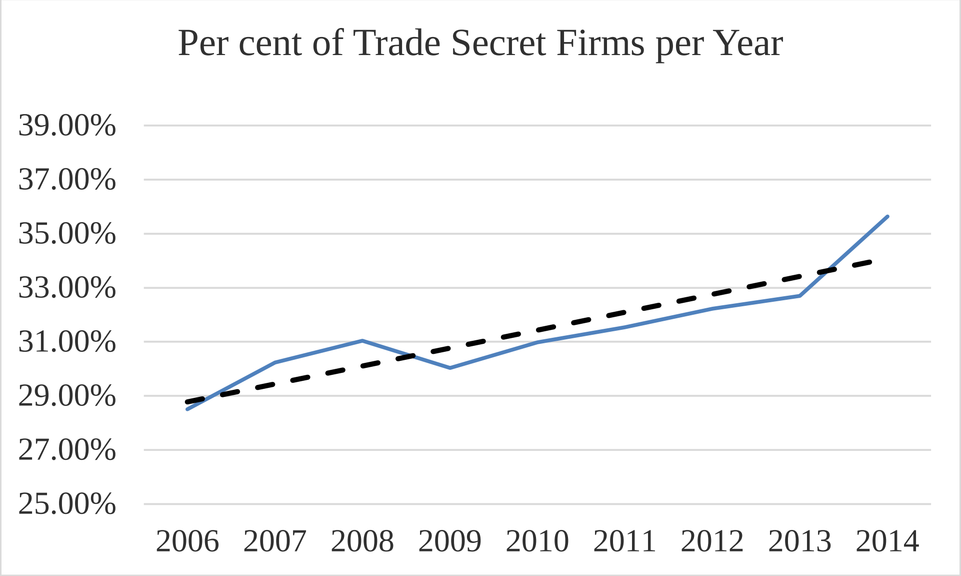

Figure 2

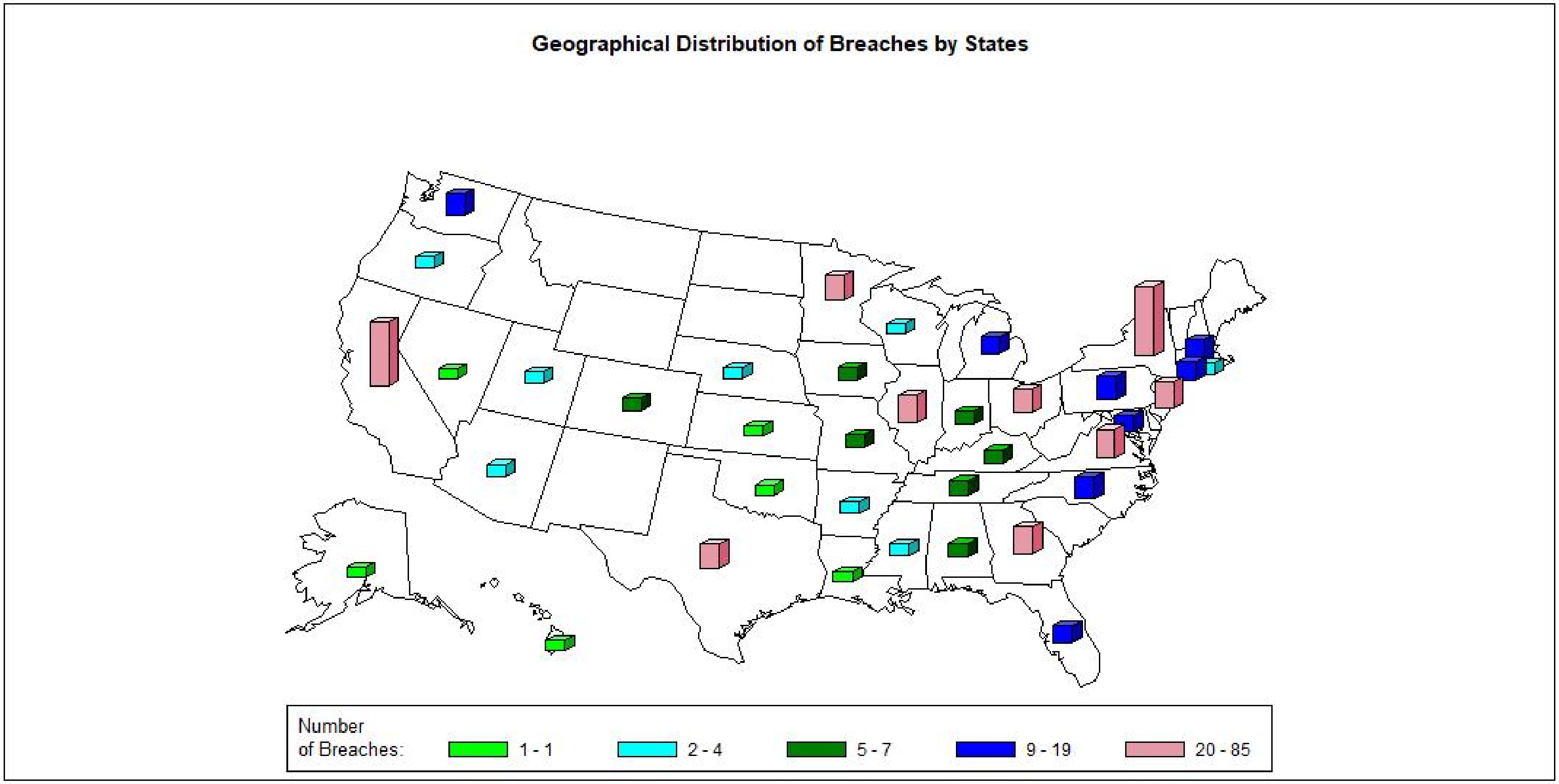
Geographical Distribution of Breaches by States
Number
of Breaches:
1 - 1
2 - 4
5 - 7
9 - 19
20 - 85

Table 1

Panel A: Industry Distribution of Breaches and Trade Secret Firms

SIC Code	Industry Group	Obs	%	Breach	%	Trade Secret	%
0000-0999	Agriculture, Forestry, And Fishing	148	0.37%	1	0.20%	29	0.23%
1000-1999	Mining and Construction	2,613	6.53%	8	1.57%	186	1.48%
2000-3999	Manufacturing	15,273	38.19%	101	19.76%	7,082	56.47%
4000-4999	Transportation, Communications, Electric, Gas, And Sanitary Services	2,931	7.33%	48	9.39%	474	3.78%
5000-5999	Wholesale Trade and Retail Trade	3,292	8.23%	77	15.07%	642	5.12%
6000-6999	Finance, Insurance, And Real Estate	9,343	23.26%	170	33.27%	851	6.78%
7000-8999	Services	6,102	15.26%	105	20.54%	3,248	25.90%
9000-9999	Public Administration	290	0.73%	1	0.20%	30	0.24%
	Total	39,992	100%	511	100.00%	12,542	100.00%

The industry group division is based on SIC manual available at https://www.osha.gov/pls/imis/sic_manual.html.

Table 1 (Continued)

Panel B: Year Distribution of Reported *Breaches* and Unique Firms

Year	Number of *Breaches* Reported	Number of Unique *Breach* Firms
2007	47	39
2008	65	55
2009	49	43
2010	74	57
2011	50	46
2012	57	56
2013	78	76
2014	79	67
2015	92	72
Total	591	511

Table 2 Descriptive statistics

This table contains descriptive statistics. All continuous variables are winsorized at the 1st and 99th percentiles. See Variable Appendix for variable definitions.

Panel A: Summary statistics

Variables	**N**	**Mean**	**Std. Dev.**	**P25**	**P50**	**P75**
Data Breach	39,992	0.013	0.112	0.000	0.000	0.000
Trade Secret	39,992	0.314	0.464	0.000	0.000	1.000
Cyber Defense	39,992	0.788	0.988	0.000	0.693	1.386
Cyber Vulnerability	39,992	0.644	0.815	0.000	0.000	1.386
SIZE	39,992	6.157	2.343	4.573	6.293	7.765
BTM	39,992	0.730	0.710	0.296	0.547	0.912
AGE	39,992	2.724	0.784	2.197	2.773	3.258
ROA	39,992	-0.051	0.341	-0.033	0.019	0.071
LOSS	39,992	0.332	0.471	0.000	0.000	1.000
R&D	39,992	0.046	0.108	0.000	0.000	0.035
Advertising	39,992	0.009	0.025	0.000	0.000	0.003
BIG 4	39,992	0.618	0.486	0.000	1.000	1.000
Log(Audit Fee)	39,992	13.384	1.444	12.384	13.480	14.342
IT Deficiency	39,992	0.009	0.092	0.000	0.000	0.000
10-K Length	39,992	10.961	0.551	10.637	10.939	11.261
Log(FOG-INDEX)	39,992	3.129	0.069	3.091	3.127	3.161
RETAIL	39,992	0.082	0.275	0.000	0.000	0.000
FINANCIAL	39,992	0.234	0.423	0.000	0.000	0.000

Table 2 Descriptive statistics (Continued)

Panel B: Mean and median descriptive statistics split on *Trade Secret*

	Trade Secret = 1		**Trade Secret = 0**		**Difference in Means**		**Difference in Median**	
	N=12542		**N=27450**					
Variables	**Mean**	**Median**	**Mean**	**Median**	**t-stat**	**p-value**	**z-stat**	**p-value**
Data Breach	0.015	0.000	0.012	0.000	2.279	0.002	2.278	0.002
Cyber Defense	0.556	0.000	0.894	0.693	-32.171	0.000	-32.354	0.000
Cyber Vulnerability	0.727	0.000	0.606	0.000	13.789	0.000	13.411	0.000
SIZE	5.648	5.565	6.390	6.619	-29.696	0.000	-33.907	0.000
BTM	0.564	0.403	0.805	0.616	-31.968	0.000	-41.978	0.000
AGE	2.576	2.639	2.792	2.833	-25.831	0.000	-28.159	0.000
ROA	-0.114	0.016	-0.023	0.019	-25.161	0.000	-13.734	0.000
LOSS	0.435	0.000	0.285	0.000	29.972	0.000	29.642	0.000
R&D	0.100	0.049	0.022	0.000	71.566	0.000	100.305	0.000
Advertising	0.011	0.000	0.008	0.000	9.886	0.000	14.266	0.000
BIG 4	0.676	1.000	0.591	1.000	16.333	0.000	16.279	0.000
Log(Audit Fee)	13.533	13.628	13.316	13.400	13.952	0.000	13.845	0.000
IT Deficiency	0.010	0.000	0.008	0.000	1.648	0.099	1.648	0.099
10-K Length	11.018	10.970	10.934	10.921	14.193	0.000	13.429	0.000
Log(FOG-INDEX)	3.131	3.127	3.129	3.127	2.068	0.039	2.791	0.005
RETAIL	0.051	0.000	0.097	0.000	-15.354	0.000	-15.310	0.000
FINANCIAL	0.068	0.000	0.309	0.000	-54.917	0.000	-52.957	0.000

Table 2 Descriptive statistics (Continued)

Panel C: Mean and median descriptive statistics split on *Breach*

	Breach = 1		**Breach = 0**		**Difference in Means**		**Difference in Median**	
	N=511		**N=39481**					
Variables	**Mean**	**Median**	**Mean**	**Median**	**t-stat**	**p-value**	**z-stat**	**p-value**
Trade Secret	0.360	0.000	0.313	0.000	2.279	0.023	2.278	0.023
Cyber Defense	1.703	1.386	0.776	0.693	21.187	0.000	15.778	0.000
Cyber Vulnerability	1.497	1.609	0.633	0.000	23.998	0.000	20.613	0.000
SIZE	9.565	9.796	6.113	6.261	33.550	0.000	29.016	0.000
BTM	0.633	0.464	0.731	0.547	-3.086	0.002	-3.509	0.001
AGE	3.216	3.332	2.718	2.773	14.290	0.000	14.253	0.000
ROA	0.044	0.039	-0.053	0.018	6.348	0.000	7.489	0.000
LOSS	0.125	0.000	0.335	0.000	-9.997	0.000	-9.985	0.000
R&D	0.020	0.000	0.046	0.000	-5.540	0.000	-4.788	0.000
Advertising	0.015	0.001	0.009	0.000	6.236	0.000	11.939	0.000
BIG 4	0.949	1.000	0.614	1.000	15.557	0.000	15.510	0.000
Log(Audit Fee)	15.392	15.550	13.358	13.459	32.038	0.000	28.617	0.000
IT Deficiency	0.004	0.000	0.009	0.000	-1.155	0.248	-1.155	0.248
10-K Length	11.486	11.395	10.954	10.934	21.810	0.000	18.700	0.000
Log(FOG-INDEX)	3.166	3.144	3.129	3.127	12.022	0.000	10.225	0.000
RETAIL	0.151	0.000	0.081	0.000	5.662	0.000	5.659	0.000
FINANCIAL	0.333	0.000	0.232	0.000	5.328	0.000	5.326	0.000

Table 3 Pearson (upper) and Spearman (lower) Correlation

		[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**	**[15]**	**[16]**	**[17]**
Data Breach	[1]	-	**0.011**$^{\#}$	0.005	**0.105***	**0.119***	**0.165***	**-0.015***	**0.071***	**0.032***	**-0.050***	**-0.028***	**0.031***	**0.078***	**0.158***	-0.006	**0.108***	**0.060***
Trade Secret	[2]	**0.011**$^{\#}$	-	**0.032***	**-0.159***	**0.069***	**-0.147***	**-0.158***	**-0.128***	**-0.125***	**0.148***	**0.337***	**0.049***	**0.081***	**0.070***	**0.008**	**0.071***	**0.010**$^{\#}$
Trade Secret Index	[3]	0.004	**0.032***	-	**-0.047***	**0.024***	**0.019***	**-0.040***	**0.042***	0.006	-0.008	**0.011**$^{\#}$	**0.018***	**0.038***	**0.020***	0.006	**-0.012**$^{\#}$	**-0.032***
Cyber Defense	[4]	**0.079***	**-0.162***	**-0.027***	-	**0.342***	**0.506***	**0.101***	**0.077***	**0.090***	**-0.168***	**-0.174***	**-0.116***	**0.164***	**0.313***	0.001	**0.439***	**0.227***
Cyber Vulnerability	[5]	**0.103***	**0.067***	**0.020***	**0.290***	-	**0.341***	**-0.032***	**0.033***	**0.100***	**-0.125***	**-0.116***	**0.086***	**0.176***	**0.272***	**0.009**	**0.300***	**0.182***
SIZE	[6]	**0.145***	**-0.170***	**0.020***	**0.499***	**0.334***	-	**0.019***	**0.295***	**0.375***	**-0.401***	**-0.341***	**-0.053***	**0.541***	**0.788***	0.005	**0.585***	**0.359***
BTM	[7]	**-0.018***	**-0.210***	**-0.028***	**0.147***	-0.006	**0.099***	-	**-0.013**$^{\#}$	**0.078***	**0.126***	**-0.189***	**-0.073***	**-0.176***	**-0.131***	-0.004	**-0.012**$^{\#}$	-0.005
AGE	[8]	**0.071***	**-0.141***	**0.047***	**0.056***	**0.024***	**0.281***	**0.047***	-	**0.239***	**-0.205***	**-0.127***	0.001	**0.188***	**0.317***	-0.002	**0.038***	**0.055***
ROA	[9]	**0.037***	**-0.069***	**-0.016***	**0.033***	**0.059***	**0.288***	**-0.157***	**0.216***	-	**-0.546***	**-0.459***	**0.025***	**0.172***	**0.267***	-0.006	**0.062***	**0.062***
LOSS	[10]	**-0.050***	**0.148***	-0.006	**-0.176***	**-0.128***	**-0.399***	**0.013***	**-0.205***	**-0.816***	-	**0.345***	0.008	**-0.172***	**-0.242***	**0.035***	**-0.072***	**-0.076***
R&D	[11]	**-0.024***	**0.502***	**0.033***	**-0.267***	**-0.124***	**-0.300***	**-0.302***	**-0.034***	**-0.147***	**0.249***	-	**-0.035***	**-0.012**$^{\#}$	**-0.128***	**-0.010**$^{\#}$	**-0.045***	**-0.078***
Advertising	[12]	**0.060***	**0.071***	**0.067***	**-0.096***	**0.096***	-0.005	**-0.039***	**0.017***	**0.060***	**-0.038***	**0.035***	-	**0.039***	**0.037***	0.003	**-0.045***	**-0.062***
BIG 4	[13]	**0.078***	**0.081***	**0.042***	**0.158***	**0.183***	**0.539***	**-0.134***	**0.171***	**0.204***	**-0.172***	**0.045***	**-0.027***	-	**0.685***	0.003	**0.379***	**0.218***
Log(Audit Fee)	[14]	**0.143***	**0.069***	**0.028***	**0.287***	**0.271***	**0.772***	**-0.087***	**0.287***	**0.264***	**-0.240***	**0.008**	**0.021***	**0.706***	-	**0.054***	**0.552***	**0.308***
IT Deficiency	[15]	-0.006	**0.008**	0.006	0.002	**0.011**$^{\#}$	0.002	0.001	-0.002	**-0.028***	**0.035***	0.008	**0.010**	0.003	**0.056***	-	**0.031***	**0.009**
10-K Length	[16]	**0.094***	**0.067***	**-0.010**	**0.435***	**0.313***	**0.588***	**0.024***	0.008	-0.008	**-0.064***	**-0.064***	**-0.054***	**0.393***	**0.556***	**0.036***	-	**0.533***
Log(FOG-INDEX)	[17]	**0.051***	**0.014***	**-0.033***	**0.276***	**0.242***	**0.385***	**0.050***	**0.039***	-0.001	**-0.078***	**-0.109***	**-0.085***	**0.228***	**0.319***	**0.014***	**0.506***	-

Pearson correlation is presented in upper triangle, while the Spearman correlation is presented in lower triangle. All significant correlations (10% level) are in bold, and higher significance levels are denoted with # for 5% level or * for 1% level. See Variable Appendix for variable definitions.

Table 4 Regression Results: Trade Secrets and Breach Incidents

	BREACH		*BREACH (Exclude Financials)*		*BREACH (Financials only)*	
VARIABLES	Estimate	P-value	Estimate	P-value	Estimate	P-value
CONSTANT	-15.189***	(0.000)	-12.625***	(0.000)	-19.281***	(0.000)
Trade Secret	**0.324*** **	**(0.009)**	**0.273** **	**(0.041)**	**0.391* **	**(0.068)**
Cyber Defense	-0.127*	(0.050)	-0.261***	(0.001)	0.056	(0.633)
Cyber Vulnerability	0.515***	(0.000)	0.458***	(0.000)	0.377***	(0.006)
SIZE	0.612***	(0.000)	0.691***	(0.000)	0.691***	(0.000)
BTM	-0.065	(0.641)	-0.244	(0.330)	-0.032	(0.845)
AGE	-0.028	(0.761)	-0.078	(0.417)	0.212	(0.126)
ROA	0.108	(0.890)	0.071	(0.923)	2.630	(0.294)
LOSS	-0.150	(0.413)	-0.089	(0.691)	-0.165	(0.608)
R&D	3.102***	(0.003)	2.460**	(0.026)	7.734	(0.626)
Advertising	10.709***	(0.000)	9.767***	(0.000)	23.942***	(0.000)
BIG 4	-0.196	(0.422)	-0.626**	(0.013)	0.606	(0.305)
Log(Audit Fee)	0.434***	(0.000)	0.274***	(0.007)	0.450***	(0.000)
IT Deficiency	-0.660	(0.343)	-0.821	(0.405)	0.193	(0.836)
10-K Length	0.160	(0.265)	0.063	(0.719)	0.201	(0.420)
Log(FOG-INDEX)	-0.898	(0.275)	-0.465	(0.645)	-0.782	(0.589)
RETAIL	0.994***	(0.000)	0.831***	(0.000)	-	-
FINANCIAL	0.257	(0.178)	-	-	-	-
Year FE	YES		YES		YES	
Observations	39,992		30,649		9,343	
Area under ROC	0.90		0.89		0.94	
Likelihood Ratio	-1954.92		-1421.60		-476.41	
Pseudo R-squared	0.285		0.241		0.410	

This table contains regression results for the relation between firm's reliance on trade secret and its propensity to be attacked by hackers in year t+1. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

Table 5 Cross-Sectional Variations in the Relation between Trade Secrets and Breach Incidents

	BREACH		*BREACH*		*BREACH*	
VARIABLES	Estimate	P-value	Estimate	P-value	Estimate	P-value
CONSTANT	-15.658***	(0.000)	-12.002***	(0.000)	-14.951***	(0.000)
Trade Secret	**1.343*** **	**(0.008)**	**0.823*** **	**(0.000)**	**0.650*** **	**(0.000)**
Trade Secret×AGE	**-0.324** **	**(0.037)**				
Trade Secret×EMP			**-0.011*** **	**(0.000)**		
Trade Secret×HHI					**-1.462** **	**(0.019)**
Cyber Defense	-0.131**	(0.044)	-0.094	(0.160)	-0.126*	(0.052)
Cyber Vulnerability	0.499***	(0.000)	0.475***	(0.000)	0.513***	(0.000)
SIZE	0.619***	(0.000)	0.536***	(0.000)	0.625***	(0.000)
BTM	-0.051	(0.717)	-0.006	(0.966)	-0.064	(0.647)
AGE	0.100	(0.399)	-0.135	(0.156)	-0.022	(0.810)
ROA	0.197	(0.800)	0.221	(0.753)	0.093	(0.905)
LOSS	-0.131	(0.473)	-0.105	(0.562)	-0.156	(0.398)
R&D	2.938***	(0.005)	2.054*	(0.054)	3.036***	(0.004)
Advertising	10.662***	(0.000)	9.725***	(0.000)	10.883***	(0.000)
BIG 4	-0.234	(0.340)	0.076	(0.769)	-0.190	(0.435)
Log(Audit Fee)	0.430***	(0.000)	0.237***	(0.007)	0.409***	(0.000)
IT Deficiency	-0.620	(0.370)	-0.463	(0.493)	-0.632	(0.354)
10-K Length	0.164	(0.251)	0.222	(0.118)	0.158	(0.271)
Log(FOG-INDEX)	-0.885	(0.279)	-1.147	(0.168)	-0.934	(0.258)
RETAIL	1.008***	(0.000)	0.533***	(0.008)	0.975***	(0.000)
FINANCIAL	0.285	(0.129)	0.424**	(0.024)	0.266	(0.164)
EMP			0.015***	(0.000)		
HHI					0.597	(0.166)
Year FE	YES		YES		YES	
Observations	39,992		39,992		39,992	
Area under ROC	0.90		0.90		0.90	
Likelihood Ratio	-1951.87		-1910.77		-1950.94	
Pseudo R-squared	0.287		0.302		0.287	

This table contains cross-sectional variations for the relation between firm's reliance on trade secret and its propensity to be attacked by hackers in year t+1. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

Table 6 Trade Secret Disclosure Initiation and Cessation and Effects on Breaches

	BREACH		*BREACH*		*BREACH*	
VARIABLES	Estimate	P-value	Estimate	P-value	Estimate	P-value
CONSTANT	-15.121***	(0.000)	-15.468***	(0.000)	-15.123***	(0.000)
Trade Secret Initiation	**0.476*****	**(0.004)**			**0.471*****	**(0.006)**
Trade Secret Cessation			**-0.131**	**(0.595)**	**-0.093**	**(0.712)**
Cyber Defense	-0.138**	(0.031)	-0.141**	(0.028)	-0.138**	(0.031)
Cyber Vulnerability	0.537***	(0.000)	0.537***	(0.000)	0.538***	(0.000)
SIZE	0.606***	(0.000)	0.607***	(0.000)	0.606***	(0.000)
BTM	-0.068	(0.621)	-0.074	(0.599)	-0.068	(0.622)
AGE	-0.057	(0.529)	-0.070	(0.442)	-0.057	(0.530)

This table contains regression results for the relation between firms initiate or cease disclosure of the existence of trade secrets and its propensity to be attacked by hackers in year t+1. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

ROA	0.227	(0.764)	0.159	(0.842)	0.228	(0.764)
LOSS	-0.148	(0.416)	-0.149	(0.415)	-0.147	(0.421)
R&D	3.699***	(0.000)	3.682***	(0.000)	3.703***	(0.000)
Advertising	10.820***	(0.000)	10.929***	(0.000)	10.823***	(0.000)
BIG 4	-0.205	(0.402)	-0.202	(0.408)	-0.205	(0.402)
Log(Audit Fee)	0.456***	(0.000)	0.459***	(0.000)	0.456***	(0.000)
IT Deficiency	-0.711	(0.308)	-0.653	(0.345)	-0.712	(0.308)
10-K Length	0.152	(0.294)	0.185	(0.195)	0.151	(0.296)
Log(FOG-INDEX)	-0.931	(0.255)	-0.935	(0.252)	-0.930	(0.256)
RETAIL	0.965***	(0.000)	0.968***	(0.000)	0.966***	(0.000)
FINANCIAL	0.220	(0.252)	0.211	(0.272)	0.220	(0.252)
Year FE	YES		YES		YES	
Observations	39,992		39,992		39,992	
Area under ROC	0.90		0.90		0.90	
Likelihood Ratio	-1955.46		-1959.10		-1955.38	
Pseudo R-squared	0.285		0.284		0.285	

Table 7 Controlling for Endogenous Choices in Voluntary Disclosure

Dependent Variable				BREACH			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CONSTANT	-13.648***	-15.243***	-15.529***	-15.360***	-14.520***	-15.145***	-15.031***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Trade Secret	**0.320****	**0.319****	**0.306****	**0.347*****	**0.303****	**0.330*****	**0.313****
	(0.019)	**(0.011)**	**(0.016)**	**(0.004)**	**(0.015)**	**(0.007)**	**(0.026)**
Long-Term Assets	0.287						0.544
	(0.536)						(0.275)
Stock Turnover		0.123***					0.067
		(0.000)					(0.106)
Stock Return Volatility			3.138***				0.737
			(0.002)				(0.668)
# of Analysts				0.354***			0.378***
				(0.000)			(0.000)
Institutional Own					0.501*		0.011
					(0.065)		(0.969)
SEC Letter on Cyber Defense						0.511***	0.347
						(0.001)	(0.233)
SEC Letter on Cyber Vulnerability						0.140	-0.160
						(0.774)	(0.841)
Cyber Defense	-0.215***	-0.117*	-0.126*	-0.121*	-0.132**	-0.146**	-0.187**
	(0.007)	(0.070)	(0.052)	(0.063)	(0.039)	(0.025)	(0.018)
Cyber Vulnerability	0.463***	0.496***	0.514***	0.480***	0.522***	0.494***	0.455***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SIZE	0.638***	0.628***	0.649***	0.520***	0.637***	0.613***	0.536***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BTM	-0.298	-0.227	-0.233	0.050	-0.120	-0.081	-0.528*
	(0.256)	(0.144)	(0.126)	(0.710)	(0.403)	(0.564)	(0.059)
AGE	-0.155	0.000	-0.010	-0.020	-0.051	-0.035	-0.135
	(0.121)	(1.000)	(0.914)	(0.828)	(0.587)	(0.700)	(0.202)
ROA	0.161	-0.267	0.026	-0.172	0.030	0.164	-0.438
	(0.839)	(0.774)	(0.978)	(0.818)	(0.975)	(0.836)	(0.643)
LOSS	-0.125	-0.311	-0.293	-0.063	-0.163	-0.142	-0.069
	(0.584)	(0.116)	(0.142)	(0.729)	(0.405)	(0.437)	(0.773)

Table 7 Controlling for Endogenous Choices in Voluntary Disclosure (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
R&D	2.619**	2.457**	2.736**	1.985*	3.124***	3.079***	1.049
	(0.028)	(0.030)	(0.013)	(0.080)	(0.004)	(0.003)	(0.434)
Advertising	9.434***	9.522***	10.044***	10.276***	10.047***	10.690***	8.165***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BIG 4	-0.562**	-0.294	-0.127	-0.347	-0.230	-0.201	-0.676***
	(0.026)	(0.249)	(0.616)	(0.160)	(0.361)	(0.410)	(0.010)
Log(Audit Fee)	0.366***	0.420***	0.404***	0.455***	0.411***	0.424***	0.404***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
IT Deficiency	-0.822	-1.223	-1.272	-0.411	-1.248	-0.625	-
	(0.408)	(0.222)	(0.201)	(0.553)	(0.217)	(0.364)	-
10-K Length	0.059	0.124	0.153	0.144	0.146	0.154	0.031
	(0.731)	(0.396)	(0.301)	(0.313)	(0.322)	(0.279)	(0.858)
Log(FOG-INDEX)	-0.448	-0.772	-0.795	-0.823	-1.083	-0.830	-0.117
	(0.659)	(0.356)	(0.343)	(0.315)	(0.192)	(0.313)	(0.911)
RETAIL	0.948***	0.967***	0.992***	0.915***	0.974***	0.986***	0.898***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FINANCIAL	0.761**	0.315*	0.286	0.336*	0.274	0.210	0.778**
	(0.019)	(0.100)	(0.129)	(0.081)	(0.144)	(0.275)	(0.012)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,214	33,763	33,763	39,992	35,226	39,992	25,703
Area under ROC	0.89	0.89	0.89	0.90	0.89	0.90	0.88
Likelihood Ratio	-1456.22	-1842.05	-1849.17	-1936.90	-1903.75	-1950.31	-1351.94
Pseudo R-squared	0.252	0.278	0.275	0.292	0.275	0.287	0.250

This table contains regression results for the relation between firm’s reliance on trade secret and its propensity to be attacked by hackers in year t+1 after we add additional control variables that are key determinants of cyber related disclosures. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

Table 8 Regression Results: Instrumental Variable Approach

	IV First Stage		*IV Second Stage*	
VARIABLES	Estimate	P-value	Estimate	P-value
CONSTANT	-1.558***	(0.000)	-0.656	(0.795)
Trade Secret Index	0.047***	(0.008)		
Predicted(Trade Secret)			**2.151*** **	**(0.000)**
Cyber Defense	-0.045***	(0.000)	0.069*	(0.058)
Cyber Vulnerability	0.086***	(0.000)	-0.055	(0.570)
SIZE	-0.056***	(0.000)	0.250***	(0.000)
BTM	-0.023***	(0.000)	0.029	(0.446)
AGE	-0.073***	(0.000)	0.151***	(0.000)
ROA	0.098***	(0.000)	-0.272*	(0.064)
LOSS	0.016*	(0.060)	-0.076	(0.114)
R&D	1.041***	(0.000)	-1.480*	(0.059)
Advertising	0.118	(0.552)	2.201*	(0.097)
BIG 4	0.005	(0.696)	-0.117	(0.106)
Log(Audit Fee)	0.081***	(0.000)	-0.055	(0.551)
IT Deficiency	-0.051*	(0.053)	-0.085	(0.663)
10-K Length	0.095***	(0.000)	-0.159**	(0.023)
Log(FOG-INDEX)	0.084	(0.141)	-0.357	(0.144)
RETAIL	-0.132***	(0.000)	0.510***	(0.000)
FINANCIAL	-0.160***	(0.000)	0.381***	(0.000)
Year FE	YES		YES	
Observations	39,992		39,992	
Adj R Square	0.235		-	
Area under ROC	-		0.90	
Likelihood Ratio	-		-22642.10	

This table contains regression results for the relation between firm's reliance on trade secret and its propensity to be attacked by hackers in year t+1 by using the instrumental variable approach. The instrumental variable is the *Trade Secret Index*, measured at the state where the firm headquartered at. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

Table 9 Regression Results: Industry and Firm Fixed Effects

	BREACH		*BREACH*		*BREACH*	
VARIABLES	Estimate	P-value	Estimate	P-value	Estimate	P-value
CONSTANT	-15.189***	(0.000)	-15.115***	(0.000)	-0.043	(0.449)
Trade Secret	**0.324*** **	**(0.009)**	**0.282** **	**(0.039)**	**0.007** **	**(0.037)**
Cyber Defense	-0.127*	(0.050)	-0.035	(0.616)	0.001	(0.295)
Cyber Vulnerability	0.515***	(0.000)	0.244***	(0.001)	0.000	(0.807)
SIZE	0.612***	(0.000)	0.652***	(0.000)	0.001	(0.329)
BTM	-0.065	(0.641)	0.025	(0.839)	0.001	(0.223)
AGE	-0.028	(0.761)	0.121	(0.249)	-0.004	(0.426)
ROA	0.108	(0.890)	0.082	(0.932)	0.001	(0.473)
LOSS	-0.150	(0.413)	-0.274	(0.160)	-0.001	(0.441)
R&D	3.102***	(0.003)	4.109***	(0.000)	-0.001	(0.881)
Advertising	10.709***	(0.000)	9.829***	(0.000)	0.047	(0.335)
BIG 4	-0.196	(0.422)	-0.178	(0.472)	-0.005**	(0.018)
Log(Audit Fee)	0.434***	(0.000)	0.410***	(0.000)	0.002*	(0.095)
IT Deficiency	-0.660	(0.343)	-0.900	(0.221)	-0.002	(0.510)
10-K Length	0.160	(0.265)	0.308**	(0.037)	-0.002	(0.256)
Log(FOG-INDEX)	-0.898	(0.275)	-1.604*	(0.072)	0.016	(0.375)
RETAIL	0.994***	(0.000)	-	-	-0.005	(0.334)
FINANCIAL	0.257	(0.178)	-	-	-0.003	(0.619)
Year FE	YES		YES		YES	
Industry FE	-		YES		-	
Firm FE	-		-		YES	
Observations	39,992		38,290		39,992	
Area under the ROC	0.90		0.91		-	
Likelihood Ratio	-1954.92		-1817.12		-	
Pesudo/Adj. R-squared	0.285		0.330		0.179	

This table contains regression results for the relation between firm's reliance on trade secret and its propensity to be attacked by hackers in year t+1 after we add additional fixed effect estimators. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.

Table 10 Regression Results: Propensity Score Matched Samples

	BREACH	
VARIABLES	Estimate	P-value
CONSTANT	-4.452	(0.563)
Trade Secret	**0.707****	**(0.021)**
Cyber Defense	-0.233	(0.175)
Cyber Vulnerability	0.716***	(0.000)
SIZE	0.561***	(0.002)
BTM	-0.034	(0.921)
AGE	0.223	(0.382)
ROA	-2.071**	(0.019)
LOSS	-1.403**	(0.024)
R&D	1.547	(0.576)
Advertising	10.714**	(0.019)
BIG 4	0.589	(0.405)
Log(Audit Fee)	0.323	(0.176)
10-K Length	0.038	(0.932)
Log(FOG-INDEX)	-3.811	(0.131)
RETAIL	1.199**	(0.046)
FINANCIAL	0.492	(0.270)
Year FE	YES	
Observations	5,080	
Area under ROC	0.93	
Likelihood Ratio	-237.84	
Pseudo R-squared	0.325	

This table contains regression results for the relation between firm’s reliance on trade secret and its propensity to be attacked by hackers in year t+1 by using the propensity score matching techniques. All continuous variables are winsorized at the 1st and 99th percentiles. *p*-values using robust, firm-clustered standard errors are in parentheses. *, ** and *** indicate two-tailed significance at the 10%, 5%, and 1% levels, respectively. See Variable Appendix for variable definitions.