Exploring Cyber Risk Contagion - A Boundless Threat

Abstract: As the complexity and severity of cyber risk continue to expand, businesses face greater systemic risk from cyber threats. Cyber risk is likely contagious given the increasing interconnectedness of the web-based global economy. Using a unique dataset, the *SAS OpRisk Global Data*, our research empirically examine contagion among cyber-attacks based on a flexible modeling framework that we develop to accommodate the interdependence of entities and their risk exposures. This paper provides new insights on cyber risk contagion and can serve as an easily implementable stepping-stone for businesses, insurers, regulators, and academics to analyze cyber risk contagion.

Keywords: Cyber Risk, Contagion, Copula, Cyber Insurance

1. Introduction

The threat of cyber risk is ubiquitous and increasing. FBI notifies over 3,000 U.S. companies each year, from financial institutions to defense contractors to mega retailers, that they were victims of cyber security breaches (Segal 2016). In a public statement on December 14, 2016, Yahoo's Chief Information Security Officer reported a security breach that are "associated with more than one billion user accounts," subsequent to a separate security breach report back in September 2016, in which 500 million accounts were affected. According to PwC's 2014 *Global Economic Crime Survey*, an astounding 19% of U.S. organizations have claimed losses between $50,000 and $1 million, and 7% of U.S. organizations lost over $1 million due to cybercrime in the previous year. The *Center for Strategic and International Studies* has estimated the annual cost of cybercrime and economic espionage to the world economy at more than $445 billion, or almost 1 percent of the global GDP[1].

Cyber risk is likely contagious given the increasing interconnectedness of the web-based global economy. The leading technology research firm Gartner forecasts about 21 billion connected devices worldwide by 2020, up more than 300 percent from today (Gartner 2015). A more connected world and every organization in it are increasingly

[1] https://www.csis.org/news/report-cybercrime-and-espionage-costs-445-billion-annually

vulnerable to computer system failures such as the Y2K problem, and contagious attacks from viruses and hackers such as the recent WannaCry ransomware attack, motivated by malice, monetary or political incentives, inflicting physical, financial, and reputational damages.

Not surprisingly, cyber risk and its management have significant implications for the global (re)insurance industry (c.f. Gordon et al. 2003; Bodin et al. 2008). According to the 2016 *RIMS Cyber Survey*, risk transfer by insurance is the primary risk management method used by organizations worldwide: nearly 70% of the respondents chose to transfer their cyber risks and over 80% of those purchased a standalone cyber insurance policy, a near 30% increase from 2015. As greater understanding of cyber risk is being accumulated, cyber risk contagion is a chief concern for (re)insurers because the interconnectedness of cyber risk exposures can be a major impediment to the insurability and market formation of cyber security risks. If contagion is indeed a concern, its impact on cyber insurance product design, actuarial pricing and risk management is paramount for (re)insurance companies. Given the high demand for cyber insurance products, it is essential for the (re)insurance industry to fully understand the nature of the risk exposure.

1.1 Challenges of Cyber Risk and Its Management

In the early history of cyber security, typical hackers' focus was mostly on fame and recognition. That focus has quickly shifted to achieving financial gains or political goals against targeted organizations. Nowadays, professional and elite hackers often maintain global operations and many belong to well-organized and profit-motivated groups hired and paid to perform illegal hacking (Romanosky et al. 2014). To achieve these goals, a large variety of sophisticated methods and tactics have been developed to exploit vulnerability in the targets' cyber systems.

Social engineering and phishing are perhaps the most commonly reported forms of cyber attacks through ostensibly legitimate email attachments, links, software downloads or other operating system vulnerabilities. With one single casual click from the victim, hackers may be able to breach the computer system, evade detection tools, and leverage its vulnerabilities. In this process, malware, spyware and ransomware are often introduced to the target's system. Malware is an all-encompassing term for a variety of malicious

software including Trojans, viruses and worms with intent that typically steals data or destroys something on the computer. Spyware specializes on tracking keystrokes to get passwords or electronically spying in order to gain unauthorized access to confidential information, sometimes staying undetected for weeks or longer. Most recently, we have seen a worldwide ransomware attack in 2017 in which case the WannaCry ransomware locked down the targeted Microsoft Windows operating systems and demanded ransom payments in the Bitcoin cryptocurrency to gain access back to the system. Some hostile cyber attacks don't even require any type of malicious software to run on the system. For instance, hackers my launch brute force attack using sophisticated algorithms to simply crack the password-protection of the target system. Another popular form of attack, the Denial of Service attack (DDoS) attack, would focus on overloading the server with high volumes of data in order to disrupt the website or bring down the network.

Due to the heterogeneous, sophisticated, and dynamic nature of cyber risks, it is increasingly challenging for organizations to effectively reduce the risk of being compromised and protect their own cyber integrity (Gordon et al. 2011). The extent of cyber losses can range from nuisance damage to catastrophic damage that seriously erodes data integrity, compromise host and client information, and reduce system availability. More importantly, while cyber risk contagion is a very real threat through emails, mobile apps, website operations, operating systems, electronic payment systems, online databases, cloud servers, and shared online storage, the true extent of the contagion risk has yet to be assessed, let alone being fully managed.

Cyber risk contagion also has important implications for cyber risk management through the use of insurance. Very recently, Kwon (2018) examines how the current insurance market has been dealing with cyber risk and concludes that the industry is still in dire need of basic infrastructure support to continue operations in the physical-cyber world of risk. According to an NAIC report[2], while there is no standard form on which the insurance industry as a whole underwrites cyber coverage, the available cyber liability policies often include coverage on both contagious risk and noncontagious risks, which may have very different implications for insurance pricing. We independently reviewed

[2] http://www.naic.org/cipr_topics/topic_cyber_risk.htm

popular cyber insurance contract provided by insurers such as AIG and Farmers and confirmed such observations. The noncontagious cyber risks are more subject to the law of large numbers and therefore can be priced in a fashion similar to many conventional P&C insurance products. However, the contagious cyber risks remain difficult to quantify due to the lack of actuarial data specifically identifying such risk exposures and proper modeling of their inter-dependence. Therefore, in this paper we attempt to fill in the gap in thoroughly understanding the presence and the extent of cyber risk contagion as well as developing practical modeling tools for assessing and managing cyber risk contagion.

1.2 Literature Review

Despite its growing importance, cyber risk has been a subject of very limited academic research in the insurance literature. Eling, Schnell and Schnell (2016) provide an overview of existing literature on cyber risks. They summarized seven core topics for cyber risk and cyber risk insurance, including definition and categorization, costs and consequences, data availability, risk management strategies, contagion and systemic risk nature, and cyber risk modeling. They, among other studies, noted that one particularly important challenge is the lack of cyber risk modeling frameworks that can capture the various unique aspects of cyber risk exposures and facilitate subsequent empirical, practical and policy discussions.

Most of the existing literature on cyber risk focuses on the economic incentives of self-protection and insurance risk transfer in light of important issues such as moral hazard, adverse selection, and interdependent risks (Hofmann and Ramaj 2011; Öğüt et al. 2011). Böhme and Schwartz (2010) provides a critical literature survey of existing economic models for cyber insurance, discussing the challenges of a viable insurance market for cyber risks and encouraging further theoretical and empirical research to improve the understanding of this important topic. Existing empirical studies are mostly restricted to the use of aggregate survey data and rely upon conceptual frameworks to identify and organize the sources of operational cyber risk (c.f. Mukhopadhyay et al. 2013; Marotta et al. 2015). In particular, Beiner et al. (2014) study the insurability of cyber risks. In addition to a literature review on cyber risk insurability, they suggest that cyber risk losses differ substantially from other operational risk losses and more research is needed to better understand cyber risks in order to develop cyber insurance products. Due to the ever-increasing importance of cybersecurity, the actuarial society has also conducted extensive

research on cyber threats to businesses and the opportunities and challenges for the insurance market, and has jointly produced series of essays[3] and practical guidelines to help actuarial professionals consider this issue (see for example, Solomon 2017; Maxwell 2017; Shang 2017; Dionisi 2017).

A strand of research also discusses the correlated nature of cyber risk exposures. Böhme and Kataria (2006) make use of "honeypots" data from 2003 to 2005 (hosts placed on the internet to attract malicious activities) to provide some evidence on cyber risk contagion. Using SANS data on threats to various components of a firm's information system from 2003 to 2011, Baldwin et al. (2012) develop and estimate a vector equation system of threats to ten important IP services and find strong evidence on cyber risk contagion. Similarly, Wang and Kim (2009) conduct an empirical study of cyber attacks across 62 countries during the period of years 2003-2007 and find strong evidence for the spatial autocorrelation of cyber attacks across countries over time. Shang (2017) point out that cyber risk is more contagious than traditional operational risk and set new challenges to the insurance industry. These studies shed light on the interconnected nature of cyber risk exposures and suggest that any cyber risk modeling approach should capture this feature.

Despite these preliminary efforts, there has been little attempt to examine the patterns and implications of cyber risk contagion that are practically relevant for insurance companies. Zurich Insurance Company and Atlantic Council (2014) in their insightful whitepaper compares cyber risk contagion to the subprime mortgages contagion that prompted the most recent financial crisis in the U.S. Just like the subprime mortgages, the heavily interconnected nature of cyber risk exposures and the common underlying driving forces make it highly susceptible to the domino effects of failures. Although the recent financial crisis has heightened awareness of risk contagion and promoted abundant academic research on systemic risk in financial institutions (c.f. Duan and Wei 2009; Cummins and Weiss 2014), similar research on cyber risk contagion and systemic cyber risks in insurance industry is scarce. Eling and Pankoke (2016) review extant research on

[3] Cybersecurity: Impact on Insurance Business and Operations, 2017, Society of Actuaries (SOA), the Casualty Actuarial Society (CAS), and the Canadian Institute of Actuaries (CIA).

systemic risk in the insurance context. Their study reveals virtually no theoretical or empirical research on cyber risk contagion from either academia or practitioner organizations. This paper adapts methodologies and empirical approaches from the existing literature on data science and financial systemic risks, such as clustering method and the factor copulas method (cf., Billio et al. 2012; Oh and Patton 2017), to develop a framework for modeling and empirically analyzing cyber risk contagion.

1.3. Objective

The aim of this research is to provide the first systematic discussion of cyber risk contagion and close the gap in the risk management and insurance literature. We propose a model framework that can serve as a stepping-stone for businesses, insurers, regulators, and academics in developing their own models. Specifically, we propose and illustrate a two-step method for modeling cyber risk contagion that is flexible to accommodate specific concerns of the end users. As such, this research can serve as a critical starting component for organizations and (re)insurers to gradually build cyber risks into a broader ERM framework. We also benefit from a unique dataset, the *SAS OpRisk Global Data*, to analyze cyber risk and empirically examine contagion among cyber-attacks.

The remainder of this article is organized as follows. Section 2 introduces the *SAS OpRisk Global Data* and describes our data and variables. Section 3 discusses how to refine the dataset for cyber risk contagion analysis. Section 4 builds the empirical method and presents a case study. Model and analysis insights are subsequently discussed. Section 5 concludes the paper and discusses future research.

2. Data Description

2.1. Introduction to *SAS OpRisk Global Data*

The *SAS OpRisk Global Data* is the world's largest and most comprehensive collection of publicly reported operational losses in excess of US$100,000 (Wei et al. 2018). In our analysis, we use the database updated to October 2017. It documents more than 34,360 events across all industries worldwide and provides up to 50 descriptive and financial features. These events span across many industry sectors, ranging from agriculture to manufacturing to financial services. Relevant background information is provided for the events such as the name, country, industry sector of the business and the specific business

lines involved. The starting and end year of the event together with the year of settlement are also documented. Important financial features such as the value of the loss (both original and CPI adjusted) are also available in the data set, including a finer breakdown into items such as legal liability and restitution. Despite these rich characteristics, the *SAS OpRisk Global Data* is still new to the insurance literature.[4] This data set provides the broadest possible statistical sample from which to model cyber risk factors, probabilities, and costs. In addition, we use financial market data from CRSP in the Wharton Research Data Services (WRDS) to supplement the main data in our analysis.

Following the existing literature (c.f. Cebula and Young 2010; Biener and Eling 2014; Eling and Wirfs 2015) and consistent with the operational risk frameworks in Basel II and Solvency II, we define cyber risk as a subgroup of *"operational risk to information and technology assets that have consequences affecting the confidentiality, availability or integrity of information or information systems"* (Cebula and Young 2010), with detailed categories shown in Table 1. By structuring cyber risk as subcategories of operational risk, we can clearly identify cyber risks from the established framework of operational risks in *SAS OpRisk Global Data*.

[Insert Table 1]

Box 1 provides a sample description of a recent prominent cyber-attack on the insurance industry we extracted from the *SAS OpRisk Global Data*.

[4] To our best knowledge, Biener, Eling and Wirfs (2015) is the only published research using this dataset in the insurance literature.

Box 1 Example of a Cyber-attack description extracted from the SAS OpRisk Global Data

> In June 2017, Anthem Inc, a US insurance company, reported that it would pay $115M to settle a class action lawsuit due to a data breach. The company discovered the breach in late January 2015 and reported it to regulators and the public a few days later. Further investigation revealed the hacking lasted from February 18, 2014 to January 30, 2015. An employee at an Anthem subsidiary opened an email containing malicious files that allowed hackers remote access to his computer. The hackers were then able to access over 90 systems, including Anthem's data warehouse. They stole 78.8M consumer records containing names, social security numbers, and other personally identifiable information. Insurance regulators in seven states investigated the breach. They found that Anthem employed reasonable security measures prior to the incident. The company also had a detailed Incident Response Plan that enabled it to respond quickly and effectively once it discovered the breach. As part of the settlement Anthem guaranteed to provide a certain level of funding for information security and would need to implement or maintain data security system changes. Proceeds from the settlement would provide two years of credit monitoring services for the data breach victims; cover out-of-pocket expenses consumers incurred because of the data breach; and provide cash compensation for consumers already enrolled in credit monitoring.

As the example shows, the event description is textual in nature. Therefore, we applied one of the recent developments in the big data analysis, the text mining method implemented by the Python programming language, to identify cyber risk events from keyword strings. This method ensures the accuracy, repeatability and scalability of our identification.

2.2. Initial Identification Method based on Biener, Eling and Wirfs (2015)

Because of very limited prior empirical analysis using this type of cyber risk data, we first applied an initial identification method in the recent literature to extract cyber risk events from the *SAS OpRisk Global Data* for validation. More specifically, we closely follow Biener, Eling and Wirfs (2015) to use a set of broadly defined keyword strings. Three criteria (i.e., critical asset, actor and outcome) in combination are deemed to be relevant for a cyber risk incident. We search in the descriptions of each observation for a combination of keywords, where each combination consisted of one word from each group/criteria (i.e., three-word combinations for the keyword strings) in our sample data. We then checked all identified observations individually for their affiliation with cyber risk and if necessary we excluded those irrelevant incidents from the cyber risk data set. Table 2 describes the keywords for each of the criteria and further group the keywords under "actor" into four categories (categories 1-4) to capture the different natures of the events, including "actions by people," "systems and technical failure," "failed internal processes," and "external events."

[Insert Table 2]

To better understand the quality of the identified cyber risk data set, especially in the context of cyber risk contagion analysis, we provide two examples to illustrate the strength and weakness of the keyword search string method previously described. In the first example, a typical incident of cyber risk involving the Bank of Brazil was identified and included in this data set. The event description in the original data base is reproduced as follows. *"In October 2004, Banco do Brasil, a Brazilian financial institution, reported an estimated loss of $.1M (.29M BRL) due to an online phishing scam that used a Trojan horse virus to attack the company's online ecommerce site. Typically, in a phishing scam internet users enter their user name and password into a fake website that looks identical to the company's site that they are trying to access. This fake website is only online for several days. It records all the user names and passwords that were entered into it so that whoever runs the site can access the real site to transfer money out of the visiting persons' accounts. A Trojan horse virus is a derivative of a typical phishing scam. However, a person does not have to enter their information into a fake site. Instead a program is unknowingly downloaded onto a computer when internet users click on a bogus site and scroll through the page to find out what the site is about. The virus downloaded then monitors the activity of the internet users and records all of their user names and passwords using a key logger. The key logger then sends the information back to the scammer so that the persons' bank accounts can be accessed. Fifty-three people have been arrested by the police who are believed to have been involved in the scam. Eighteen people charged had previously been charged with similar offenses. Banks have lost $30M overall from the Brazilian Trojan horse virus. Other banks that also experienced losses were Banco Bradesco SA, Banco Itau Holding Financeira SA, Caixa Economica Federal, HSBC, and Unibanco."* This event clearly has implications for studying cyber risk contagion possibly across different types of financial institutions and beyond.

Another instance identified in the data set exhibits different characteristics because while there is contagion within the (large) network of the specific financial institution impacted, there is unlikely any contagion across different companies in a larger setting. The description of this event from the original database is reproduced below. *"In March 2005, Barclays Bank, a UK financial institution and the primary subsidiary of Barclays*

PLC, reported that it lost an estimated $.33M (.18M GBP) due to ATMs malfunctioning. On March 27, 2005, from 2am to 5pm, customers were unable to withdraw money from approximately 1500 cash machines after a computer breakdown stopped them from accessing their accounts. Telephone and internet banking was also out of service, but customers were able to make purchases with their cards. The cause of the computer glitch was unknown; however, speculations were that it was caused by the clocks going forward or by a piece of IT hardware. Although the glitch was resolved by 4pm, internet banking remained offline until 5pm. The 1500 Barclays cash points that were out of service represented half of the bank's southern network. The northern network was not affected, as it resided on a different server."

These two examples showcase that, not all identified cyber risk incidents are equally relevant for the purpose of our study. There is clearly a trade-off in this identification method. On the one hand, the keyword search string method has the advantage of comprehensively including all aspects of a possible cyber incident. However, because of its all-encompassing nature, many different types of cyber events are mixed together, resulting in a highly heterogeneous sample. More importantly, upon second screening of the outcomes of the initial keyword search, we find that certain identified event, while appropriately classified as cyber risk, does not seem to have implications for cyber risk contagion. This calls for a more sophisticated process of defining and screening for cyber incidents for the use of our study. For example, some identified incidents result from fraudulent activities of one particular employee or physical damage to certain computers and network equipment. While these qualify as cyber risk events according to the widely accepted broad definition in Biener, Eling and Wirfs (2015), they don't necessarily have any potential for contagion within a company or across many different companies. We address this concern by creating a more elaborated method to identify cyber risk incidents and further refine the data set to study contagion.

It is worth noting that we also considered an alternative identification method by using the event category and subcategory definitions provided in the *SAS OpRisk Global Data* to identify cyber risk incidents. Three event categories (or subcategories) are considered relevant for identifying possible cyber risk incidents: (1) Event = internal fraud and subcategory = unauthorized activity, (2) Event =external fraud and subcategory = systems

security, and (3) Event = business disruption and system failures, in which (2) and (3) better identify contagious cyber risk than (1). This method has its own limitations that the event category classification is pre-determined by the data vendor and hence is subject to any potential bias therein. In addition, only a relatively small set of activities were described and employed when classifying incidents into event categories and subcategories. This might lead to an incomplete set of relevant cyber risk incidents. For this reason, we do not focus on this identification method in the subsequent analysis.[5]

3. Refined Data Set for Cyber Risk Contagion

3.1. Refined Keyword Search

Because cyber risk is a very broadly defined concept, there has not been much work as of yet focusing on narrowing down its definition for the purpose of studying cyber risk contagion. To help advance understanding of contagion, we propose a refined data extraction method to identify a more accurate set of data points pertaining to the contagious cyber risks and base all subsequent analysis upon this refined data set.

Recall that the Biener, Eling and Wirfs (2014) method relies on the set of broadly defined keyword strings based on the three criteria (i.e., critical asset, actor and outcome) to identify the relevant cyber risk events as described in Table 1. While all of the actor categories are related to cyber risk, we consider only a subset of it to be potentially contagious. More specifically, we consider the "systems and technical failure", "external events", and the highlighted "actions by people" to be more prone to contagious cyber risks and hence decompose the actor category into noncontagious and contagious groups. We again applied the textual mining analysis and Python programming language to conduct the new keyword strings extraction. Only the keywords under "Actor Category" in Table 3 are considered by us to be pertaining to contagious cyber risks and hence are included in the refined keyword search.

[Insert Table 3]

3.2. Descriptive Statistics

[5] We conducted parallel descriptive analysis by using the event categories definition. To save space, we did not tabulate these results but note that the results do show some commonality across these two methods.

While the *SAS OpRisk Global Data* covers events from earlier years, we decided to only use extracted events that started in the year 1990. This is because the use of computers and internet only started to become ubiquitous since the 1990's and focusing on more recent data helps us obtain a more relevant and homogenous data set for our analysis. Our descriptive analysis also confirmed that cyber incidents since 1990 count for over 95% of all identified cyber events in our data set.

With the refined cyber risk data set identified by the new keyword search, we have conducted an analysis of firm characteristics to understand the differences between companies that have had a cyber risk incident and those that have not had cyber risk event but have had other operational risk events to be included in the *SAS OpRisk Global Data*. From Table 4, we can see that companies that have had experienced cyber risk incidents have more employees and experienced bigger loss amount.

[Insert Table 4]

Overall, our descriptive analysis of the cyber risk incidents shows that consistent with our intuitive understanding, cyber risk exposures are widespread across industries, business lines, and countries. To conserve space, tables containing the full set of descriptive analysis results are relegated to the Appendix (Tables A1-A4). Based on the keyword search method, we identified a total of 491 cyber risk incidents that are potentially relevant for our analysis, about 30% of which are results of "actions of people," i.e., category 1 as defined in Table 3, 63% of which are results of system and technical failure, i.e., category 3, and 6.3% of which are results of external events, i.e., category 4. We can observe a U-shaped trend in the frequency of cyber risk incidents over the 28 years period (i.e., 1990-2017) that we have collected data on, with a larger number of events being recorded during the years 1997-2010. This pattern can be accounted for by two reasons. First, the increased awareness and efforts on cyber security may have prevented some cyber incidents in the more recent years. Second, sometimes cyber incidents can take years to uncover (e.g., the infamous hack of email accounts at Yahoo!) and it could simply be because that some incidents in the more recent years have not been discovered and included in the data base yet. Indeed, when we examined the length of the documented cyber events, we find that

over 60% of the events last more than one year and nearly 15% of the events last five years and longer.

Our data set also shows that cyber events occur to companies in a wide array of countries although the intensity differs greatly across these countries. A total of 82 countries have at least one recorded cyber event in our data set and sample period. The United States by far has the largest number of recorded events, with other developed economies such as the United Kingdom, Germany, Japan, France, Canada, and Australia closely following behind. It is interesting to note that the largest developing countries are also quite susceptible to cyber risk, with India ranked the 3rd in the number of cyber incidents. While our study will mainly focus on the U.S. market, the nature of many cyber risks suggests the possibility of cross-country contagion. Our model framework can potentially extend to analyzing this type of contagion; detailed analysis of this cross-country contagion might be an interesting research avenue for future studies.

Cyber risk also affects many different industry sectors. Our descriptive analysis shows that 17 industry sectors have experienced at least one cyber incident in our sample period. However, cyber risk exposure seems to be extremely unevenly distributed across different industry sectors. Financial services industry has a much greater exposure to cyber risk than any other industries. In fact, financial services industry accounts for almost 50% of all cyber risk incidents in our data set. Our finding is consistent with Verizon (2017) survey and Kopp et al. (2017), which also show that the finance industry has by far the most incidents with confirmed data losses. Retail, manufacturing, information, professional services and utilities also account for a large portion of the cyber risk events. The larger amount of cyber risk events can naturally lead to a higher chance of cyber risk contagion.

A unique variable available in the data set is whether multiple firms are impacted by the same event. Out of the 491 identified cyber risk events, 129 or 26% of all events have impacted more than one firm. This finding provides initial evidence on the potential contagion effects of cyber risk incidents. While this variable alone is insufficient to characterize fully how cyber risk contagion is formed or what types of companies/incidents are most susceptible to contagion, it can be used to validate our analysis of cyber risk contagion, which will be presented in a subsequent section. The data set also provides

information on whether a single event has resulted in multiple losses. This is relevant when the same incident has resulted in multiple lawsuits, say, in multiple states of the United States. About 16.5% of all incidents result in multiple losses. In our analysis, we aggregate all the losses associated with the same event to correctly identify the impact of one event.

4. Empirical Models

Due to the nature of cyber risk exposures, cyber risk contagion is an important concern for many companies. Different firms are subject to different types of cyber risks at different levels and to different extent, and yet they may be connected from being situated within the same supply chain or value net. In addition, companies may use the same underlying technology platform and/or security software, so cyber attacks on one system may lead to simultaneous attacks on many different companies. Therefore, it is important for the firms and insurers to understand the types of cyber risk exposures before further examining the contagion of the cyber risks.

As complexity grows exponentially with increased number of entities, cyber risk space needs to be broken down into smaller, more manageable clusters in order to examine the formation of the contagion risk and explore its implications for the (re)insurance industry. Additionally, a critical component of our analysis is to evaluate the dependence or co-integration between entities, which also presents a tremendous challenge due to both the extremely sparse data and the high dimensionality of the cyber risk space. To address this challenge, we propose to first use clustering techniques to reduce dimensionality and then use the factor copula model to assess inter-dependence among entities for their cyber risk exposures.

4.1. The Clustering Model for Cyber Risk Data

One natural categorization of the cyber risk exposures is to use the subcategories as defined in Cebula and Young (2010) and Biener, Eling and Wirfs (2014), or the event categories and subcategories as defined in *SAS OpRisk Global Data*. However, upon careful examination of the events, it is clear that these natural choices of categories are too broadly defined and do not take many firm characteristics into consideration. Consequently, they cannot accurately capture the unique features of different types of cyber risks that may have different implications for firm risk managers and/or insurers.

Therefore, we propose to adopt clustering techniques from the machine learning literature to simplify the complex cyber risk graphs and reduce dimensionality. There is a wide set of classification techniques to select from, including cluster analysis, logistic regression, decision tree, support vector machine, and neural network methods, readily available from various statistical packages (such as SAS and R) for robustness and validation. Due to the fast-changing nature of cyber attacks, we need to place a heavier weight on unsupervised learning methods (i.e., methods that do not need to rely on a known "left-hand-side" variable, such as cluster analysis) so that our modeling framework can be easily updated to accommodate new data and patterns by insurance companies in practice on an ongoing basis.

For this reason, we build our classification model using the unsupervised method cluster analysis. Cluster analysis is perhaps one of the most commonly used unsupervised learning methods (c.f. Gan 2013). Under this class of models, data is partitioned according to certain "similarity" and "dissimilarity" measures. The choice of specific measures of "similarity" (or "dissimilarity") is critical. Commonly used clustering methods include Ward's method (which minimizes the within-cluster variance), K-means method (reassignment to the nearest centroid at each iteration), Average linkage (where distance is defined as the average distance between all pairs of members of the two clusters), among others. Many simulations and empirical analyses have shown that there is often not a rule of thumb in choosing the exact type of similarity measure as the performance of different cluster analysis methods depend heavily on the nature of the underlying data to be classified. When one is unsure of the underlying shape and distribution of the clusters in the data set, a nonparametric method is more conservative such as the "density linkage" method that uses nonparametric probability density estimates to find the clusters. For this reason, we chose to use the two-stage density linkage method available in the SAS statistical package to conduct our cluster analysis of the cyber risk data. The two-stage density linkage is a modification of density linkage that ensures all points are assigned to modal clusters before the modal clusters are permitted to join.

While cluster analysis is a popular and easy to implement method and application software (e.g., SAS) have ready-to-use programs to implement it, there are several common considerations in applying cluster analysis. First, cluster analysis can result in an uneven

split of the sample, which may not be desirable in certain applications requiring either a pre-defined number of clusters or a more even sample split. Second, in applications where there are pre-specified classes (such as the classes of fraud, non-fraud in the context of fraud detection), cluster analysis does not suggest a correspondence between these pre-specified classes and the identified clusters. However, these considerations are not causes for concern in the context of cyber risk contagion modeling because we do not have a pre-determined set of clusters. Our main purpose is to group cyber risk incidents by their characteristics and reduce dimensionality, which is precisely what cluster analysis does. In addition to being flexible enough to more easily adaptable to the ever-changing landscape of cyber risks, the unsupervised nature of cluster analysis also sidesteps constraints imposed by the rather limited understanding on properly defining the subcategories within the domain of cyber risk exposure.

4.2. The Within and In-Between Dependence of Cyber Risk Clusters

Upon examining the available characteristics for their relevance, the cyber risk incidents are grouped into three clusters based on the following characteristics: Country of legal entity, Country of incident, First year of event, Industry sector code, Assets (size), and Net income (profitability). Since most of the companies are based in the U.S. and most events have occurred in the U.S, we denote the country of legal entity and country of incident to be 1 if in the U.S., and 0 otherwise.

[Insert Table 5 & 6]

The cluster analysis results show case that by properly identifying relevant characteristics in cyber risk events, one can effectively group similar events while distinguishing between groups of events that exhibit significantly different patterns. In Table 5 we provide a set of descriptive statistics for each cluster and comparisons of the clusters. We can easily see that the three clusters are indeed quite different in terms of firm size, industry and many firm-based characteristics. For example, cluster 3 contains larger, longer lasting events that occur to much less profitable companies. These differences are evident in the two-sample t-tests we have conducted across clusters, which is available upon request. Table 6 confirms that while there is indeed some correspondence between

the three clusters identified based on event and firm characteristics and the event risk categories and/or activities defined in the *SAS OpRisk Global Data*, the cluster analysis reveals more subtle differences between different cyber risk events. For example, while "external fraud" is considered to be one general event risk category, these events are unevenly distributed across the three identified clusters, suggesting that when modeling these external fraud risks, one should consider also the specific risk characteristics, such as those revealed by our cluster analysis.

Cyber risk events may exhibit dependence within clusters and/or in-between clusters. We hypothesize that firms within each clusters would be more subject to contagious cyber risks than firms between different clusters. One natural way to validate our hypothesis is to take advantage of the available information contained in the variable "multiform impacted" in the *SAS OpRisk Global Data* and test if multiple firms that are impacted by the same event are indeed included in the same cluster. As described previously, since 1990, there are 491 events identified as cyber risk incidents and 129 of which have impacted multiple firms. Our validation results confirm that firms within each cluster would be more subject to contagious cyber risks than firms between different clusters at 91.5% accuracy.

4.3. The Factor Copula Model

Based on our results, we can capture cyber risk dependence among entities within each of the clusters developed. A very common and intuitive way to model the dependence is to use copulas. Copulas have been studied in both actuarial science and finance to examine dependencies among risks (c.f. Frees and Valdez 1998; Venter et al. 2007; Ai, Brockett and Wang 2017). In this paper, we propose to extract useful information from financial prices to enrich the sparse cyber risk data and by taking advantage of a new statistical development in factor copula models based on a latent factor structure (Zhang and Jiao 2012; Oh and Patton 2013). A factor copula model is generated by the following structural equation

$$X_i = \gamma_i Z + \varepsilon_i, \varepsilon_i \sim iid, Z \perp \varepsilon_i \ \forall i,$$

$$X \equiv [X_1, \dots, X_N]' \sim F_X = C(F_{X_1}(x_1), \dots, F_{X_N}(x_N); \theta),$$

where the X_i are latent variables, Z is the common factor and the ε_i are idiosyncratic factors.

The majority of our sample consists of public firms with financial prices that contain valuable information on cyber risks. Recent research by Lange and Burger (2017) shows that data breach has impact on the total returns and volatility of the affected companies' stock. In addition, Ahern (2013) and Foucault and Fresard (2014) illustrate that a firm's stock price learns and incorporates available market information from its network such as peers or supply chain streams. Accordingly, we hypothesize that cyber risk is reflected in the risk premium and hence stock price of the company, and a cyber risk event may affect the risk premium and stock price of its network as suggested in Ahern (2013) and Foucault and Fresard (2014). Therefore, we can match the *SAS OpRisk Global Data* with financial data of public firms for our subsequent analysis. Our proposed factor copula model approach is particularly attractive for cyber risk contagion modeling because it significantly enhances the available sparse data with public information from the market, and has a very flexible dependence structure for modeling systemic risks in the high dimensional space. The available real world data will be used to tune and validate the model.

4.4. A Simple Case Study of Target and Home Depot Data Breach

We now illustrate the use of factor copulas with a simple case study of Target and Home Depot data breaches to examine the impact of contagious cyber risk exposures. Without loss of generality and for ease of presentation, we showcase the features of the factor copulas model using this simplified, representative, and tractable setup of Target and Home Depot data breaches. Generalizations and extensions concerning more firms can be relatively easily derived.

We choose retail companies for our case study due to their susceptibility to cyber attacks as shown in previous descriptive analysis. The retail companies tend to share similar cyber risk because of the use of the same payment card systems and therefore are exposed to cyber risk contagion. Indeed the previous cluster analysis has identified both companies in the same cluster and as described below, the Target and Home Depot data

breaches are a typical example of cyber risk contagion. The two companies' *Point-of-Sale* systems were compromised by similar exploitation methods and the use of stolen third-party vendor credentials and RAM scraping malware were instrumental in the success of both data breaches. Target data breach was disclosed by Brian Krebs on December 18, 2013 with 40 million payment cards stolen (Krebs, 2014). Ever since then, occurrences of similar retail data breaches have been on the rise, including Neiman Marcus and Michael's in January 2014; Sally Beauty Supply in March 2014; P.F. Chang's in June 2014; Goodwill Industries in July 2014; SuperValu and The UPS Store in August 2014. Until the Home Depot data breach, the Target breach was the largest retail breach in U.S. history. The Home Depot data breach topped that by having 56 million payment cards stolen on September 2, 2014 when law enforcement and some banks contacted them about signs of the compromise (Krebs, 2014). The impact of these data breaches on each of the companies was significant. After the Target data breach, its posted quarterly profits was 46 percent below the expected profits (Gertz, 2014). Target and Home Depot stock prices both took a significant hit as well when the breach happened.

We make use of information in the stock returns of these publicly traded companies around the cyber attacks to study the contagion risk. Following the financial systemic risk literature (Zhang and Jiao 2012), we examine the dependence relationships between fluctuations on stock returns by using copula models conditional on the common factors found through the factor analysis and the marginal impact due to cyber risk. We focus on the Principal Component Analysis (PCA) method to extract the common factors that are responsible for the co-variation among the observed variables. The principal components are able to account for most of the variation in the observed stock returns. More specifically, we define r_i (i=1, 2) as the daily returns on Target and Home Depot during our sample period from December 18, 2013 (Target event date) to September 2, 2014 (Home Depot event date). If these returns were normally distributed, the joint distribution of them should be bivariate normal. However, a well-documented observation in the academic literature is that the probability distribution of financial series tends not be normal. Therefore, we followed the suggestions of Hull (2009) to transform the returns into normalized variables X_i (i=1, 2) using $X_i = \Phi^{-1}[F_i(r_i)]$, where Φ^{-1} is the inverse of the

cumulative standard normal distribution and F_i are the cumulative distribution functions for respective returns. In this transformation, the new variables X_i are constructed to have a standard normal distribution with mean equal to zero and standard deviation equal to one. After transforming the non-Gaussian returns into normally distributed variables, we find the common factors of these variables with the factor loadings and the percentages that the common factors account for in the underlying data. These results are exhibited in Table 7.

[Insert Table 7]

Note that this transformation is percentile to percentile so the correlations among the returns can be measured by the ones among the new variables. In the two-factor model,

$$X_i = \alpha_i F_1 + \beta_i F_2 + \sqrt{(1 - \alpha_i - \beta_i)}\, Z_i$$

where F_1 and F_2 are two common factors (latent factors from PCA or other factor analysis) affecting returns for both Target and Home Depot which include the impact of cyber risks, and Z_i s have independent standard normal distributions. The α_i and β_i are constant parameters between -1 and +1. The correlation between X_i and X_j is thus $\alpha_i \alpha_j + \beta_i \beta_j$.

[Insert Table 8]

The calculated correlations from the factor copula model are referred to as the copula correlations. Both unconditional correlations and copula correlations between Target and Home Depot returns are reported in Table 8. The difference between the two factors copula correlations and the unconditional correlation thus represents the isolated marginal effect by cyber risk. Based on our results, we can see that without using the factor copula model, the unconditional linear correlations of the Target returns and the Home Depot returns is about 0.4027. When using the factor copula model, we identified a significant increase of copula correlations to 0.7580, which indicates the impact of contagious cyber risks during the sample period. These results suggest that ignoring the increase of correlation due to cyber risk contagion may have significant impact to insurers and investors.

The two-stage model framework we have proposed in this paper is easily adaptable by insurers and businesses to build their own model for cyber risk contagion, based on either actual cyber risk incidents or simulated incidents for the analysis. In doing so, selected characteristics and clustering techniques can be used to group the large set of cyber risk

incidents into a desired number of groups. A factor copulas model can then be applied to each of these clustered groups to further model the dependence among different entities. Lastly, the difference between the unconditional correlations and the copula correlations can be used to assess the existence and the extent of the contagion effects. While the modeling process does involve choices that should be made according to the specific scenarios, the two-stage modeling approach is general and flexible to be used in many different settings. As a first step to propose such a modeling framework, there are naturally numerous avenues that future research can explore along. We discuss some of these future research opportunities in the Conclusion.

5. Conclusion

The cyber risk landscape is evolving rapidly and cyber security is one of the key concerns to modern organizations. As the complexity and severity of cyber risk continue to expand, businesses face greater systemic risk from cyber threats. Modeling and empirically examining the interconnected risk exposures will help reduce vulnerability of individual organizations and hence the entire economic system. At the same time, it represents a great opportunity as well as a significant challenge for the cyber insurance providers.

In this paper, we provide new modeling insights on cyber risk contagion and illustrate a two-step method based on cluster analysis and the factor copulas approach. The proposed framework is simple and flexible to accommodate specific concerns of the end users and can serve as a stepping-stone for businesses, insurers, regulators, and academics to develop their own models. This research can also serve as a critical starting component for organizations and (re)insurers to gradually build cyber risks into a broader ERM framework.

There are many intuitive ways to extend the current modeling framework. For example, in the current analysis, we use a sample of identified cyber risk incidents. However, further research can also adopt the propensity score matching method to include entities that have yet to experience a cyber attack, or use the Monte Carlo simulations method based on the initial analysis to increase the amount of usable data for actuarial pricing and risk management purposes. The proposed factor copula model is also flexible enough to allow

fat tail dependence and asymmetric dependence during recession or market boom and can be combined with semi-parametric marginal distributions.

Acknowledgements

This work was supported by the Casualty Actuarial Society through a research grant in 2017–2018. The authors thank the Casualty Actuarial Society for their very helpful support and feedback on improving this paper. All remaining errors are our own.

References

Ahern, K.R., 2013. Network centrality and the cross section of stock returns. Working paper.

Ai, J., Brockett, P.L. and Wang, T., 2017. Optimal enterprise risk management and decision making with shared and dependent risks. Journal of Risk and Insurance, 84(4), pp.1127-1169. Baldwin A, Gheyas I, Ioannidis C, Pym D, Willams J. 2012. Contagion in Cybersecurity Attacks. In: Workshop of Economics of Information Security (WEIS 2012).

Biener, C., Eling, M. and Wirfs, J.H., 2015. Insurability of cyber risk: An empirical analysis. The Geneva Papers on Risk and Insurance Issues and Practice, 40(1), pp.131-158.

Billio, M., M. Getmansky, A. W. Lo., and L. Pelizzon, 2012, "Econometric Measures of Connectenness and Systemic Risk in the Finance and Insurance Sectors," Journal of Financial Economics 104 (3): 535-559.

Bodin, L.D., Gordon, L.A. and Loeb, M.P., 2008. Information security and risk management. *Communications of the ACM*, *51*(4), pp.64-68.

Böhme, R., Kataria, G., 2006, Models and Measures for Correlation in Cyber-Insurance. WEIS.

Böhme, R., Schwartz, G., 2010, Modeling Cyber-Insurance: Towards a Unifying Framework. WEIS.

Cebula, J. J. and Young, L. R. (2010): A Taxonomy of Operational Cyber Security Risks, Technical Note CMU/SEI-2010-TN-028, Software Engineering Institute, Carnegie Mellon University.

Chapper, J.R., 2016. Worldwide Threat Assessment of the US Intelligence Community. Office Of The Director Of National Intelligence Washington DC.

Cummins, J.D. and Weiss, M.A., 2014. Systemic risk and the US insurance sector. Journal of Risk and Insurance, 81(3), pp.489-528.

Dionisi S. 2017. Determining the Likelihood of a Cybersecurity Failure for use in Cybersecurity Insurance Pricing. In Essay Collection of Cybersecurity: Impact on Insurance Business and Operations, Society of Actuaries (SOA), the Casualty Actuarial Society (CAS), and the Canadian Institute of Actuaries (CIA).

Duan, J.C. and Wei, J., 2009. Systematic risk and the price structure of individual equity options. Review of Financial Studies, 22(5), pp.1981-2006.

Eling, M. and Pankoke, D., 2016. Systemic risk in the insurance sector: Review and directions for future research. Risk Management and Insurance Review, 19(2), pp. 249-284.

Eling, M. and Wirfs, J.H., 2015. Modelling and management of cyber risk. Working paper.

Eling, M., Eling, M., Schnell, W. and Schnell, W., 2016. What do we know about cyber risk and cyber risk insurance?. *The Journal of Risk Finance*, *17*(5), pp.474-491.

Frees, E.W. and Valdez, E.A., 1998. Understanding relationships using copulas. *North American actuarial journal*, *2*(1), pp.1-25.

Foucault, T. and Fresard, L., 2014. Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, *111*(3), pp.554-577.

Gan, G., 2013. Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, *53*(3), pp.795-801.

Gartner, November 10, 2015. Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015. http://www.gartner.com/newsroom/id/3165317

Gordon, L.A., Loeb, M.P. and Sohail, T., 2003. A framework for using insurance for cyber-risk management. *Communications of the ACM*, *46*(3), pp.81-85.

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

Hofmann, A. and Ramaj, H., 2011. Interdependent risk networks: The threat of cyber attack. International Journal of Management and Decision Making, 11(5-6), pp.312-323.

Kopp, E.A., Kaffenberger, L. and Wilson, C.L., 2017. Cyber Risk, Market Failures, and Financial Stability.

Krebs, B. (2014, May 14). *The Target Breach, By the Numbers*. Retrieved from krebsonsecurity.com/2014/05/the-target-breach-by-the-numbers/

Krebs, B. (2014, September 14). *Home Depot: 56M Cards Impacted, Malware Contained*. Retrieved from krebsonsecurity.com/2014/09/home-depot-56m-cards-impactedmalware-contained/

Kwon, W.J., 2018. The Insurance Business in Transition to the Cyber-Physical Market: Communication, Coordination and Harmonization of Cyber Risk Coverages. Lange, R. and Burger, E.W., 2017. Long-term market implications of data breaches, not. *Journal of Information Privacy and Security*, *13*(4), pp.186-206.

Marotta, A., Martinelli, F., Nanni, S. and Yautsiukhin, A., 2015. A Survey on Cyber-Insurance.

Maxwell, L. 2017. Cybersecurity and the Insurance Market. In Essay Collection of Cybersecurity: Impact on Insurance Business and Operations, Society of Actuaries (SOA), the Casualty Actuarial Society (CAS), and the Canadian Institute of Actuaries (CIA).

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A. and Sadhukhan, S.K., 2013. Cyber-risk decision models: To insure IT or not?. Decision Support Systems, 56, pp.11-26.

Öğüt, H., Raghunathan, S. and Menon, N., 2011. Cyber Security Risk Management: Public Policy Implications of Correlated Risk, Imperfect Ability to Prove Loss, and Observability of Self-Protection. Risk Analysis, 31(3), pp.497-512.

Oh, D., and Patton, A. J., 2017. Modelling dependence in high dimensions with factor copulas. Journal of Business & Economic Statistics 1(35).

Oh, D., and Patton, A.J., 2013. Simulated method of moments estimation for copula-based multivariate models. Journal of the American Statistical Association, 108(502), 689-700.

PwC's Global Economic Crime Survey, 2014. PwC. Savor, P. and Wilson, M., 2016. Earnings announcements and systematic risk. The Journal of Finance, 71(1), pp.83-138.

RIMS, 2016. 2016 RIMS Cyber Survey.

Romanosky, S., Hoffman, D. and Acquisti, A., 2014. Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, *11*(1), pp.74-104.

Segal, A., 2016. The Hacked World Order: How Nations Fight, Trade, Maneuver, and Manipulate in the Digital Age. Public Affairs.

Shang K. 2017. Embedding Cyber Risk in Risk Management: An Insurer's Perspective. In Essay Collection of Cybersecurity: Impact on Insurance Business and Operations, Society of Actuaries (SOA), the Casualty Actuarial Society (CAS), and the Canadian Institute of Actuaries (CIA).

Solomon M. 2017. Cyber Risk is Opportunity. In Essay Collection of Cybersecurity: Impact on Insurance Business and Operations, Society of Actuaries (SOA), the Casualty Actuarial Society (CAS), and the Canadian Institute of Actuaries (CIA).

Venter, G., Barnett, J., Kreps, R. and Major, J., 2007. Multivariate copulas for financial modeling. *Variance*, *1*(1), pp.103-119.

Verizon, 2017, "Data Breach Investigations Report 2017," Verizon Enterprise. http://www.verizonenterprise.com/resources/reports/rp_DBIR_2017_Report_en_xg.pdf

Wang, Q.H. and Kim, S.H., 2009, May. Cyber Attacks: Cross-Country Interdependence and Enforcement. In WEIS.

Wei, L., Li, J. and Zhu, X., 2018. Operational Loss Data Collection: A Literature Review. *Annals of Data Science*, pp.1-25.

Xu, M. and Hua, L. 2017. Cybersecurity Insurance: Modeling and Pricing, SOA.

Zhang, S. and Jiao, F., 2012. Factor copula models and their application in studying the dependence of the exchange rate returns. International Business Research 5(2), 3-12.

Zurich Insurance Company and Atlantic Council, 2014. Risk Nexus: Beyond Data Breaches: Global Interconnections of Cyber Risk.

Table 1 Categories of Cyber Risk

Category		Description	Elements
Actions of people			
1.1	Inadvertent	unintentional actions taken without malicious or harmful intent	mistakes, errors, omissions
1.2	Deliberate	actions taken intentionally and with intent to do harm	fraud, sabotage, theft, and vandalism
1.3	Inaction	lack of action or failure to act upon a given situation	lack of appropriate skills, knowledge, guidance, and availability of person to take action
Systems and technology failures			
2.1	Hardware	risks traceable to failures in physical equipment	failure due to capacity, performance, maintenance, and obsolescence
2.2	Software	risks stemming from software assets of all types, including programs, applications, and operating systems	compatibility, configuration management, change control, security settings, coding practices, and testing
2.3	Systems	failures of integrated systems to perform as expected	design, specifications, integration, and complexity
Failed internal processes			
3.1	Process design and/or execution	failures of processes to achieve their desired outcomes due to poor process design or execution	process flow, process documentation, roles and responsibilities, notifications and alerts, information flow, escalation of issues, service level agreements, and task hand-off
3.2	Process controls	inadequate controls on the operation of the process	status monitoring, metrics, periodic review, and process ownership
3.3	Supporting processes	failure of organizational supporting processes to deliver the appropriate resources	staffing, accounting, training and development, and procurement
External events			
4.1	Hazards	events, both natural and of human origin, over which the organization has no control and that can occur without notice	weather event, fire, flood, earthquake, unrest
4.2	Legal issues	risk arising from legal issues	regulatory compliance, legislation, and litigation
4.3	Business issues	risks arising from changes in the business environment of the organization	supplier failure, market conditions, and economic conditions
4.4	Service dependencies	risks arising from the organization's dependence on external parties	utilities, emergency services, fuel, and transportation

Source: Biener and Eling (2014)

Table 2 Keyword Search Strings based on Three Criteria

Critical Asset	**Actor Category**	**Actor Category (cont.)**	**Outcome**
account	*(1) Actions by people*	*(2) Systems and technical failure*	availability
accounting system	administrator	defect	available
address	deadline	hardware	breach
code	denial of service, DoS	loading	breakdown
communication	destruction	malicious code	confidential
computer	devastation	software	congestion
computer system	employee	stress	constrain
confidential	extortion	system crash	control
confidential document	forgot, forget, forgotten	delete	
consumer information	hacker, hacked	*(3) Failed internal processes*	deletion
data	hacking	unauthorized access	disclosure
disk	human error	disorder	
document	infect	*(4) External events*	disruption
file	infection	Blizzard	disturbance
hard-disk	infiltrate	Earthquake	encryption
hard-drive	infiltrated	Eruption	espionage
homepage	key logger	Explosion	failure
info(rmation)	lapse	Fire	false
information system	logic bomb	Flood	falsification
internet site	maintenance	Hail	falsified
names	malware	heat wave	falsifying
network	manager	Hurricane	incompatibility
numbers	manipulate	Lightning	incompatible
online banking	miscommunication	natural catastrophe	incomplete
payment system	mistake	Outage	integrity
PC	misuse	pipe burst	interruption
personal information	omission	Riot	limit
phone	online attack	Smoke	lose
purchase information	oversight	Storm	loss
record	phish	Thunder	lost
reports	phishing	Tornado	malfunction
server	spam	Tsunami	missing
site	Trojan	Typhoon	modification
social security number	vandalism	Unrest	modified
stored information	virus	Utilities	modify
tablet	worm	War	overload
trade secret	Weather	publication	
webpage	Wind	restrict	
website	sabotage		
steal			
stole			
theft			

Source: Biener, Eling and Wirfs (2015)

Table 3 Refined Keyword Search Strings based on Three Criteria

Actor Category	**Actor Category**	**Actor Category**
(1) Actions by people	*(2) Systems and technical failure*	*(4) External events*
extortion	defect	Blizzard
hacker, hacked	hardware	Earthquake
hacking	loading	Eruption
infect	malicious code	Explosion
infection	software	Fire
infiltrate	stress	Flood
infiltrated	system crash	Hail
key logger	delete	heat wave
logic bomb		Hurricane
malware		Lightning
online attack		natural catastrophe
phish		Outage
phishing		pipe burst
spam		Riot
Trojan		Smoke
virus		Storm
worm		Thunder
Weather		Tornado
Wind		Tsunami
sabotage		Typhoon
		Unrest
		Utilities
		War
		publication
		restrict

Table 4 Summary Statistics on Firm Characteristics for Non-cyber and Cyber Operational Risk events

Panel A: Firms that have not had a cyber risk event

Variable	Obs.	Mean	Std Dev	P5	Median	P95
Assets (Millions)	25476	291,786	588,559	113	28,238	1,864,660
Employee	26054	61,574	122,719	94	21,126	264,200
Equity (Millions)	25275	28,565	57,230	10	5,705	149,137
Liability (Millions)	29217	40	506	0	0	106
Net Income (Millions)	25673	2,310	6,703	-1,012	417	14,099
Revenue (Millions)	26496	33,765	84,084	22	8,596	119,190
Loss Amount (Millions)	29219	71	675	0	4	209

Panel B: Firms that have had a cyber risk event

Variable	Obs.	Mean	Std Dev	P5	Median	P95
Assets (Millions)	462	249,129	566,158	240	29,181	1,698,155
Employee	468	75,036	118,064	224	24,200	323,000
Equity (Millions)	460	28,138	49,712	31	6,170	118,826
Liability (Millions)	491	64	689	0	0	113
Net Income (Millions)	462	2,400	5,835	-1,521	410	13,831
Revenue (Millions)	475	34,465	66,854	58	8,138	136,821
Loss Amount (Millions)	491	104	722	0	4	368

Table 5 Summary Statistics for the Three Identified Cyber Risk Clusters

	Cluster 1					
Variable	Obs.	Mean	Std Dev	P5	Median	P95
Event Length	264	2.07	3.03	0	1	9
Loss Amount	264	86.85	354.91	0.15	5.75	420.00
Country of Legal Entity	264	0.74	0.31	0.78	0.78	0.78
Country of Incident	264	0.74	0.00	0.74	0.74	0.74
Standardized Industry Sector Code	264	0.04	1.33	-0.47	0.11	0.16
Standardized Assets	264	-0.01	1.05	-0.44	-0.41	3.00
Standardized Net Income	264	0.04	0.98	-0.64	-0.34	1.96
	Cluster 2					
Variable	Obs.	Mean	Std Dev	P5	Median	P95
Event Length	169	1.59	2.19	0	1	7
Loss Amount	169	139.17	1143.72	0.11	3.20	264.20
Country of Legal Entity	169	-1.02	0.68	-1.27	-1.27	0.78
Country of Incident	169	-1.33	0.16	-1.34	-1.34	-1.34
Standardized Industry Sector Code	169	-0.03	0.28	-0.47	0.11	0.11
Standardized Assets	169	0.05	0.97	-0.44	-0.34	2.49
Standardized Net Income	169	-0.02	1.10	-0.67	-0.31	2.10
	Cluster 3					
Variable	Obs.	Mean	Std Dev	P5	Median	P95
Event Length	23	3.17	4.37	0	2	15
Loss Amount	23	169.28	292.57	0.38	21.00	940.00
Country of Legal Entity	23	-1.27	0.00	-1.27	-1.27	-1.27
Country of Incident	23	0.74	0.00	0.74	0.74	0.74
First year of event	23	0.18	0.87	-1.29	0.30	1.19
Standardized Industry Sector Code	23	-0.14	0.31	-0.47	0.08	0.16
Standardized Assets	23	-0.14	0.65	-0.44	-0.39	0.62
Standardized Net Income	23	-0.26	0.39	-0.69	-0.39	0.20

Note: The variables of Industry Sector Code, Assets and Net Income are standardized for analysis.

Table 6 Cluster Analysis Results vs. Event Risk Categories

Event Risk Category	Cluster 1	Cluster 2	Cluster 3
Business Disruption and System Failures	13	19	0
Clients, Products & Business Practices	97	50	16
Damage to Physical Assets	3	2	1
Employment Practices and Workplace Safety	4	1	0
Execution, Delivery & Process Management	22	4	3
External Fraud	94	74	3
Internal Fraud	31	19	0

Table 7 Principal Components Analysis for Common Factors in Stock Returns

Principal Component Analysis		
Component	Common Factor 1	Common Factor 2
Normalized Target Return	0.7071	-0.7071
Normalized Home Depot Return	0.7071	0.7071
Variance	1.4027	0.5973
Variance Percentage	70.14%	29.86%

Table 8 Unconditional and Copula Correlations for Stock Returns

Panel A: Unconditional Correlation			Panel B: Copula Correlation		
	Target	Home Depot		Target	Home Depot
Target	1	0.4027	Target	1	0.7580
Home Depot	0.4027	1	Home Depot	0.7580	1

Appendix Additional Descriptive Statistics for the Cyber Contagion Risk Data Set

Table A1 Count of Cyber Risk Events by Event Starting Year

Event Starting Year	Total Count of Event	Cyber Risk Events	Actor Category 1	Actor Category 2	Actor Category 4
1990	637	0	0	0	0
1991	621	1	0	1	0
1992	682	3	1	2	0
1993	708	1	0	1	0
1994	753	3	1	2	0
1995	894	3	0	3	0
1996	875	1	0	1	0
1997	1044	3	0	3	0
1998	1274	8	1	7	0
1999	1524	3	0	3	0
2000	1803	2	0	2	0
2001	1750	9	1	8	0
2002	1566	15	1	13	1
2003	1519	12	0	11	1
2004	1680	29	11	17	1
2005	1830	14	2	11	1
2006	1837	28	6	18	4
2007	2045	27	8	16	3
2008	1751	32	9	22	1
2009	1296	61	21	38	2
2010	1189	38	15	17	6
2011	857	25	5	19	1
2012	574	35	15	18	2
2013	524	31	16	15	0
2014	268	41	11	29	1
2015	149	24	8	12	4
2016	54	28	13	14	1
2017	6	14	6	6	2

Table A2 Length of Events

Length of Events	Count of Cyber Events	Actor Category 1	Actor Category 2	Actor Category 4
0	223	102	113	8
1	84	28	47	9
2	52	9	39	4
3	40	3	34	3
4	20	1	17	2
5	16	1	13	2
6	18	4	13	1
7	13	0	11	2
8	4	0	4	0
9	6	1	5	0
10	7	1	6	0
11	1	0	1	0
12	2	1	1	0
13	1	0	1	0
14	1	0	1	0
15	2	0	2	0
17	1	0	1	0

*Note that length = 0 means the event occurrence period is within a year

Table A3 Top 10 Countries of Companies that Had Cyber Risk Events

Country of Legal Entity	Count of Cyber Events	Actor Category 1	Actor Category 2	Actor Category 4
United States	304	83	199	22
United Kingdom	30	10	20	0
India	20	3	17	0
Germany	17	0	16	1
Japan	17	3	13	1
France	12	6	3	3
Canada	8	4	4	0
Australia	7	2	5	0
Brazil	7	5	2	0
Sweden	6	2	4	0

Table A4 Industry Sectors of Companies that Had Cyber Risk Events

Industry Sector	Count of Cyber Events	Actor Category 1	Actor Category 2	Actor Category 4
Financial Services	226	99	107	20
Manufacturing	104	9	90	5
Information	88	19	66	3
Retail Trade	23	14	9	0
Professional, Scientific and Technical Services	18	1	17	0
Utilities	7	2	4	1
Public Administration	4	1	3	0
Transportation and Warehousing	4	2	1	1
Administrative and Support, Waste Management and Remediation Services	3	0	3	0
Construction	3	0	3	0
Mining	3	1	2	0
Agriculture, Forestry, Fishing and Hunting	2	0	1	1
Health Care and Social Assistance	2	0	2	0
Accommodation and Foodservices	1	1	0	0
Non-Profit Organizations	1	1	0	0
Other Services (except Public Administration)	1	1	0	0
Wholesale Trade	1	0	1	0