Florackis, Chris; Louca, Christodoulos; Michaely, Roni; Weber, Michael

Working Paper

Cybersecurity Risk

CESifo Working Paper, No. 8760

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Florackis, Chris; Louca, Christodoulos; Michaely, Roni; Weber, Michael (2020) : Cybersecurity Risk, CESifo Working Paper, No. 8760, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/229578

CESifo WORKING PAPERS

8760
2020

December 2020

Cybersecurity Risk

Chris Florackis, Christodoulos Louca, Roni Michaely, Michael Weber

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8760*

Cybersecurity Risk

Abstract

We develop a novel firm-level measure of cybersecurity risk using textual analysis of cybersecurity-risk disclosures in corporate filings. The measure successfully identifies firms extensively discussing cybersecurity risk in their 10-K, displays intuitive relations with quantitative measures of cybersecurity risk disclosure language, exhibits a positive trend over time, is more prevalent among industries relying more on information technology systems, correlates with several characteristics linked to firms hit by cyber attacks and, importantly, predicts future cyber attacks. Stocks with high exposure to cybersecurity risk exhibit high expected returns on average, but they perform poorly in periods of increasing attention to cybersecurity risk.

JEL-Codes: G140, G320.

Keywords: cyber attacks, risk disclosures, textual analysis, stock returns.

Chris Florackis
University of Liverpool / United Kingdom
C.Florackis@liverpool.ac.uk

Christodoulos Louca
Cyprus University of Technology / Cyprus
Christodoulos.Louca@cut.ac.cy

Roni Michaely
University of Geneva / Switzerland
Roni.Michaely@unige.ch

Michael Weber
University of Chicago, Booth School of Business, Chicago / IL / USA
Michael.Weber@chicagobooth.edu

This Version: December 3rd 2020
We gratefully acknowledge valuable comments from several seminar and conference participants. Weber also gratefully acknowledges financial support from the University of Chicago Booth School of Business and the Fama Research Fund.

1. Introduction

Cybersecurity risk is the risk of financial loss, disruption, or damage to the reputation of a firm as a result of a failure in its information technology systems due to external attacks (Institute of Risk Management).[1] Examples of cybersecurity risk include the risk of losing sensitive data, disruption in a firm's network, systems, and services, and physical electronic damage. Cybersecurity risk is currently considered one of the top global concerns for firm executives and market participants in advanced economies,[2] which is not surprising given the increase in major cyber attacks in recent years.[3] Despite substantial investments in information security systems, firms remain highly exposed to cybersecurity risk,[4] with possible losses amounting to $6 trillion annually by 2021.[5] Although the attacks and the possible preventive measures are well studied in the literature, whether a firm's exposure to cybersecurity risk is priced in financial markets remains unclear.

We propose a novel firm-level measure of cybersecurity risk for all listed firms in the U.S. and examine whether heterogeneity in cybersecurity risk is priced in the cross section of stock returns. We find that portfolios of firms with high exposure to cybersecurity risk outperform other firms by up to 8.3% per year in terms of equal-weighted (7.9% value-weighted) returns. Our measure of cybersecurity risk is a robust return predictor, and its effect is not subsumed by standard return predictors in Fama-MacBeth regressions. A cybersecurity-mimicking portfolio performs poorly in times of heightened cybersecurity risk and investors' concerns about data breaches.

[1] IRM Cyber Risk: Executive Summary, pp. 1-16.
[2] See "The Global Risks Report 2019" (14th Edition), World Economic Forum, and the 2017 survey from DTCC Systemic Risk Barometer (Bouveret, 2018, provides further information).
[3] According to a recent report from the Center for Strategic and International Studies and McAfee, the amount lost to cybercrime every year is experiencing a rapidly increasing trend (valued at nearly 1% of global GDP for 2018).
[4] Gartner, a global research and advisory firm, for example, estimates worldwide spending on information security products of $124 billion in 2019, representing an increase of 8.8% relative to 2018.
[5] See "Cybercrime Damages $6 Trillion by 2021. In addition, Steve Morgan, Founder of Cybersecurity Ventures and Editor-in-Chief at Cybercrime Magazine, suggests the cybercrime damage costs could potentially double during the COVID-19 outbreak period.

To construct our measure, we use firms that were subject to cyber attacks as a training sample, and then compare the wording and language in the relevant risk-disclosure section in annual reports of the attacked firms with that of all other firms. Specifically, we first extract the discussion on cybersecurity risk in the "Item 1A. Risk Factor" section from firms' 10-K, which contains information about the most significant risk factors for each firm on Edgar over the period 2007-2018. Second, we identify a sample of firms that have been subject to a major cyber attack (involving lost personal information by hacking or malware-electronic entry by an outside party) in any given year. We argue these firms have high cybersecurity risk, and they serve as our training sample. Third, we estimate the similarity of each firm's cybersecurity-risk disclosure with past cybersecurity-risk disclosures of firms in the training sample (i.e., from the one-year period prior to the firm's filing date).[6] The higher the measured similarity in cybersecurity risk disclosure for our sample firms and firms in the training sample, the greater the exposure to cybersecurity risk.

We validate our measure in several ways. First, firms that score high on our measure (i.e., top 5 firms) emphasize cybersecurity risk in their 10-K filings more than firms with low scores (i.e., bottom 5 firms). For instance, top 5 firms typically mention that the increasing sophistication of hackers makes defending against cybersecurity attacks difficult, despite investments in preventive systems. Firms with the 5 lowest scores instead tend to emphasize that they can adequately deal with cybersecurity risk through preventive measures. Moreover, these firms typically do not devote a separate section to cybersecurity risks in their 10-Ks.[7]

Second, firms with higher scores provide lengthier and more comprehensive cybersecurity-risk disclosures in their 10-Ks, discuss legal consequences associated with cybersecurity risk, use more precise language, and use more negative words in their

[6] Other studies that use document similarity to extract meaning from text collections include, among others, Hoberg and Phillips (2010; 2016), Brown and Tucker (2011), Hoberg and Maksimovic (2015), Lang and Stice-Lawrence (2015), and Lowry, Michaely, and Volkova (2020).
[7] We find similar patterns when we consider more firms (e.g. top 10 *vs.* bottom 10).

discussions, which potentially lowers their exposure to litigation risk (Loughran and McDonald, 2011).

Third, high-score firms actively manage their exposure to cybersecurity risk through real actions. Within our sample, a non-negligible number of firms purchase cyber insurance policies; notably, our measure is positively correlated with the presence of cyber insurance policies, supporting the view that firms use cyber insurance to partially protect against claims that may arise due to cyber attacks.

Fourth, our measure exhibits an increasing trend over time, especially after 2011, when the SEC issued for the first time specific disclosure obligations relating to cybersecurity risks and cyber incidents. The estimated cybersecurity score for the average firm in our sample increased from 0.153 in 2011 to 0.454 in 2018. Importantly, this increase is not only an average effect; whereas 49.03% of our sample firms exhibit zero cybersecurity risk in 2011, only 10.59% of them exhibit zero cybersecurity risk in 2018. Overall, these results are consistent with the recent growth in the number and significance of successful cyber attacks against major organizations, as well as firms' increasing vulnerability to cyber attacks.

Fifth, our measure is particularly high in industries that rely heavily on information technology system to perform their operations, which makes them more vulnerable to cyber attacks (e.g., the Telephone & Television Transmission, Business Equipment, and Money Finance industries). According to our calculations, these industries exhibit a high cyber-attack incident rate (see also Romanosky, 2016).

Sixth, our measure correlates with firm characteristics that previous research linked to firms hit by cyber attacks. For example, in line with Kamiya et al. (2020), our measure relates cross sectionally with firm characteristics such as size, age, profitability, growth opportunities and tangibility. It is also positively associated with other characteristics that likely indicate vulnerability to cyber attacks such as R&D expenditures and the presence of trade secrets.

Seventh, based on the theoretical premise that any kind of bad news should induce negative asymmetry in stock returns (see, e.g., Campbell and Hentschel, 1992), we posit that exposure to cybersecurity risk should result in negative returns once the risk materializes; that is, we expect large negative stock returns when firms are subject to cybersecurity attacks. Consistent with this view, our measure is positively associated with (negative) asymmetries in stock returns.

Finally, and most directly, we show firms with higher cybersecurity risk scores are more likely to experience a future cyber-attack. In economic terms, one standardized unit increase in our cybersecurity risk score increases the probability of a future cyber attack by 92.70%. Taken together, our firm-level measure of cybersecurity risk has features that one would expect for firms indeed being exposed to the risk of cyber attacks.

Although the measure we propose has properties that one would associate with a heightened risk of cyber attacks, another way to validate the measure is to check whether cybersecurity is priced in the cross section of stock returns. Accordingly, we sort stocks into portfolios based on their cybersecurity-risk score and track their future returns over time. Firms with high cybersecurity risk exhibit higher future returns. Specifically, an equal-weighted portfolio that goes long stocks with high cybersecurity risk and shorts stocks with low cybersecurity risk earns a statistically significant excess return of 66 to 69 basis points per month, or 8.3% per year; similar results exist for value-weighted portfolios (7.9% per year). High cybersecurity-risk portfolios differ from low cybersecurity-risk portfolios in terms of several firm- and 10-K-specific characteristics. Through bivariate portfolio sorts, we confirm the premium remains robust across all sub-samples of stocks sorted by size, book-to-market, profitability, institutional ownership, illiquidity, idiosyncratic volatility, risk-section length, and 10-K readability (Fog-Index). We also show the excess returns of high versus low

cybersecurity-exposure stocks is larger when we exclude firms that partially insure against cyber attacks.

We also examine the cross-sectional relation between cybersecurity risk and stock returns by running stock-level Fama-MacBeth (1973) regressions and document a strong positive relation between cybersecurity risk and stock returns. Interestingly, we find that cybersecurity risk predicts cross sectional variation in stock returns up to 12 months into the future. Accordingly, the predictability is not a short-term phenomenon.

Finally, we introduce a cybersecurity-risk factor and test its economic and statistical significance for the full sample and for important subsample periods (i.e., upon the occurrence of events that increase attention to cybersecurity risk). If our measure accurately captures cybersecurity risk and it is a priced source of risk, then high-cybersecurity-risk stocks should perform poorly and significantly worse than low-cybersecurity-risk stocks on the days of intense attention toward cybersecurity risk. To perform the analysis, we resort to daily data and identify days of increasing attention to cybersecurity risk based on abnormal search volume index (SVI) of the search topics "Hacker" and "Data Breach" in Google Trends. We find that the cybersecurity-risk factor exhibits poor performance during periods of increasing attention to cybersecurity risk, although generally it performs well throughout our sample period.

This study contributes to the literature in several ways. First, it adds to a growing literature extracting important economic information utilizing text as data.[8] For example, Baker, Bloom and Davis (2016) use newspaper articles to develop an index of economic-policy uncertainty. More recently, Hassan, Hollander, van Lent, and Tahoun (2019) and Sautner, van Lent, Vilkov, and Zhang (2020) utilize text from earnings conference calls to develop firm-level measures of political risk and climate-change exposure, respectively. Other studies use text from financial

[8] Gentzkow, Kelly, and Taddy (2019) provide an introduction of text in economic research whereas Loughran and Mcdonald (2016) discuss commonly applied textual-analysis methods and provides an excellent review of the literature.

reports, such as 10-Ks and 10-Qs. Cohen, Malloy, and Nguyen (2020) link changes in the language of financial reports to future firm operations. Hoberg and Maksimovic (2015) and Buehlmaier and Whited (2018) use the management's discussion and analysis section to obtain measures of financial constraints. Finally, Frésard, Hoberg, and Phillips (2020) link product descriptions with vertically-linked product descriptions from the Bureau of Economic Analysis to construct measures of vertical relatedness. Most relevant to our work are the studies that extract information from the risk-factor disclosures section in 10-Ks. For example, Campbell et al. (2014) find that risk-factor disclosures are not "boilerplate" and are positively associated with post-disclosure market-based measures of firm risk. We are the first to focus on cyber-related risk disclosures and examine whether these convey useful information about firm exposure to cyber threats and the associated costs, rather than focusing on overall risk exposure. Using an estimation procedure that is transparent, objective, and easily implementable, we contribute to this literature by developing, for the first time, a firm-level measure of cybersecurity risk.

Second, we add to the asset-pricing literature by showing that cybersecurity risk is priced in the cross section of stocks. Our tests show that stocks of firms exposed to high cybersecurity risk earn higher expected returns. Further, we observe a degree of commonality in cybersecurity risk among US stocks, especially around periods of increasing attention to cybersecurity risk. This finding is consistent with the view that cybersecurity risk is priced as a systematic risk factor and investors require a premium to hold stocks exposed to high cybersecurity risk.

Finally, we also add to the literature focusing on the implications of cyber attacks on the attacked firms. For example, several studies focus on the valuation impact of cyber attacks (see, e.g., Hilary, Segal, and Zhang, 2016; Johnson, Kang, and Lawson, 2017; Amir, Levi, and Livne, 2018; Lending, Minnick, and Schorno, 2018; and Tosun, 2020); other studies focus on

how firms adjust their financial, investment, governance, and risk-management policies following costly cyber attacks (see, e.g., Akey, Lewellen, and Liskovich, 2020; Kamiya et al., 2020). Instead of focusing only on attacked firms, we resort to cyber-related disclosures for the population of US traded firms and assess their cybersecurity-risk exposures.

The remainder of the study is organized as follows. In section 2, we present our data, develop the cybersecurity-risk measure, and provide descriptive statistics. In section 3, we evaluate our measure and its ability to capture cybersecurity risk. Section 4 presents results on the relation between cybersecurity risk and stock returns, and section 5 provides various robustness tests. Finally, section 6 concludes.

2. Data and Methods

2.1 Data

We combine several databases to construct the sample. We use the Center for Research in Security Prices (CRSP) to obtain stock returns, Standard and Poor's Compustat Industrial Annual (CIA) to obtain financial information, Thomson-Reuters 13F database to obtain information on institutional ownership, BoardEx to obtain corporate governance-related information, SEC Edgar for annual filings, and Privacy Rights Clearinghouse (PRC) to collect data on cyber attacks.[9] The final sample with complete information covers the period 2007-2018 and consists of 5,534 firms with 35,308 firm-year observations.

2.2 Cybersecurity-risk Disclosures

We use a web-crawling algorithm to download all "10-K," 10-K405," "10-HSB," or 10-KSB40" filings, excluding amended documents from SEC Edgar and extract the fiscal year,

[9] PRC is a non-profit organization that aims to increase consumers' awareness of privacy protection (for more details, see https://privacyrights.org/).

the central index key (CIK), and the cybersecurity-risk disclosures from "Item 1A. Risk Factor" section.[10] Appendix A provides detailed information about our disclosure-extraction procedure and examples about its reliability. Note we exclude all firms that do not have an "Item 1A. Risk Factor" section; these firms are typically small, as defined by SEC Regulation S-K Item 10, and are not required to provide information about risk factors. Furthermore, like Hoberg and Phillips (2010), we exclude firms that incorporate the "Item 1A. Risk Factor section" by reference.[11] Finally, we link each firm's cybersecurity-risk disclosures with the CIA database using the fiscal year, the CIK, and the mapping table from the WRDS SEC Analytics suite.

2.3 Training Sample

We obtain from PRC information about firms that were subject to a data breach, a short description of the incident, the date the event was made public, the type of breach, the type of organization, and, if available, the number of records that were affected. Following Kamiya et al. (2020), we exclude incidents on governments, educational institutions, and non-profit organizations, and focus only on cyber attacks that involve lost personal information by hacking or malware-electronic entry by an outside party. We collect information on all recorded cyber attacks, and manually search news articles from Factiva to cross reference the information and to identify which cyber attacks attracted the attention of global news outlets (e.g., CNBC, Financial Times, Wall Street Journal) or are covered in major Newswires (e.g., AP, Bloomberg, Reuters). We call such cyber attacks "major" and use them as our training sample. Using only major attacks ensures the cybersecurity-risk estimation approach only employs information that is widely disseminated and available to investors (nevertheless, we repeat our experiment using all incidents of cyber attacks as the training sample, and the results

[10] According to SEC Regulation S-K Item 305 publicly listed firms must disclose "Item 1A. Risk Factor" section in 10Ks since December 1, 2005.
[11] One example of such a 10-K is as follows:
https://www.sec.gov/Archives/edgar/data/72971/000007297114000337/wfc10k_20131231.htm

are unchanged). Out of a total of 175 cyber attacks identified during the period 2005-2018 with available cybersecurity-risk disclosures in Item 1A. Risk Factors, 69 are classified as major cyber attacks that attracted extensive media coverage. These major cyber attacks span the period 2006-2018 and correspond to 54 firm-year cyber attacks (e.g., a firm may exhibit more than one cyber attack in a given year). Note the first cyber attack appears in 2006; therefore, given that our cybersecurity-risk measure utilizes past disclosures of firms that have been subject to cyber attacks, the earliest year we can estimate cybersecurity risk is 2007. Finally, we manually link the names of these firm-year cyber attacks in the PRC database with firm names in CRSP and CIA.

2.4 Cybersecurity-risk Measure

Given the growing dependence of firms on information technology systems to perform their operations, the risk associated with cybersecurity has increased over time. As a result, firms have to provide qualitative information about how cybersecurity risk affects their operations (SEC Regulation S-K Item 305). The SEC issued specific guidelines in 2011 and 2018, instructing public companies to inform their investors about material cybersecurity risks and incidents in a timely, comprehensive, and accurate manner (see, SEC, CF Disclosure Guidance: Topic No. 2 Cybersecurity, October 13, 2011; and updated SEC, Commission Statement and Guidance on Public Company Cybersecurity Disclosures, February 21, 2018). The guidelines apply to both the attacked companies and companies that are subject to material cybersecurity risks but may not yet have been the target of a cyber attack.

We use the textual information about cybersecurity risk in "Item 1A. Risk Factors" to create the cybersecurity-risk measure. The measure is based on how similar each firm's cybersecurity-risk disclosure is to *past* cybersecurity-risk disclosures of firms that have been subject to cyber attacks; that is, firms in our training sample. The idea behind the measure is firms more

vulnerable to cyber attacks are actually attacked, express this heightened risk ex-ante in their disclosure and that firms that use similar words to describe risk exposure and exposure management, exhibit similar levels of cybersecurity risk. This approach is quite common in information processing and has been recently applied in finance and economics. For instance, Hoberg and Phillips (2010) estimate product-market language similarity between firms, and Hoberg and Maksimovic (2015) estimate the similarity of firms' liquidity and capitalization resources relative to a training set of financially constrained firms.

After excluding certain types of words (e.g., pronouns, conjunctions, stop words, common words and/or articles, compound words, words that refer to geographic locations or names, and words with frequency less than 10), we store the text in separate word vectors. In line with previous work, these vectors are based on word roots rather than actual words. We identify word roots using a web-crawling algorithm and *https://www.merriam-webster.com/*. The universe of all words in the sample is 3,210 and the top 20 most common words in the text include: "security," "system," "information," "result," "business," "breach," "data," "operation," "customer," "service," "failure," "loss," "financial," "damage," "computer," "include," "technology," "disruption," "reputation," "unauthorized". Then, for each firm, we populate the vector of 3,210 words with the count of the number of times each word appears in the cybersecurity-risk disclosures and use this vector to measure the similarity between any two 10-K documents.

Next, for each firm and year, we consider the N_{t-1} firms that have been subject to cyber attacks during the one-year period ending at the firm's filing date (training sample).[12] For each firm and year, we then calculate the cosine similarity ($CS_{i,n,t}$) and the Jaccard similarity ($JS_{i,n,t}$) of the cybersecurity-risk disclosures with all N_{t-1} disclosures of firms that have been subject

[12] If no cyber attack occurs in the previous one-year period, we look for cyber attacks in the previous two-year period.

to a cyber attack (i.e., for each firm and year, we have N_{t-1} such similarities). Cosine similarity is defined as the cosine angle between two text vectors, whereas Jaccard similarity is defined as the size of the intersection divided by the size of the union of the two vectors (see Hanley and Hoberg, 2010 and Cohen, Malloy and Nguyen, 2020, for more information). Both cosine and Jaccard similarities are bounded between (0,1), and greater values imply a firm's words overlap more with the vector of words for firms that have been subject to past cyber attacks (i.e., more similar cybersecurity-risk disclosures). Finally, we define the cybersecurity risk for each firm and year as the average cosine or Jaccard similarity across all N similarities:

$$Cybersecurity\ Risk_{i,t} = \sum_{n=1}^{N} \frac{CS_{i,n,t}}{N} \quad [1]$$

$$Cybersecurity\ Risk_{Jaccard_{i,t}} = \sum_{n=1}^{N} \frac{JS_{i,n,t}}{N} \quad [2]$$

3. Validation

In this section, we describe the output of our measure and present various tests to verify that it captures exposure to cybersecurity risk.

3.1 Excerpts from Cybersecurity-risk Disclosures

Our measure utilizes cybersecurity-risk-related disclosures from Item 1A. Risk Factors in firms' 10-Ks. To gain some intuitive understanding of the relevance of its content, Panel A (B) of Table 1 compares excerpts from cybersecurity-risk disclosures, focusing on the five firms with top scores and the five with the lowest positive score.

Seeing that firms with the highest scores emphasize risk in their discussions in 10-K (see Panel A of Table 1) is reassuring; for instance, the firm with the highest score (Walgreens Boots Alliance Inc) acknowledges the businesses it interacts with, and the firm itself, have experienced threats to their data and systems. Other firms highlight the difficulty and

impossibility of defending against every risk, because the techniques used to attack change frequently, and attacks can originate from a wide variety of sources, which creates a risk of cybersecurity incidents. As the excerpts in Panel B suggest, the discussions of firms with the lowest score are quite different. For example, the firm with the lowest score (Weyerhaeuser Co) mentions their service providers and the firm itself employ *adequate* security measures, whereas other firms simply discuss cyber attacks *in conjunction* with other risks. Overall, firms with the lowest score appear to believe that, through preventive measures, they can adequately deal with cybersecurity risk, and that cybersecurity risk is not important enough for explicit and separate discussion.

This evidence suggests firms with high values of our measure indeed discuss threads of cybersecurity-risk extensively in their risk disclosures, whereas firms with low scores manage these risks and threads adequately and face little risk.

3.2 Cybersecurity-risk-disclosure Language

Another way to verify that our measure captures variation in exposure to cybersecurity risk is to directly study how it correlates with certain language features of the risk-disclosure section. Intuitively, we would expect firms facing a higher threat of cybersecurity risk to spend more time discussing these risks relative to other risks. Table 2 reports the results. We find a positive correlation of the measure with the number of cybersecurity-risk-disclosure sentences (*CRD Sentences (#)*) and the ratio of the number of cybersecurity-risk-disclosure sentences scaled by the number of sentences in the Item 1A. Risk Factors section (*CRD Sentences (Ratio)*) (0.57 and 0.43, respectively). This finding suggests firms with higher scores tend to have more comprehensive disclosures and perceive cybersecurity risk as a more important source of risk (compared with other types of risks that firms face).

We also use a collection of predefined words constructed by Loughran and McDonald (2011) to extract additional information about certain attributes of cybersecurity-risk disclosures. Specifically, our basic premise is that managers anticipating cybersecurity-risk challenges will communicate their concerns to shareholders; doing so would help them lower their exposure to litigation risk. Consistent with this view, we find firms with higher scores discuss significant legal consequences (*Litigious words*), use more precise language (*Precise words*), and use more negative words (*Negative words*) in their relevant discussions; the corresponding correlations between our measure and the variables *Litigious words, Precise words,* and *Negative words* are modest but positive (0.13, 0.08, and 0.03, respectively), and they are all statistically significant at the 1% level.

Further, we would expect firms subject to more cybersecurity-risk exposure to actively manage their exposure through real actions. One such real action is to use a cyber insurance policy. By looking for the word "insurance" in the cybersecurity-risk disclosures, we identify a non-negligible number of firms that explicitly mention insurance policies (8.43% of all firm-years in our sample). The vast majority of these firms (80%) have high above median cybersecurity-risk scores. We manually read all disclosures in which the word "insurance" appears. In almost all cases, firms mention their insurance policy only partially protects them against claims that may arise due to cyber attacks. For example, Apple Inc in its cybersecurity-risk disclosures for fiscal year 2017 states, "While the Company maintains insurance coverage that is intended to address certain aspects of data security risks, such insurance coverage may be insufficient to cover all losses or all types of claims that may arise". Likewise, Verizon Communications Inc in its cybersecurity-risk disclosures for fiscal year 2017 states, "The potential costs associated with these attacks could exceed the insurance coverage we maintain". Our measure is positively correlated (0.16) with the existence of an insurance policy (*Cyber Insurance*).

Overall, our findings reveal correlations between our cybersecurity-risk measure and a series of quantitative measures of language, which are computed using a different methodology (i.e., by using word lists). Therefore, these results further suggest our measure captures exposure to cybersecurity risk.

3.3 Time-series and Industry Properties

Figure 1 presents the yearly average value of our measure as well as the number of successful cyber attacks per year. The figure shows a positive time trend, especially after 2011, when the SEC issued the first cybersecurity-disclosure requirements. The observed increased in the average score is notable, increasing from 0.153 in 2011 to 0.454 in 2018. In addition, whereas 49.03% of the firm-years exhibit zero cybersecurity risk in 2011, only 10.59% do in 2018. Overall, this period is characterized by increasing concerns over cybersecurity risk, which largely originate from the large number of successful cyber attacks against public firms (e.g., 32 incidents in 2014, up from 11 in 2010 and 9 in 2012). A simple correlation between our measure and the percentage of cyber attacks per year is 0.72, suggesting the time-series properties of our measure aligns well with the percentage of cyber attacks.

Figure 2 presents the average value of the cybersecurity-risk measure across the 12 Fama and French industries. The measure exhibits considerable across-industry differences. Cybersecurity risk is more pronounced in Telephone and Television Transmission; Wholesale, Retail and Some Services; Business Equipment and Money Finance sectors. All of these industries rely on information technology systems, which makes them more vulnerable to cyber attacks. Indeed, 125 cyber attacks or 71.4% of the total number of cyber attacks occur in these industries. Firms in more “traditional” industries such as Energy, Oil and Gas; and Manufacturing exhibit much lower cybersecurity risk and fewer cyber attacks.

Taken together, both the time-series and industry variation of our measure has features that intuitively relate to cybersecurity risk: the average exposure and the number of firms exposed to cybersecurity risk increases over time, and firms in industries that are more reliant on information technology are more exposed to cybersecurity risk than other firms.

3.4 Firm and 10-K Characteristics

Table 3 presents descriptive statistics of our measure as well as various firm, industry, 10-K, and corporate-governance characteristics. We detail all variable definitions in Appendix B. To mitigate the impact of outliers, we winsorize the continuous variables in the sample at the 1st and 99th percentiles (by year). The results show the average score (*Cybersecurity Risk Index*) is 0.24. Because our measure is based on cybersecurity-risk disclosures, and several firms in our sample started providing such disclosures after the SEC specific guidelines in 2011, the 25th percentile is zero.[13] In addition, the 75th percentile of our measure is 0.45. Overall, as expected, the distribution of our measure is not normal and exhibits a positive skewness.

We then employ a linear regression model to examine how our measure relates to firm characteristics. The model also includes industry, 10-K, and corporate-governance characteristics. Table 4 reports the results. In Model 1, we control for industry and year fixed effects, whereas in Model 2, we control for firm and year fixed effects. Including firm fixed effects removes the impact from possible boilerplate or generic cybersecurity-risk disclosures that could lead to highly sticky scores across time for the same firm. Finally, standard errors are clustered at the firm level. The results show a positive association between our measure and firm size (*Firm Size (ln)*), growth opportunities (*Tobin's Q*), and profitability (*ROA*). These results indicate the score is higher for typically more visible firms. Firms with higher scores are also younger (*Firm Age (ln)*), have trade secrets (*Secrets*), and spend more on research and

[13] Excluding firm-year observations prior to 2011 from the analyses, does not alter our main conclusions.

development (*R&D Expenditures*). Naturally, such firms are expected to be more vulnerable to cyber attacks.

For the remaining variables, our measure is negatively related to industry cash-flow volatility (*Cash Flow Volatility (Industry)*), which suggests risky innate operations that characterize certain industries, as reflected by cash-flow risk, are associated with our measured scores. Our measure is also related to 10-K and corporate-governance characteristics. Firms with higher scores also have lengthier Item 1A. Risk Factors sections (*Risk Section Length (ln)*) and less readable 10-Ks (*Readability (ln)*). These results support the view that firms with higher scores are inherently riskier firms. In addition, firms with better governance quality (e.g., those with higher institutional ownership (*Institutional Ownership*), more independent directors sitting on their boards (*Independent Directors*), and a separate risk committee (*Risk Committee*)) exhibit a higher score. These findings might indicate firms with better corporate governance are also pre-emptively more active in attempting to understand, report, and manage their risks, including the risk of litigation and cyber attacks.

Overall, these results indicate our measure is related to characteristics of firms that were successfully attacked (see Kamiya et al., 2020).[14] Therefore, these results provide additional support to the view that our measure captures exposure to cybersecurity risk.

3.5 Firm Outcomes

Finally, we verify that the measure captures cybersecurity-risk exposure by investigating whether it is associated with firm-level outcomes that are consistent with cybersecurity risk. If our measure indeed captures exposure to cybersecurity risk, we would expect a higher likelihood of such an event materializing, which would result in negative stock returns. A

[14] We obtain similar results when using Jaccard similarity as a proxy for cybersecurity risk (see Table IA.1 of Internet Appendix).

negative stock market reaction may occur even in the absence of a cyber-attack directed against a firm's systems and operations; that is, negative returns may occur for high-cybersecurity-risk firms in times of heightened concerns over data breaches for various reasons (e.g., disclosure regulatory changes) and when investors require higher compensation for holding stocks with high exposure to cybersecurity risk. Accordingly, we expect that firms with high cybersecurity risk should have negative asymmetries in stock returns. Therefore, we estimate a linear regression where the dependent variable is the negative coefficient of skewness of weekly returns (*NCSKEW*) (Chen, Hong, and Stein, 2001; Lettau, Maggiori, and Weber, 2014). The main explanatory variable is our cybersecurity-risk measure. Control variables include firm, industry, 10-K, and corporate-governance characteristics. In addition, we include time and industry fixed effects. All continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. In Model 1 of Table 5, the results show our measure positively correlates with *NCSKEW*. As a robustness test, we use extreme sigma as an alternative measure of negative asymmetries in stock returns. Extreme sigma is the negative of the worst deviation of firm-specific weekly returns from the average firm-specific weekly returns divided by the standard deviation of firm-specific weekly returns (*EXTR_SIGMA*) (Andreou, Louca, and Petrou, 2017). In Model 2 of Table A, the results continue to support a positive and statistically significant association between our measure and *EXTR_SIGMA*. In Table IA.2 of the Internet Appendix, we repeat this analysis and obtain similar findings using the cybersecurity-risk measure constructed through Jaccard similarity.

Our next test focuses on whether our measure forecasts future cyber attacks. We estimate a logit regression in which the dependent variable equals 1 if a firm experiences a cyber-attack in a given year, and 0 otherwise. The key explanatory variable is our (one-year) lagged measure of cybersecurity risk. Like before, we control for firm, industry, 10-K, and corporate-governance characteristics, and time and industry fixed effects. All continuous variables are

standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. Table 6 presents the results. In Panel A, the dependent variable is based on all cyber attacks reported in the PRC database for which we have complete data. Model 1 only adds year and industry fixed effects, and Model 2 additionally controls for various firm, industry, 10-K, and corporate-governance characteristics. Furthermore, it includes an indicator variable of whether a firm has been subject to a cyber-attack in the past (*Previous Attack Dummy*), which controls for the fact that the history of past attacks may be a good predictor of future attacks. The results show a positive and statistically significant association between our measure and the probability of experiencing a cyber attack. In terms of economic importance, one standardized unit increase in our measure increases the probability of a cyber attack by 92.70%. These results are not surprising; by construction, our cybersecurity-risk measure is likely to contain forward-looking information. The reason is the cybersecurity-related disclosures, which we use to construct cybersecurity-risk exposure, are legally required to be accurate and current, and thus are likely to reflect the top management's view about exposure to cybersecurity risk. As a result, our measure contains unique information that is incremental relative to both past attacks and other known accounting determinants of cyber attacks, such as firm size or profitability (which tend to be backward looking).

As a robustness test, we repeat the analysis after redefining the dependent variable; rather than using all the cyber attacks reported in PRC database, we focus on major attacks that attracted attention from global news outlets (e.g., CNBC, Financial Times and the Wall Street Journal) or are covered in major Newswires (e.g., AP, Bloomberg, Reuters). Panel B of Table 6 reports the results; our measure retains a positive and statistically significant coefficient, and therefore, it also predicts major future cyber attacks.

Finally, we repeat our analysis after focusing on non-major cyber attacks (those that did not attract attention from major Newswires). Accordingly, in Panel C of Table 6, our dependent

variable takes the value of 1 if a firm experiences a non-major cyber-attack in year *t*, and 0 otherwise. Notably, firms with no experience of major attacks are not included in our training sample used to construct our cybersecurity-risk measure. Therefore, this analysis offers an out-of-sample setting for our predictive exercise. The results of Panel C confirm the predictive ability of our measure for future (non-major) attacks. These findings are further validated via predictive regressions based on Jaccard similarity as a proxy for cybersecurity risk (see Table IA.3 of the Internet Appendix).

4. Cybersecurity Risk and Stock Returns

In the previous section, we show our measure is correlated with both language and real actions, consistent with managing exposure to cybersecurity risk. In addition, it displays intuitive time-series, industry, and firm characteristics that are associated with the probability of cyber attacks. Consistent with these results, our measure is significantly associated with negative asymmetries in stock returns and predicts future cyber attacks. After providing evidence that our measure captures exposure to cybersecurity risk, in this section, we examine whether the stock market prices cybersecurity risk in the cross section of returns.

Specifically, we conjecture that investors may demand compensation for bearing cybersecurity risk; that is, they may require a higher expected return from a firm exposed to high cybersecurity risk. We first use univariate portfolio-level analyses to examine the return performance of firms exposed to high and low cybersecurity risk. Second, we conduct bivariate portfolio sorts to better understand whether exposure to cybersecurity risk is more prevalent in certain subsamples of stocks with different characteristics. Third, we present Fama-McBeth (1973) cross sectional regression results to ensure we are not simply capturing exposure to other well-known risk factors. Finally, we investigate the time-series variation of a cybersecurity-risk factor.

4.1 Univariate Portfolio-level Analysis

We implement the portfolio analysis as follows. We first assign firms into three tercile portfolios according to their exposure to cybersecurity risk. Portfolio 1 includes stocks with the lowest exposure to cybersecurity risk. Given the nature of the data, Portfolio 1 may consist of firms with no cybersecurity-risk disclosures in their 10-Ks. The remaining stocks are then assigned into Portfolio 2 and Portfolio 3 based on the median values of cybersecurity risk.[15] Our objective is to test whether stocks in Portfolio 3 (high-cybersecurity-risk stocks) outperform those in Portfolio 1 portfolio (low-cybersecurity-risk stocks). For our benchmark tests, we start in December 2007 and construct portfolios at the end of each quarter (quarterly rebalancing).[16] We then track the performance of the three portfolios and compute monthly returns in excess of the risk-free rate over the period of March 2008 - March 2019.[17] We calculate both equal-weighted (ew) and value-weighted (vw) monthly portfolio returns. We report average excess portfolio returns as well as portfolio alphas adjusted for market risk (CAPM alphas) or, alternatively, for market, size (SMB), value (HML), and momentum (MOM) factor exposures according to Carhart's (1997) FFC model (FFC alphas), as well as alphas adjusted for market, size and value, profitability (RMW) and investment (CMA) factor exposures according to the Fama and French's (2015) model (five-factor alphas).

The results are presented in Table 7. The average portfolio returns increase from 0.17% to 0.84% from low- to high-cybersecurity-risk stocks for the equal-weighted portfolios, indicating a monthly average difference of 0.67% between the two portfolios. The difference is also statistically significant at the 1% level with a Newey-West *t*-statistic of 4.54.[18] The

[15] We have chosen tercile portfolios for our benchmark analysis so that the three portfolios have a similar number of firms. In a series of robustness tests, we also present results based on quartile, quintile, and decile portfolios (see section 5).
[16] For robustness purposes, we present in section 5 results based on monthly and yearly rebalancing.
[17] Due to data availability (i.e., small number of firms assigned in each portfolio in January 2008 and February 2008), our portfolio analysis starts in March 2008.
[18] We use 12 lags for the calculation of standard errors. Our results are stronger when we use fewer lags, such as six or four.

corresponding return differential is slightly lower for the case of value-weighted returns (0.61% per month), but it remains statistically significant at the 1% level. Controlling for the market, Fama-French-Carhart, and Fama-French (2015) risk factors does not affect our findings. For example, the FFC (five-factor) alpha for the long-short portfolio is 0.69% (0.66%) per month with a *t*-statistic of 4.80 (4.38) for the case of equal-weighted portfolios. The results based on value-weighted returns yield slightly smaller return differences across the two portfolios, but these differences remain both economically and statistically significant (e.g., the five-factor alpha for the long-short portfolio is 0.57% per month with a *t*-statistic of 3.58). Overall, the results imply firms with high cybersecurity risk exhibit higher future excess returns and positive alphas net of well-known risk factors. We further validate the results using the alternative measure of cybersecurity risk, constructed through Jaccard similarity (see Table IA.4 of the Internet Appendix); therefore, the findings are not sensitive to the method used to measure the degree of similarity in cybersecurity-risk disclosures.

Panel B of Table 7 reports the average portfolio characteristics in each cybersecurity-risk portfolio. Specifically, we present information about the number of stocks in each portfolio, well-known stock characteristics, such as size and book-to-market (Fama and French, 1992, 1993), profitability (Fama and French, 2015), institutional ownership (Weber, 2018), illiquidity (Amihud, 2002), idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang, 2006), and 10-K characteristics such as the length of Item 1A. Risk Factor disclosures and the complexity of 10-K disclosures (You and Zhang, 2009; Lehavy, Li, and Merkley, 2011). The results show Portfolio 1 includes a larger number of stocks than Portfolio 3 on average (1233 vs. 966). This is driven by the fact that a non-negligible number of firms have no cybersecurity-risk disclosures in their 10-Ks (Item 1A); by construction, no cybersecurity-risk disclosures imply zero cybersecurity risk. Firms may not report cybersecurity-risk disclosures, because they simply have no such risk concerns. Nevertheless, firms may not report cybersecurity-risk

disclosures, because of (i) low awareness of cybersecurity risk and/or (ii) poor disclosure practises. In section 5, we explicitly address these possibilities through several robustness tests and show our main result is not driven by firms with no cybersecurity-risk disclosures.

The results in Panel B of Table 7 also indicate non-negligible differences between Portfolios 1 and 3 in terms of several firm and 10-K characteristics. Specifically, the average firm in Portfolio 3, which contains high-cybersecurity-risk stocks, is larger in size and exhibits higher profitability, institutional ownership, length of Item 1A, and lower book-to-market, illiquidity, and readability than the average firm in Portfolio 1, which contains low-cybersecurity-risk stocks. These differences, which are also statistically significant between Portfolios 1 and 3, motivate us to conduct bivariate portfolio tests to examine whether the excess returns of high-cybersecurity-risk stocks are confined to subsamples of firms with certain characteristics.

4.2 Bivariate Portfolio-level Analysis

In this section, we use bivariate portfolio sorts. Specifically, starting from December 2007, we sort stocks at the end of each quarter in ascending order on the basis of their cybersecurity risk and allocate them into three groups (low-cyber-risk stocks, middle group and high-cyber-risk stocks), and we also independently sort stocks into ascending order according to several firm- and 10-K-level characteristics. Specifically, we allocate them into two portfolios (low and high) based on median values for each of the following characteristics: market value, book-to-market, return-on-assets (ROA), institutional ownership, illiquidity, idiosyncratic volatility, risk-section length, and readability. The intersection of the above classifications yields several double-sorted portfolios. We track the performance of the intersection portfolios over the following quarter until they are rebalanced, and report results in Table 8. Specifically, we directly report the excess returns of high- versus low-cybersecurity-risk portfolios within each

subsample sorted by another firm characteristic. The outperformance of high-cybersecurity-risk stocks persists in all combinations of stocks and remains statistically significant in the vast majority of cases. These results hold for different measures of excess returns (i.e., average return and five-factor alpha) and both equal- and value-weighted portfolio returns. These results ensure our findings are not contained within a small subsample of stocks, and alleviate concerns that exposure to cybersecurity risk captures other well-known risk proxies.

4.3 Cross-sectional Regressions with Individual Securities

The previous portfolio-level analysis may mask some relevant information: First, controlling for multiple effects jointly is difficult (Freyberger, Neuhierl and Weber, 2020), and second, through portfolio aggregation, it throws away a significant amount of information in the cross section of stock returns. Therefore, we also test the cross sectional relation between cybersecurity risk and subsequent stock returns, using Fama-MacBeth (1973) regressions. For each month of our sample, we run cross sectional regressions of excess stock returns on lagged cybersecurity-risk exposure and a series of characteristics. Specifically, we control for beta, size, and book-to-market (Fama and French, 1992), momentum (Jegadeesh and Titman, 1993), short-term reversal (Jegadeesh, 1990), illiquidity (Amihud, 2002), coskewness (Harvey and Siddique, 2000), idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang, 2006), asset growth and profitability (Fama and French, 2015), and demand for lottery-like stocks (Bali, Cakici, and Whitelaw, 2011). We also control for 10-K characteristics such as the length of Item 1A. Risk Factors and the degree of readability of the 10-K. Table 9 reports the average slope coefficients estimated from these monthly regressions as well as their *t*-statistics computed using Newey-West standard errors. To interpret the economic significance of our findings, all explanatory variables are standardized (demeaned and divided by their standard deviations).

In Model 1, we only include our cybersecurity-risk index in the regressions and find the time-series average of the cross sectional coefficients is 0.30% (with a Newey-West adjusted *t*-statistic of 6.28); therefore, a one standard deviation increase in cybersecurity risk increases returns by 0.30% per month. Model 2 controls for a series of additional stock- firm-level characteristics (as listed above), and in Model 3, we additionally control for the length of Item 1A. Risk Factors and the readability of 10-K filings. The results show the coefficient estimate of cybersecurity risk remains positive and significant, although the magnitude of the cybersecurity risk effect is reduced.

In the last five columns of Table 9 (Models 4 to 8), we assess the long-term (up-to-12 month) predictive power of the cybersecurity-risk proxy. The results show that controlling for all firm characteristics and risk factors, cybersecurity risk predicts monthly cross-sectional variation in stock returns up to 12 months into the future. This finding suggests the predictability is not a short-term phenomenon.

Finally, we also re-estimate the Fama-MacBeth regressions using Jaccard similarity as our key explanatory variable and proxy for cybersecurity risk. The results, which are reported in Table IA.5 of the Internet Appendix, are very similar to the ones reported in Table 9.

4.4 A Cybersecurity-risk Factor and its Time-series Variation

So far, we have documented that stocks exposed more to cybersecurity risk have higher expected returns, and exposure to cybersecurity risk predicts the cross-sectional variation in individual stock returns. To the extent that our measure accurately captures cybersecurity risk, and this is a priced source of risk, then high-cybersecurity-risk stocks should perform poorly and significantly worse than low-cybersecurity-risk stocks on the days of intense attention toward cybersecurity risk. To test whether this conjecture holds, we first form a simple cybersecurity-risk factor using a method similar to the one proposed by Fama and French

(1993). At the end of each month, we sort all stocks into two groups based on market value (using the median market value as a cut-off point). We then independently sort all stocks into three groups based on our cybersecurity-risk measure using the 30th and 70th percentiles as cut-off points. The cybersecurity-risk factor portfolio is calculated as the average return of the two value-weighted high-cybersecurity-risk portfolios minus the average return of the two value-weighted low-cybersecurity-risk portfolios. As Fama and French (1993) note, this ensures the constructed factor captures returns associated with a risk premium (in our case cybersecurity risk) while maintaining neutrality to market capitalization.

For the analysis of the time-series variation of the factor, we resort to daily data and calculate daily returns of the factor over the period March 2008 to March 2019, that is, 2789 daily returns. We are interested in examining the performance of the cybersecurity-risk factor, especially during days of intense attention toward cybersecurity risk. We identify these days based on abnormal search volume index (SVI) in Google Trends. The SVI measures the intensity on "search terms" or "search topics" during a time period, and it is considered a reliable measure of revealed investor attention and demand for information (Drake, Roulsstone, and Thornock, 2012; Da, Engelberg, and Gao, 2011). "Search topics" are a collection of related "search terms"; therefore, we focus on "search topics" because potentially capture attention more comprehensively. We identify the following relevant topics: hacker, data breach, cyber-attack, cyber insurance, cybersecurity, cyber security regulation and hacking. However, not all topics exhibit the same intensity. After comparing them, we find that the average intensity of hacker in our sample period is 19.14 whereas for data breach, it is 15.01; all the remaining topics exhibit substantially less intensity. As a result, to gauge when investors have increasing concerns over cybersecurity risk, we use the topics hacker and data breach.

More specifically, we estimate the following regression model:

$$CRF_t = a + \beta \times High\_Google\_SVI\_dummy_t + \gamma_i \times \boldsymbol{X}_t + error, \quad [3]$$

where CRF is the cybersecurity-risk factor, "*High_Google_SVI_dummy*" is a dummy variable that takes the value of 1 on days with high SVI, and 0 otherwise, and *X* is a vector of commonly used (daily) risk factors, namely, market, size, value, momentum, operating profitability, and investment factors (see Carhart, 1997; Fama and French, 2015). We construct the variable *"High_Google_SVI_dummy"* as follows: we first download the monthly SVI for our sample period and take its first difference.[19] We then identify the months with the highest attention on cybersecurity risk (top decile of the distribution), and through another Google Trends search for each of these months, we identify the day with the highest SVI.[20] We present results using the highest SVIs identified using both hacker and data breach topics. However, the results are robust if we consider independently the highest SVIs identified by hacker or data breach topics. Our event window covers the period [0,+1].

We estimate equation [3] using alternative models and present the results in Table 10. Models 1 includes *High_Google_SVI_dummy* as the only explanatory variable. Model 2 controls for the market risk factor (CAPM specification); in Model 3, we add the size, value and momentum factors (FFC specification), and Model 4 controls for the five risk factors proposed by Fama and French (2015) (Fama-French five-factor specification). Based on the findings, the cybersecurity-risk factor exhibits, on average, positive returns over the sample period; the daily estimate for the constant term α is positive (at 0.0002, which implies an annualized return above 5% per year) and is statistically significant (at the 1% level) in all models. Importantly, the estimate for β is consistently negative and statistically significant, which suggests the cybersecurity-risk factor exhibits negative returns on days with major concerns on cybersecurity risk. These results suggest firms with high exposure to cybersecurity risk earn high returns on average, but they perform poorly when concerns about cybersecurity

[19] Google Search Trends provide daily data only for query period shorter than nine months.
[20] For non-trading days with the highest monthly SVI, we use the first subsequent trading day.

risk are heightened. The premium that high-cybersecurity-risk stocks earn compensates risk-averse investors for holding high-cybersecurity-risk stocks, which significantly underperform in times of heightened cybersecurity risk and investors' concerns about data breaches.

Finally, in Panels B and C, we re-estimate equation [3] after replacing the variable *High_Google_SVI_dummy* with the variable *High_Google_SVI_dummy + 5 days (High_Google_SVI_dummy + 1 month)*, which moves the event window a trading week (month) after the actual peak of the SVI index. For these placebo events, we find no evidence of underperformance of the cybersecurity-risk factor, ensuring we are not capturing any other events coincidentally close in time.

5. Portfolio Analysis: Robustness Tests

This section presents several robustness tests. First, we check whether the outperformance of stocks exposed to high cybersecurity risk is more pronounced in the latter period of our sample, in particular, after SEC's 2011 guidance for public-disclosure obligations with respect to cybersecurity risk and cyber incidents. Consistent with the view that concerns over cybersecurity risk have increased in the post-2011 period, the results in Panel A of Table 11 show the excess return and five-factor alpha for the long-short portfolio are higher than those reported in Table 7 for the entire period (i.e., the five-factor alpha increases to 0.65% - up from 0.57%- for the case of value-weighted portfolios).

In Panel B of Table 11, we assess the outperformance of high-cybersecurity-risk stocks after excluding from the sample all firms that use cyber insurance as a form of (partial) protection against cybersecurity risk. As mentioned above, we identify these firms by searching for the word "insurance" in their cybersecurity-risk disclosures and looking at their relevant discussions. By construction, these firms are more likely to be classified as high-cybersecurity-risk firms (P3) than low-cybersecurity-risk firms (P1) (i.e. P1 only includes firms with no cybersecurity-related disclosures in the early years of our sample). Such classification may be

problematic because these firms are at least partially protected against claims that may arise due to cyber attacks, which suggests investors should be less concerned about their exposure to cybersecurity risk. Consistent with this reasoning, we find the performance of the long-short portfolio increases after the exclusion of firms with cyber insurance from our sample (i.e., the five-factor alpha increases to 0.65% - up from 0.57% - for the case of value-weighted portfolios).

In Panel C, we perform a similar exercise after excluding from the analysis all firms that experienced major attacks and that we used for the construction of our cybersecurity-risk measure; that is, all firms in the training sample. These are, by construction, high-cybersecurity-risk firms and their exclusion has a direct effect on the composition of the portfolio with the highest-cybersecurity-risk stocks (P3). Indeed, we observe a decline in the spread of the long-short portfolio, especially for the case of value-weighted returns. Nevertheless, the documented premium still remains robust and statistically significant (at 1% for the case of equal-weighted portfolios and 5% for the case of value-weighted portfolios). This result suggests our findings are not driven by firms that experienced major cyber attacks.

In Panel D, we repeat our analysis for monthly and annual rebalancing of our portfolios. Once again, our results are robust and very similar to those reported in Table 7.

We then check whether the outperformance of stocks exposed to high cybersecurity risk is driven by certain industries. To do so, we repeat the portfolio analysis (as in section 4.1) 12 times after excluding each industry at a time, to remove any potential abnormal impact of a particular industry group. Panel E of Table 11 presents the estimates on the performance (both in terms of average return and five-factor alpha) of the spread strategy, that is, long the portfolio with the highest-cybersecurity-risk stocks (P3) and short the portfolio with the lowest-cybersecurity-risk stocks (P1). In all cases, we find a positive and statistically significant

premium of high-cybersecurity-risk stocks, both in terms of equal-weighted and value-weighted returns. Therefore, the results are not driven by any particular industry.

Then, we deal with the fact that a non-negligible number of firms in the sample have no cybersecurity-risk related disclosures in their 10-Ks (Item 1A). This feature of the data is concentrated in the early years of the sample (i.e., 2008-2011) and results in the assignment of zero cybersecurity risk for such firm-years. Given that in the earlier years of the sample, cybersecurity risk was arguably not so prevalent, we can assume that these firms have indeed relatively low levels of cybersecurity risk. However, a non-cybersecurity-risk disclosure may also be driven by (i) low awareness of cybersecurity risk and/or (ii) poor disclosure practices. The subperiod analysis in Panel A helps (partly) in dealing with this potential measurement problem in the cybersecurity-risk measure, because the non-disclosure problem largely disappears in the January 2012-March 2019 period. As an additional test, we replace all firm-year observations of cybersecurity risk with zero values, with the median industry value in the corresponding year. To capture risk exposure as accurately as possible, we use four-digit SIC codes for the industry classification. The results, as reported in Panel F of Table 11, are qualitatively similar to those based on the original cybersecurity-risk measure. As our final test, we assume a firm's exposure to cybersecurity risk is likely to be persistent across time, and hence backfilled all zeros in the measure with the first available non-zero observation of each firm. A complication, however, with this approach is that, on average, the cybersecurity risk increases across time; therefore, given the non-disclosure problem is concentrated in the early years of the sample, backfilling cybersecurity risk artificially "inflates" the exposure to cybersecurity risk for firms that do not report cybersecurity-risk disclosures in a certain year relative to firms that do. Nevertheless, as shown in Panel G, the results remain largely unaffected; although the spread is lower, reflecting perhaps the noisier measure of cybersecurity risk. Interestingly, when we focus on more "extreme" portfolios to calculate the

spread (i.e., quartile, quintile and decile portfolios), the return spread increases in magnitude (e.g., 0.53% per month using a five-factor alpha for decile portfolios, up from 0.29% per month for tercile portfolios). Overall, these results, along with the fact the more extreme portfolio classifications help distinguish more clearly between low- and high-cybersecurity-risk stocks, lead to the conclusion that the results are not driven by firm-years with no cyber-related disclosures in 10-Ks.

6. Conclusions

We construct a novel firm-level measure of cybersecurity risk using textual analysis of cybersecurity-risk disclosures in Item 1A. Risk Factors in 10-K statements and use it to examine whether cybersecurity risk is priced in the cross section of stock returns. We show that the measure successfully identifies firms that discuss risk extensively, and that it displays intuitive relations with quantitative measures based on cybersecurity-risk-disclosure language. In addition, the measure displays interesting time-series and cross sectional characteristics. For instance, it exhibits a positive trend over time, and it is more prevalent among industries that rely on information technology systems. We also find the measure correlates with several characteristics linked to firms hit by cyber attacks, such as size, age, growth opportunities, asset tangibility, R&D expenditures, and the presence of trade secrets. Finally, we find the measure is positively associated with (negative) asymmetries in stock returns and it also predicts the probability of experiencing a future cyber attack. Overall, these results support the view that our measure captures exposure to cybersecurity risk.

In financial markets, cybersecurity risk is priced in the cross section of stock returns. Specifically, a portfolio long on firms with high-cybersecurity-risk and short on low-cybersecurity-risk stocks earns a statistically significant 66-69 (56-66) basis points per month - up to 8.3% (7.9%) - in equal-weighted (value-weighted) returns over the following year.

Fama-MacBeth cross sectional regressions confirm a positive and statistically significant association between future individual stock returns and our cybersecurity-risk measure. A factor-mimicking portfolio calculated as the difference in the return of stocks with high and low cybersecurity risk performs poorly around periods of increasing investor attention to cybersecurity risk but earns a high premium during other times. These results support the predictions of asset-pricing theory that investors require compensation for bearing cybersecurity risk.

Our study opens several avenues for future research. The cybersecurity-risk measure and its underlying methodology, which is transparent, easily implementable, and comprehensively covers the population of US firms that file 10-K reports in Edgar, enables a systematic analysis on cybersecurity risk and its implications for firm value, corporate policies, and firm operations.

References

Akey, P., Lewellen, S., Liskovich, I., & Schiller, C. (2020) Hacking corporate reputations. Rotman School of Management Working Paper No. 3143740, Available at SSRN: https://ssrn.com/abstract=3143740.

Amihud, Y. (2002) Illiquidity and stock returns: cross section and time-series effects. *Journal of Financial Markets 5:*31-56.

Amir, E., Levi, S., & Livne, T. (2018) Do firms underreport information on cyber attacks? Evidence from capital markets. *Review of Accounting Studies* 23:1177-1206.

Andreou, P. C., Louca, C., & Petrou, A. P. (2017) CEO age and stock price crash risk. *Review of Finance* 21:1287-1325.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006) The cross-section of volatility and expected returns. *Journal of Finance* 61:259-299.

Baker, S.R., Bloom, N., & Davis, S.J. (2016) Measuring economic policy uncertainty. *Quarterly Journal of Economics* 131:1593-1636.

Bali, T.G., Cakici, N., & Whitelaw, R.F. (2011) Maxing out: Stocks as lotteries and the cross section of expected returns. *Journal of Financial Economics* 99:427-446.

Bouveret, A. (2018) Cyber risk for the financial sector: A framework for quantitative assessment. *IMF Working paper* 143. Available at SSRN: https://ssrn.com/abstract=3203026.

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

Buehlmaier, M.M., & Whited, T.M. (2018) Are financial constraints priced? Evidence from textual analysis. *Review of Financial Studies* 31:2693-2728.

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

Carhart, M. (1997) On the persistence in mutual fund performance. *Journal of Finance* 52:57-82.

Chen, J., Hong, H., & Stein, J.C. (2001) Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61:345-381.

Cohen, L., Malloy, C., & Nguyen, Q. (2020) Lazy prices. *Journal of Finance* 70: 1371-1415.

Da, Z., Engelberg, J., & Gao, P. (2011) In search of attention. *Journal of Finance* 66:1461-1499.

Drake, M.S., Roulstone, D.T., & Thornock, J.R. (2012) Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research* 50:1001-1040.

Fama, E.F., & French, K.R. (1992) The cross-section of expected stock returns. *Journal of Finance* 47:427-465.

Fama, E.F., & French, K.R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3-56.

Fama E.F., & French, K.R. (2015) A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

Fama, E.F., & MacBeth, J. D. (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607-636.

Freyberger, J., Neuhierl, A., & Weber, M. (2020) Dissecting characteristics non-parametrically. *Review of Financial Studies* 33:2326-2377.

Frésard, L., Hoberg, G., & Phillips, G.M. (2020) Innovation activities and integration through vertical acquisitions. *Review of Financial Studies* 33:2937-2976.

Gentzkow, M., Kelly, B., & Taddy, M. (2019) Text as data. *Journal of Economic Literature* 57:535-74.
Hanley, K.W., & Hoberg, G. (2010) The information content of IPO prospectuses. *Review of Financial Studies* 23:2821-2864.
Harvey, C.R., & Siddique, A. (2000) Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263-1295.
Hassan, T.A., Hollander, S., van Lent, L., & Tahoun, A. (2019) Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics* 134: 2135-2202.
Campbell, J.Y., & Hentschel, L. (1992) No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31:281-318.
Hilary, G., Segal, B., & Zhang, M. (2016) Cyber-risk disclosure: Who cares? Georgetown McDonough School of Business Research Paper No. 2852519.
Hoberg, G., & Maksimovic, V. (2015) Redefining financial constraints: A text-based analysis. *Review of Financial Studies* 28:1312-1352.
Hoberg, G., & Phillips, G. (2010) Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23:3773-3811.
Hoberg, G., & Phillips, G. (2016) Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423-1465.
Jegadeesh, N. (1990) Evidence of predictable behavior of security returns. *Journal of Finance* 45:881-898.
Jegadeesh, N., & Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65-91.
Johnson, M.S., Kang, M.J. & Lawson, T. (2017) Stock price reaction to data breaches. *Journal of Finance Issues* 16:1-13.
Kamiya, S., Kang, J. K., Kim, J., Milidonis, A., & Stulz, R. M. (2020) Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*. Forthcoming.
Lang, M., & Stice-Lawrence, L. (2015) Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60:110-135.
Lehavy, R., Li, F., & Merkley, K. (2011) The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review* 86:1087-1115.
Lending, C., Minnick, C., & Schorno, P.J. (2018) Corporate governance, social responsibility and data breaches. *Financial Review* 53:413-455.
Lettau, M., Maggiori, M., & Weber, M. (2014) Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics* 114:197-225.
Loughran, T., & McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66:35-65.
Loughran, T., & McDonald, B. (2016) Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54:1187-1230.
Lowry, M., Michaely, R., & Volkova, E. (2020). Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO. *Review of Financial Studies* 33: 5510–5554.
Romanosky, S. (2016) Examining the costs and causes of cyber incidents. *Journal of Cybersecurity* 2:121-135.
Sautner, Z., van Lent, L., Vilkov, G. & Zhang, R. (2020) Firm-level climate change exposure. European Corporate Governance Institute – Finance Working Paper No. 686/2020.
Securities and Exchange Commission. (2011) CF disclosure guidance: Topic no. 2: Cybersecurity.
Securities and Exchange Commission. (2018) Commission statement and guidance on public company cybersecurity disclosures.

Tosun, O.K. (2020) Cyber attacks and stock market activity. Available at SSRN: https://ssrn.com/abstract=3190454.
Weber, M. (2018) Cash flow duration and the term structure of equity returns. *Journal of Financial Economics* 188:486-503.
You, H., & Zhang, X. J. (2009) Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14:559-586.

Appendix A

A1: Extracting Cybersecurity-risk Disclosures

Based on our reading of 500 randomly 10-K files, the relevant cybersecurity risk discussion is usually presented separately within certain paragraphs; each paragraph contains a title (in bold or italics) followed by the relevant discussion. The title/relevant discussion often contains a direct description of cybersecurity risk. For instance, the title of the relevant discussion in Apple Inc 10-Ks for fiscal year 2017 is "*There may be losses or **unauthorized access** to or releases of confidential information, including personally identifiable information, that could subject the Company to significant reputational, financial, legal and operational consequences.*". In general, firms describe the nature of their business, how/why a firm's business is exposed to cybersecurity risk, potential changes in exposure, and efforts to establish or improve security measures which mitigate cybersecurity risk. In addition, in line with the regulatory concept of "*material*", firms also provide information about internal/legal/economic consequences that may arise from cybersecurity risk. Among others, internal consequences include theft or misuse of assets, intellectual property, data and information that may arise from potential cyber attacks; legal consequences e.g. the loss of confidential information could subject the company to significant legal consequences; and finally, economic consequences i.e. information about how cybersecurity risk may affect their businesses; in particular operations, competitive positioning, reputation etc..

Below, we provide common keywords/phrases that companies use in their direct descriptions of cybersecurity risk.[21] Our algorithm is not case sensitive; thus, it avoids missing relevant keywords/phrases. In addition, to alleviate issues related to language expression, it captures all the words that "start with" the relevant keyword. For example, with the keyword attack the algorithm searches also for attacks, attacking, attacked etc. While some keywords/phrases, such as hackers clearly describe exposure to cybersecurity risk, others such as attacks may also be considered in different settings (e.g. terrorist attacks). We overcome this challenge as follows: when we have a relevant keyword/phrase that may also be used in different settings we require (i) the presence of an additional relevant hit within the same sentence and (ii) the absence of an additional irrelevant hit within the same sentence. For instance, when we find the keyword "attack" in a sentence we also require the presence of the keyword "cyber" and the absence of the keyword "terrorist".

[21] The compilation of keywords/phrases is based on (i) cybersecurity risk glossaries such as https://www.threatconnect.com/cyber-security-glossary/ and (ii) the language that firm's use to describe cybersecurity risk in the 10-Ks.

In addition, we noticed that firms may also use indirect description that may relate to cybersecurity risk. For instance, in Apple Inc 10-K for fiscal year 2017 it writes "*The Company's business requires it to use and store confidential information, including, among other things, personally identifiable information ("PII") with respect to the Company's customers and employees.*" This sentence does not contain any direct keywords/phrases of cybersecurity risk. However, it is part of the cybersecurity risk discussion as it is immediately after the title of the paragraph "*There may be losses or **unauthorized access** to or releases of confidential information, including personally identifiable information, that could subject the Company to significant reputational, financial, legal and operational consequences.*" and it is followed by the "*The Company devotes significant resources to network and data security, including through the use of encryption and other security measures intended to protect its systems and data.*" Therefore, to capture such indirect description of cybersecurity risk we create another list of indirect keywords/phrases.[22] Below we provide the list with the keywords/phrases, which the companies use in their indirect descriptions for cybersecurity risk. To ensure that our algorithm retrieves only relevant to cybersecurity risk sentences, we require first, to identify a sentence with a direct cybersecurity risk discussion. Then, we search the subsequent 10 sentences to find indirect keywords/phrases. Because the discussion is often clustered in a paragraph, it is reasonable to assume that indirect keywords/phrases are tagged to cybersecurity risk. While this approach is very successful, we noticed that occasionally it may also be noisy as it may capture discussion from the subsequent risk factor description. We reduce this noise by exploiting the presence of title fonts (bold or italics) in the subsequent risk factor to end the search; thus, we search until we find a subsequent sentence in bold or italics – if we don't find any such sentence we search up to 10 subsequent sentences.

Finally, we provide below examples on how successful the algorithm is in extracting/missing relevant sentences from the 10-Ks of Apple Inc, Abbott Laboratories, General Motors Co, and Verizon Communications Inc for the fiscal year 2017. We display sentences that the algorithm retrieves from "relevant paragraphs" (i.e., when the focus is on cybersecurity risk) and from "other paragraphs" (i.e., when the focus is not on cybersecurity risk).

[22] The compilation of keywords/phrases is based on the structure of the most comprehensive discussions of cybersecurity risk in the 10-Ks and includes descriptions of (i) company business, (ii) internal consequences, (iii) legal consequences, and (iv) economic consequences.

Keywords/Phrases

	Relevant hit if	Irrelevant hit if
1. Direct description of cybersecurity risk		
Attack	Cyber-, cyber, networks, systems, products, services, datacenter, infrastructure	Terror, war, contraband, bombs
Threat	Cyber-, cyber, networks, systems, products, services, datacenter, infrastructure	Terror, simulator, disease, legal action, competitive, competitors, substitute, patent, nuclear, life, threaten/ed
Computer, information system	Malware, virus, viruses, intrusions	
Malicious	Software, programs, third parties, attacks	fires, product sales, warranty claim/s
Breaches		Fiduciary duty/duties, covenant/s, credit, agreement/s, warranty, warranties, obligations, regulations, contract/s, resolution
Hacker, hacking, social engineering, denial of service, denial-of-service, phishing, cyber-attack, cyber attacks, cyber risk, cyber security, cybersecurity, cyber intrusions, unauthorized access, unauthorized disclosure, breach in security, security breach		
2. Indirect description of cybersecurity risk		
2.1 Company business		
Company, regular course	Business, operation, services	
Technology, technologies	Computer, information, communication, proprietary, infrastructure, reliance, digital, advances	
Information	Network, services, systems, confidential, proprietary, account	
Electronic	Network, services, systems, information	
Computer, telecommunication, third-party, infrastructure	Systems, networks, facilities	
Collect, store, transmit, retrieve, sensitive, critical, protection	Data, information	
IT environment, IT systems, operational systems, communication systems, critical infrastructure		
Security	Network, products, services, systems, devices, data, infrastructure, patches, cloud, web, email, vulnerabilities, threat, breach, penetrate, bypass, compromised, incidence, incident, circumvent, measures, portfolio, solutions, practices, standards	
Vulnerabilities	Network, products, services, systems, devices, data, infrastructure, claims	

2.2 Internal consequences	
Integrity, reliability, protect, protection, protecting, prevent, prevention, preventing, monitors, compromise, secure, failure	Network, products, services, systems, data, measures, information
Gain access	Network, systems, data, datacenter
Access, accessed, modified	Improper, improperly,
Theft, misuse, misusing, modification, destruction, lost, loss, stolen, steal, disclose, publicly disclosed	Assets, intellectual property, data, information
Investigate, remediate, remediation, recover, repair, replace	Network, products, services, systems, data, measures, efforts
Interruptions, disruptions, delays	Network, services, system
Degrade the user experience, invasion, user names, password, break-ins, terminated agreements	
2.3 Legal consequences	
Legal	Claims, actions, challenges, liability
Legislative	Actions
Regulatory	Actions, investigations, agencies
Liability	Claims
Lawsuits, litigation	
2.4 Economic consequences	
Business	Adversely, material, harm disruptive, negative
Operations, services	Disrupt
Revenues	Reduce, adversely, loss, lose
Cost	Increase, increasing, remedy
Operating results, operating margin	Harm, diminish, reduce
Earnings	Reduce, adversely
Financial	Harm, diminish, adversely, material, damage, negative
Competitive position	Harm, diminish
Reputation	Harm, damage, loss, adverse
Brand	Harm, damage

Appendix A (continued)

A2: Examples of Algorithm Extraction Ability

Number of Sentence	*Sentence as in Company's 10-K (Item 1A.Risk Factors)*	*Sentence captured (Yes/No)*	*Sentence Type*
Apple Inc (Fiscal year ended September 30, 2017)			
Text from the relevant paragraph:			
1	There may be losses or unauthorized access to or releases of confidential information, including personally identifiable information, that could subject the Company to significant reputational, financial, legal and operational consequences.	Yes	Direct: Description of Cybersecurity Risk
2	The Company's business requires it to use and store confidential information, including, among other things, personally identifiable information ("PII") with respect to the Company's customers and employees.	Yes	Indirect: Description of Company Business
3	The Company devotes significant resources to network and data security, including through the use of encryption and other security measures intended to protect its systems and data.	Yes	Indirect: Description of Company Business
4	But these measures cannot provide absolute security, and losses or unauthorized access to or releases of confidential information may still occur, which could materially adversely affect the Company's reputation, financial condition and operating results.	Yes	Direct: Description of Cybersecurity Risk
5	The Company's business also requires it to share confidential information with suppliers and other third parties.	Yes	Indirect: Description of Company Business
6	Although the Company takes steps to secure confidential information that is provided to third parties, such measures may not be effective and losses or unauthorized access to or releases of confidential information may still occur, which could materially adversely affect the Company's reputation, financial condition and operating results.	Yes	Direct: Description of Cybersecurity Risk
7	For example, the Company may experience a security breach impacting the Company's information technology systems that compromises the confidentiality, integrity or availability of confidential information.	Yes	Indirect: Description of Company Business
8	Such an incident could, among other things, impair the Company's ability to attract and retain customers for its products and services, impact the Company's stock price, materially damage supplier relationships, and expose the Company to litigation or government investigations, which could result in penalties, fines or judgments against the Company.	Yes	Indirect: Description of Legal Consequences
9	Although malicious attacks perpetrated to gain access to confidential information, including PII, affect many companies across various industries, the Company is at a relatively greater risk of being targeted because of its high profile and the value of the confidential information it creates, owns, manages, stores and processes.	Yes	Direct: Description of Cybersecurity Risk
10	The Company has implemented systems and processes intended to secure its information technology systems and prevent unauthorized access to or loss of sensitive data, including through the use of encryption and authentication technologies.	Yes	Direct: Description of Cybersecurity Risk
11	As with all companies, these security measures may not be sufficient for all eventualities and may be vulnerable to hacking, employee error, malfeasance, system error, faulty password management or other irregularities.	Yes	Direct: Description of Cybersecurity Risk
12	For example, third parties may attempt to fraudulently induce employees or customers into disclosing user names, passwords or other sensitive information, which may in turn be used to access the Company's information technology systems.	Yes	Indirect: Description of Company Business
13	To help protect customers and the Company, the Company monitors its services and systems for unusual activity and may freeze accounts under suspicious circumstances, which, among other things, may result in the delay or loss of customer orders or impede customer access to the Company's products and services.	Yes	Indirect: Description of Company Business
14	In addition to the risks relating to general confidential information described above, the Company may also be subject to specific obligations relating to health data and payment card data.	Yes	Indirect: Description of Company Business
15	Health data may be subject to additional privacy, security and breach notification requirements, and the Company may be subject to audit by	Yes	Indirect: Description of Company Business

	governmental authorities regarding the Company's compliance with these obligations.		
16	If the Company fails to adequately comply with these rules and requirements, or if health data is handled in a manner not permitted by law or under the Company's agreements with healthcare institutions, the Company could be subject to litigation or government investigations, may be liable for associated investigatory expenses, and could also incur significant fees or fines.	Yes	Indirect: Description of Legal Consequences
17	Under payment card rules and obligations, if cardholder information is potentially compromised, the Company could be liable for associated investigatory expenses and could also incur significant fees or fines if the Company fails to follow payment card industry data security standards.	Yes	Indirect: Description of Internal Consequences
18	The Company could also experience a significant increase in payment card transaction costs or lose the ability to process payment cards if it fails to follow payment card industry data security standards, which would materially adversely affect the Company's reputation, financial condition and operating results.	Yes	Indirect: Description of Economic Consequences
19	While the Company maintains insurance coverage that is intended to address certain aspects of data security risks, such insurance coverage may be insufficient to cover all losses or all types of claims that may arise.	Yes	Indirect: Description of Company Business

Relevant paragraph algorithm accuracy: The algorithm successfully extracted 19/19 sentences or 100% of the total number of sentences.

Text from other paragraphs (outside Item 1A. Risk Factors):

1	The Company may be subject to information technology system failures or network disruptions caused by natural disasters, accidents, power disruptions, telecommunications failures, acts of terrorism or war, computer viruses, physical or electronic break-ins, or other events or disruptions.	Yes	Direct: Description of Cybersecurity Risk
2	System redundancy and other continuity measures may be ineffective or inadequate, and the Company's business continuity and disaster recovery planning may not be sufficient for all eventualities.	Yes	Indirect: Description of Company Business
3	Such failures or disruptions could adversely impact the Company's business by, among other things, preventing access to the Company's online services, interfering with customer transactions or impeding the manufacturing and shipping of the Company's products.	Yes	Indirect: Description of Company Business
4	These events could materially adversely affect the Company's reputation, financial condition and operating results.	Yes	Indirect: Description of Economic Consequences

---

Abbott Laboratories (Fiscal year ended December 31, 2017)

Text from the relevant paragraph:

1	Abbott depends on sophisticated information technology systems and a cyber attack or other breach of these systems could have a material adverse effect on Abbott's results of operations.	Yes	Direct: Description of Cybersecurity Risk
2	Similar to other large multi-national companies, the size and complexity of the information technology systems on which Abbott relies for both its infrastructure and products makes them susceptible to a cyber attack, malicious intrusion, breakdown, destruction, loss of data privacy, or other significant disruption.	Yes	Direct: Description of Cybersecurity Risk
3	These systems have been and are expected to continue to be the target of malware and other cyber attacks.	Yes	Direct: Description of Cybersecurity Risk
4	In addition, third party hacking attempts may cause Abbott's information technology systems and related products, protected data, or proprietary information to be compromised.	Yes	Direct: Description of Cybersecurity Risk
5	A significant attack or other disruption could result in adverse consequences, including increased costs and expenses, problems with product functionality, damage to customer relations, lost revenue, and legal or regulatory penalties.	Yes	Direct: Description of Cybersecurity Risk
6	Abbott invests in its systems and technology and in the protection of its products and data to reduce the risk of an attack or other significant disruption, and monitors its systems on an ongoing basis for any current or potential threats and for changes in technology and the regulatory environment.	Yes	Direct: Description of Cybersecurity Risk

7	There can be no assurance that these measures and efforts will prevent future attacks or other significant disruptions to any of the systems on which Abbott relies or that related product issues will not arise in the future.	Yes	Direct: Description of Cybersecurity Risk
8	Any significant attack or other disruption on Abbott's systems or products could have a material adverse effect on Abbott's business.	Yes	Direct: Description of Cybersecurity Risk

Relevant paragraph algorithm accuracy: The algorithm successfully extracted 8/8 sentences or 100% of the total number of sentences.

Text from other paragraphs (outside Item 1A. Risk Factors): None

---

General Motors Co (Fiscal year ended December 31, 2017)

Text from the relevant paragraph:

1	Security breaches and other disruptions to information technology systems and networked products, including connected vehicles, owned or maintained by us, GM Financial, or third-party vendors or suppliers on our behalf, could interfere with our operations and could compromise the confidentiality of private customer data or our proprietary information.	Yes	Direct: Description of Cybersecurity Risk
2	We rely upon information technology systems and manufacture networked products, some of which are managed by third-parties, to process, transmit and store electronic information, and to manage or support a variety of our business processes, activities and products.	Yes	Indirect: Description of Company Business
3	Additionally, we and GM Financial collect and store sensitive data, including intellectual property, proprietary business information, proprietary business information of our dealers and suppliers, as well as personally identifiable information of our customers and employees, in data centers and on information technology networks.	Yes	Indirect: Description of Company Business
4	The secure operation of these systems and products, and the processing and maintenance of the information processed by these systems and products, is critical to our business operations and strategy.	Yes	Indirect: Description of Company Business
5	Despite security measures and business continuity plans, these systems and products may be vulnerable to damage, disruptions or shutdowns caused by attacks by hackers, computer viruses, or breaches due to errors or malfeasance by employees, contractors and others who have access to these systems and products.	Yes	Direct: Description of Cybersecurity Risk
6	The occurrence of any of these events could compromise the operational integrity of these systems and products.	Yes	Indirect: Description of Internal Consequences
7	Similarly, such an occurrence could result in the compromise or loss of the information processed by these systems and products.	Yes	Indirect: Description of Company Business
8	Such events could result in, among other things, the loss of proprietary data, interruptions or delays in our business operations and damage to our reputation.	Yes	Indirect: Description of Internal Consequences
9	In addition, such events could result in legal claims or proceedings, liability or regulatory penalties under laws protecting the privacy of personal information; disrupt operations; or reduce the competitive advantage we hope to derive from our investment in advanced technologies.	Yes	Indirect: Description of Company Business
10	We have experienced such events in the past and, although past events were immaterial, future events may occur and may be material.	No	
11	Portions of our information technology systems also may experience interruptions, delays or cessations of service or produce errors due to regular maintenance efforts, such as systems integration or migration work that takes place from time to time.	Yes	Indirect: Description of Company Business
12	We may not be successful in implementing new systems and transitioning data, which could cause business disruptions and be more expensive, time-consuming, disruptive and resource intensive.	Yes	Indirect: Description of Internal Consequences

13	Such disruptions could adversely impact our ability to design, manufacture and sell products and services, and interrupt other business processes.	Yes	Indirect: Description of Internal Consequences
14	Security breaches and other disruptions of our in-vehicle systems could impact the safety of our customers and reduce confidence in GM and our products.	Yes	Direct: Description of Cybersecurity Risk
15	Our vehicles contain complex information technology systems.	Yes	Indirect: Description of Company Business
16	These systems control various vehicle functions including engine, transmission, safety, steering, navigation, acceleration, braking, window and door lock functions.	No	
17	We have designed, implemented and tested security measures intended to prevent unauthorized access to these systems.	Yes	Indirect: Description of Company Business
18	However, hackers have reportedly attempted, and may attempt in the future, to gain unauthorized access to modify, alter and use such systems to gain control of, or to change, our vehicles' functionality, user interface and performance characteristics, or to gain access to data stored in or generated by the vehicle.	Yes	Direct: Description of Cybersecurity Risk
19	Any unauthorized access to or control of our vehicles or their systems or any loss of data could impact the safety of our customers or result in legal claims or proceedings, liability or regulatory penalties.	Yes	Direct: Description of Cybersecurity Risk
20	In addition, regardless of their veracity, reports of unauthorized access to our vehicles, their systems or data could negatively affect our brand and harm our business, prospects, financial condition and operating results.	Yes	Direct: Description of Cybersecurity Risk

Relevant paragraph algorithm accuracy: The algorithm successfully extracted 18/20 sentences or 90.00% of the total number of sentences.

Text from other paragraphs (outside Item 1A. Risk Factors):

1	We sometimes face attempts to gain unauthorized access to our information technology networks and systems for the purpose of improperly acquiring our trade secrets or confidential business information.	Yes	Direct: Description of Cybersecurity Risk
2	The theft or unauthorized use or publication of our trade secrets and other confidential business information as a result of such an incident could adversely affect our competitive position.	Yes	Indirect: Description of Company Business

---

Verizon Communications Inc (Fiscal year ended December 31, 2017)

Text from the relevant paragraph:

1	Cyber attacks impacting our networks or systems could have an adverse effect on our business.	Yes	Direct: Description of Cybersecurity Risk
2	Cyber attacks, including through the use of malware, computer viruses, dedicated denial of services attacks, credential harvesting and other means for obtaining unauthorized access to or disrupting the operation of our networks and systems and those of our suppliers, vendors and other service providers, could have an adverse effect on our business.	Yes	Direct: Description of Cybersecurity Risk
3	Cyber attacks may cause equipment failures, loss of information, including sensitive personal information of customers or employees or valuable technical and marketing information, as well as disruptions to our or our customers' operations.	Yes	Direct: Description of Cybersecurity Risk
4	Cyber attacks against companies, including Verizon, have increased in frequency, scope and potential harm in recent years.	Yes	Direct: Description of Cybersecurity Risk
5	Further, the perpetrators of cyber attacks are not restricted to particular groups or persons.	Yes	Direct: Description of Cybersecurity Risk
6	These attacks may be committed by company employees or external actors operating in any geography, including jurisdictions where law enforcement measures to address such attacks are unavailable or ineffective, and may even be launched by or at the behest of nation states.	No	
7	Cyber attacks may occur alone or in conjunction with physical attacks, especially where disruption of service is an objective of the attacker.	Yes	Direct: Description of Cybersecurity Risk

8	While, to date, we have not been subject to cyber attacks which, individually or in the aggregate, have been material to our operations or financial condition, the preventive actions we take to reduce the risks associated with cyber attacks, including protection of our systems and networks, may be insufficient to repel or mitigate the effects of a major cyber attack in the future.	Yes	Direct: Description of Cybersecurity Risk
9	The inability to operate our networks and systems or those of our suppliers, vendors and other service providers as a result of cyber attacks, even for a limited period of time, may result in significant expenses to Verizon and/or a loss of market share to other communications providers.	Yes	Direct: Description of Cybersecurity Risk
10	The costs associated with a major cyber attack on Verizon could include expensive incentives offered to existing customers and business partners to retain their business, increased expenditures on cybersecurity measures and the use of alternate resources, lost revenues from business interruption and litigation.	Yes	Direct: Description of Cybersecurity Risk
11	The potential costs associated with these attacks could exceed the insurance coverage we maintain.	No	
12	Further, certain of Verizon's businesses, such as those offering security solutions and infrastructure and cloud services to business customers, could be negatively affected if our ability to protect our own networks and systems is called into question as a result of a cyber attack.	Yes	Direct: Description of Cybersecurity Risk
13	Moreover, our increasing presence in the IoT industry with offerings of telematics products and services, including vehicle telematics, could also increase our exposure to potential costs and expenses and reputational harm in the event of cyber attacks impacting these products or services.	Yes	Direct: Description of Cybersecurity Risk
14	In addition, a compromise of security or a theft or other compromise of valuable information, such as financial data and sensitive or private personal information, could result in lawsuits and government claims, investigations or proceedings.	Yes	Indirect: Description of Company Business
15	Any of these occurrences could damage our reputation, adversely impact customer and investor confidence, and could further result in a material adverse effect on Verizon's results of operation or financial condition.	Yes	Indirect: Description of Economic Consequences

Relevant paragraph algorithm accuracy: The algorithm successfully extracted 13/15 sentences or 86.67% of the total number of sentences.

Text from other paragraphs (outside Item 1A. Risk Factors): None

---

Appendix B: Variable Definitions

This table provides definitions for the key variables used in our analysis. All names within square brackets refer to Compustat item names.

Variable	Description	Source
Beta	The market beta of individual stocks estimated using monthly returns over the previous 60 months.	CRSP
Book-to-market	Book value of common equity [ceq] divided by the market value of common equity [prcc_f x csho]	Compustat
Cash Holdings	Cash holdings is the ratio of cash and short-term investments [che] to total assets [at].	Compustat
Cash Flow Volatility (Industry)	Industry average of the standard deviation of cash flow from operations [ib + dp – dvc] to total assets [at]. The standard deviation is estimated for each firm on a rolling basis using information available in the past five years. The industry is defined at the two-digit SIC level.	Compustat
CoSkew	The coefficient estimate of the market square term from a regression of monthly excess returns on market and market square excess returns; we require at least 24 months observations for the estimation.	CRSP
CRD Sentences (#)	The number of cybersecurity risk disclosure sentences in Item 1A. Risk Factors section	10-K
CRD Sentences (Ratio)	The ratio of the number of cybersecurity risk disclosure sentences scaled by the number of sentences in Item 1A. Risk Factors section;	10-K
Cyber Insurance	A dummy variable taking the value of 1 for firms that report in their 10-K that they have cyber insurance and also explicitly state that such insurance only partially covers them against claims that may arise due to cyber attacks, and 0 otherwise.	10-K
Cybersecurity Risk Index	The cosine similarity between a firm's cyber risk disclosure and the cyber risk disclosures of firms that have been subject to a cyber-attack during the one-year period prior to the firm's current filings.	10-K
Cybersecurity Risk Index (Jaccard)	The Jaccard similarity between a firm's cyber risk disclosure and the cyber risk disclosures of firms that have been subject to a cyber-attack during the one-year period prior to the firm's current filings.	10-K
EXTR_SIGMA	The negative of the worst deviation of firm-specific weekly returns from the average firm specific weekly return divided by the standard deviation of firm-specific weekly returns.	CRSP
Firm Age	Fiscal year – the year that the firm firstly appeared in Compustat.	Compustat
Firm Size	Total assets [at].	Compustat
High Google SVI Dummy	A dummy variable taking the value of 1 on days with high Search Volume Index (SVI) of the search topics "Data Breach" and "Hacker" in Google Trends, and 0 otherwise	Google Trends
Illiquidity	The ratio of the daily absolute stock return to the daily dollar trading volume averaged within the month; for the estimation, we require at least 15 daily returns within a given month.	CRSP

Independent Directors (%)	Number of independent directors in the board to the total number of board directors.	BoardEx
Idiosyncratic Volatility	The standard deviation of the residual series derived from Fama and French's (1993) three-factor model on monthly data within the prior 5 years.	CRSP
Institutional Ownership	Number of shares held by institutional shareholders that own more than 5% of a firm's equity to total number of shares outstanding.	Thomson-Reuters 13F
Leverage	Leverage is long-term debt [dltt] plus debt in current liabilities [dlc], scaled by total assets [at]	Compustat
Litigious Words	The number of "litigious" words in cybersecurity-risk disclosures. To identify "litigious" words, we draw upon the collection of pre-defined words constructed by Loughran and McDonald (2011).	10-K & Loughran and McDonald (2011)
Max	The average of the five highest daily returns of the stock during a month	CRSP
Momentum	The cumulative return of a stock over a period of 11 months ending one day prior to month *t*	CRSP
NCSKEW	The negative of the third moment of firm-specific weekly returns for each firm in a year divided by the standard deviation of firm-specific weekly returns raised to the third power.	CRSP
Negative Words	The number of "negative" words in cybersecurity-risk disclosures. To identify "negative" words, we draw upon the collection of pre-defined words constructed by Loughran and McDonald (2011).	10-K & Loughran and McDonald (2011)
Precise Words	The number of "precise" words in cybersecurity-risk disclosures. To identify "precise" words, we draw upon the collection of pre-defined words constructed by Loughran and McDonald (2011).	10-K & Loughran and McDonald (2011)
Previous Attack Dummy	A dummy variable taking the value of 1 for firms experienced past cyber attacks, and 0 otherwise.	PRC, Factiva
Readability	File size in megabytes of the SEC "complete submission text file" for the 10-K filing.	10-K
Reversal	The stock returns over the previous month.	CRSP
Risk Committee	A dummy variable that equals 1 if the name of a firm's board committee includes "risk", and 0 otherwise.	BoardEx
Risk Section Length	Number of sentences in Item 1A. Risk Factors of the 10-K.	10-K
ROA	Operating income before depreciation [oibdp] to total assets [at].	Compustat
R&D Expenditures	R&D expenditures [xrd] to total assets [at]. Missing values are replaced with zero.	Compustat
Secrets	A dummy variable that equals 1 if in a firm's 10-K filing there is any of the key phrases "trade secret", "trade secrets", "confidential information" or "proprietary information" and within a 5-word window before or after one the previous key phrases the firm also mentions "protect", "protection" or "safeguard", and 0 otherwise	10-K
Tangibility	Total property, plant and equipment [ppent] to total assets [at].	Compustat
Tobin's Q	Total assets [at] – common/ordinary equity [ceq] + market value of equity [prcc_f x csho] to total assets [at].	Compustat

Figure 1
Cybersecurity Risk by Year

This figure displays the average value of our cybersecurity risk measure and the number of cyber attacks by year. Based on the way our measure is constructed (i.e. we measure the similarity of each firm's cyber-related disclosures with those in past disclosures of firms that have been subject to cyber attacks), 2007 is the earliest year for which we get an estimate of cybersecurity risk (see Section 2 for details).

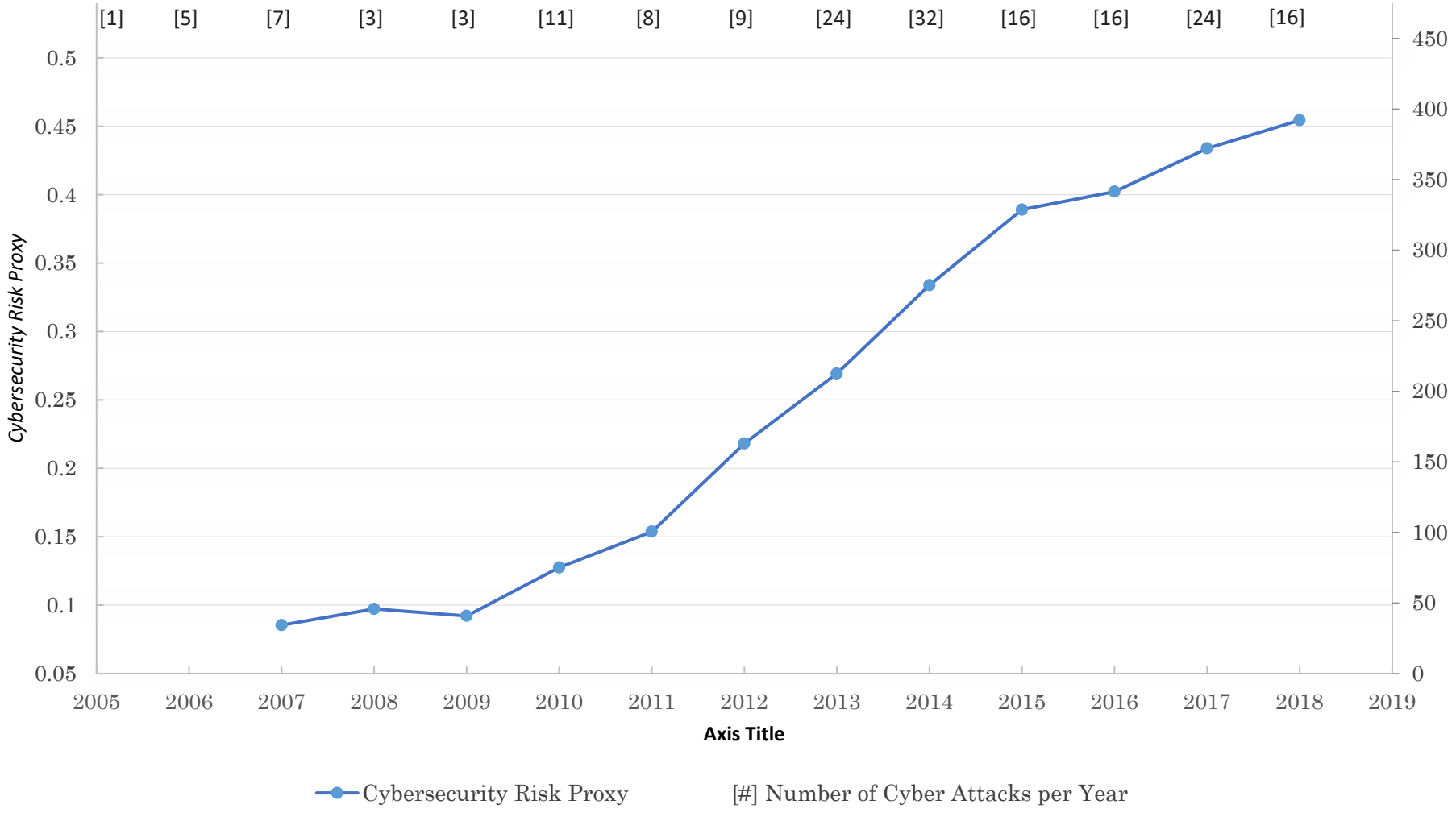

Figure 2
Cybersecurity Risk across Industries

This figure displays the average value of our cybersecurity risk measure and the number of cyber attacks by industry. Firms are classified into 12 industries according to Fama and French's 12 industry portfolios.

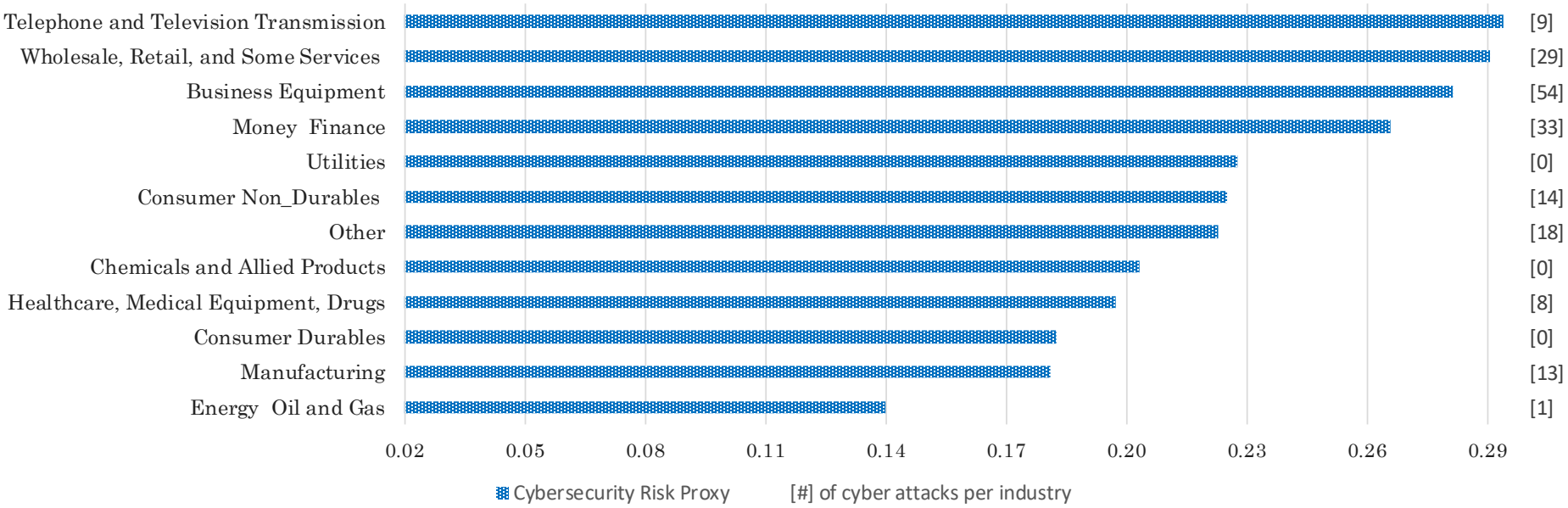

Table 1
Excerpts from Cybersecurity-risk Disclosures

Panel A: Excerpts for Firms with the Highest Cybersecurity Risk Score

Company Name	Fiscal Year	Industry	Cybersecurity Score	Text from Cybersecurity Risk Disclosures
Walgreens Boots Alliance Inc	2018	9	0.684	Like other global companies, we and businesses we interact with have experienced threats to data and systems, including by perpetrators of random or targeted malicious cyberattacks, computer viruses, worms, bot attacks or other destructive or disruptive software and attempts to misappropriate customer information, including credit card information, and cause system failures and disruptions.
Great Western Bancorp Inc	2016	11	0.683	We are not able to anticipate or implement effective preventive measures against all security breaches of these types, especially because the techniques used change frequently and because attacks can originate from a wide variety of sources.
Heritage Commerce Corp	2017	11	0.676	However, it is difficult or impossible to defend against every risk being posed by changing technologies as well as criminal intent on committing cyber-crime.
Salem Media Group Inc	2017	7	0.674	There can be no assurance that we, or the security systems we implement, will protect against all of these rapidly changing techniques.
Dexcom Inc	2017	10	0.670	Despite these efforts, threats from malicious persons and groups, new vulnerabilities and advanced new attacks against information systems create risk of cybersecurity incidents.

Panel B: Excerpts for Firms with Low Cybersecurity Risk Score

Company Name	Fiscal Year	Industry	Cybersecurity Score	Text from Cybersecurity Risk Disclosures
Weyerhaeuser Co	2015	12	0.036	We and our service providers employ what we believe are adequate security measures.
Hess Corp	2012	4	0.052	Examples of catastrophic risks include hurricanes, fires, explosions, blowouts, such as the accident at the Macondo prospect, pipeline interruptions and ruptures, severe weather, geological events, labor disputes or cyber attacks.
Wayside Technology Group Inc	2013	9	0.078	Any failure on the part of us or our vendors to maintain the security of data we are required to protect, including via the penetration of our network security and the misappropriation of confidential and personal information, could result in business disruption, damage to our reputation, financial obligations to third parties, fines, penalties, regulatory proceedings and private litigation with potentially large costs, and also result in deterioration in our employees', partners' and clients' confidence in us and other competitive disadvantages, and thus could have a material adverse impact on our business, financial condition and results of operations.
Sanderson Farms Inc	2012	1	0.109	Disruptions could be caused by a variety of factors, such as catastrophic events or weather, power outages, or cyber attacks on our systems by outside parties.
Dover Corp	2012	3	0.111	Disruptions or cybersecurity attacks, such as unauthorized access, malicious software, or other violations may lead to exposure of proprietary or confidential information as well as potential data corruption.

Table 2
Correlations

This table presents the correlation coefficients between our cybersecurity risk index and several quantitative measures based on cybersecurity risk disclosure language. CRD Sentences (#) is the number of cybersecurity risk disclosure sentences in Item 1A. Risk Factors section. CRD Sentences (Ratio) is the ratio of the number of cybersecurity risk disclosure sentences scaled by the number of sentences in Item 1A. Risk Factors section; Negative Words is the number of "negative" words in cybersecurity-risk disclosures. Precise Words is the number of "precise" words in cybersecurity-risk disclosures. Litigious Words is the number of "litigious" words in cybersecurity-risk disclosures. To identify "Negative Words", "Precise Words" and "Litigious Words", we draw upon the collection of pre-defined words proposed by Loughran and McDonald (2011). Cyber Insurance is a dummy variable taking the value of 1 for firms that report in their 10-K that they have cyber insurance and also explicitly state that such insurance only partially covers them against claims that may arise due to cyber attacks, and 0 otherwise. *** indicates statistical significance at the 1% level.

		(i)		*(ii)*		*(iii)*		*(iv)*		*(v)*		*(vi)*		*(vii)*
(i)	*Cybersecurity Risk*	1.000												
(ii)	*CRD Sentences (#)*	0.569	***	1.000										
(iii)	*CRD sentences (Ratio)*	0.443	***	0.717	***	1.000								
(iv)	*Negative Words*	0.033	***	-0.215	***	-0.133	***	1.000						
(v)	*Precise Words*	0.084	***	0.071	***	0.016	***	-0.145	***	1.000				
(vi)	*Litigious Words*	0.127	***	0.049	***	0.042	***	0.263	***	-0.071	***	1.000		
(vii)	*Cyber Insurance*	0.169	***	0.369	***	0.266	***	-0.115	***	0.003		0.004		1.000

Table 3
Descriptive Statistics

This table presents descriptive statistics for the key variables used in our analysis. Analytical variable definitions are provided in Appendix B

	Mean	*STDEV*	*P1*	*P25*	*P50*	*P75*	*P99*
Cybersecurity Risk Index	0.24	0.22	0.00	0.00	0.28	0.45	0.61
Firm Size (ln)	6.59	2.08	2.16	5.11	6.61	7.99	11.56
Firm Age (ln)	2.60	0.90	0.69	1.95	2.71	3.26	4.16
Tobin's Q	1.94	1.58	0.64	1.05	1.39	2.13	9.20
ROA	0.03	0.25	-1.08	0.01	0.08	0.14	0.42
Tanginility	0.21	0.24	0.00	0.02	0.10	0.30	0.89
R&D Expenditures	0.05	0.13	0.00	0.00	0.00	0.04	0.68
Secrets	0.30	0.46	0.00	0.00	0.00	1.00	1.00
Cash Flow Volatility (Industry)	0.10	0.07	0.00	0.05	0.08	0.12	0.34
Risk Section Length	262.61	178.71	1.00	138.00	226.00	346.00	841.00
Risk Section Length (ln)	5.26	1.04	0.69	4.93	5.42	5.85	6.74
Readability	10453409	11546923	384975	1865855	6163418	15323736	52900376
Readability (ln)	15.52	1.22	12.86	14.44	15.63	16.54	17.78
Institutional Ownership	0.20	0.16	0.00	0.06	0.18	0.30	0.65
Independent Directors	0.82	0.09	0.56	0.78	0.86	0.89	1.00
Risk Committee	0.05	0.22	0.00	0.00	0.00	0.00	1.00

Table 4
Cybersecurity Risk and Firm Characteristics

This table reports the results of linear regressions of firm characteristics on cybersecurity risk, as measured through cosine similarity (see Section 2.2 for details). All variables are defined in Appendix B. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

	Model 1	**Model 2**
Firm Size (ln)	**0.014** ***	**0.016** ***
	[13.28]	[4.70]
Firm Age (ln)	**-0.003**	**-0.041** ***
	[-1.29]	[-5.84]
Tobin's Q	**0.008** ***	**0.003** ***
	[6.73]	[3.36]
ROA	**0.062** ***	**0.033** ***
	[6.57]	[3.47]
Tanginility	**-0.085** ***	**-0.012**
	[-8.51]	[-0.58]
R&D Expenditures	**-0.006**	**0.090** ***
	[-0.32]	[4.27]
Secrets	**0.017** ***	**0.029** ***
	[4.16]	[4.24]
Cash Flow Volatility (Industry)	**-0.247** ***	**0.025**
	[-6.94]	[0.67]
Risk Section Length (ln)	**0.057** ***	**0.051** ***
	[40.80]	[20.59]
Readability (ln)	**0.007** ***	**0.004** *
	[2.92]	[1.93]
Institutional Ownership	**0.022** **	**0.011**
	[2.34]	[1.01]
Independent Directors	**0.406** ***	**0.046** **
	[5.79]	[1.98]
Risk Committee	**0.013** **	**-0.011**
	[2.09]	[-1.08]
Constant	**-0.461** ***	**-0.301** ***
	[-12.74]	[-6.60]
No of Observations	35,308	35,308
Clustered SE	Firm	Firm
Firm fixed effects	No	Yes
Industry fixed effects	Yes	No
Year fixed effects	Yes	Yes
R-Squared	0.523	0.780

Table 5
Cybersecurity Risk and (Negative) Asymmetries in Stock Returns

This table reports the results of regressions of cybersecurity risk on two different proxies for negative asymmetries in stock returns. In Model 1 we use *NCSKEW*, which equals the negative of the third moment of firm-specific weekly returns for each firm in a year divided by the standard deviation of firm-specific weekly returns raised to the third power. In Model 2, we use *EXTR_SIGMA*, which is the negative of the worst deviation of firm-specific weekly returns from the average firm specific weekly return divided by the standard deviation of firm-specific weekly returns. Cybersecurity risk is measured at the beginning of each year using cosine similarity. All variables are defined in Appendix B. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	NCSKEW	*EXTR_SIGMA*
	Model 1	Model 2
Cybersecurity Risk Index	**0.110** ***	**0.094** ***
	[3.14]	[2.91]
Firm Size (ln)	**0.048** ***	**0.026** ***
	[5.13]	[3.09]
Firm Age (ln)	**-0.031** ***	**-0.024** ***
	[-4.05]	[-3.36]
Tobin's Q	**-0.085** ***	**-0.075** ***
	[-9.59]	[-10.03]
ROA	**0.022**	**0.036** ***
	[1.48]	[2.70]
Tanginility	**-0.020** **	**-0.033** ***
	[-2.28]	[-4.04]
R&D Expenditures	**0.045** ***	**0.052** ***
	[2.95]	[3.75]
Secrets	**0.020** ***	**0.023** ***
	[2.70]	[3.33]
Cash Flow Volatility (Industry)	**0.005**	**0.002**
	[0.41]	[0.23]
Risk Section Length (ln)	**0.017** ***	**0.010** *
	[2.71]	[1.73]
Readability (ln)	**-0.015** *	**-0.010**
	[-1.89]	[-1.39]
Institutional Ownership	**0.043** ***	**0.030** ***
	[6.86]	[5.08]
Independent Directors	**-0.013** *	**-0.007**
	[-1.95]	[-1.07]
Risk Committee	**-0.033**	**-0.050** **
	[-1.55]	[-2.35]
Constant	**0.212** ***	**2.714** ***
	[9.05]	[116.3]
Clustered SE	Firm	Firm
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of Observations	24,657	24,657
R-squared	0.025	0.029

Table 6
Cybersecurity Risk and Future Cyber attacks

This table reports the results of logit regressions of cybersecurity risk (cosine similarity) on future cyber attacks. Panel A includes all cyber attacks reported in PRC database for which we have complete risk disclosure and financial data. In Panel B we restrict our attention to major cyber attacks and in particular those that attracted attention by global news outlets (e.g. CNBC, Financial Times and the Wall Street Journal) and covered in major Newswires (e.g. AP, Bloomberg, Reuters). In Panel C we restrict our attention to non-major cyber attacks (those that did not attract attention from major Newswires). Future cyber attacks are measured at time *t+1* while all independent variables are measured at time *t*. All variables are defined in Appendix B. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	Panel A: All Cyber Attacks		*Panel B: Major Cyber Attacks*		*Panel C: Non-major Cyber Attacks*	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Cybersecurity Risk Index	**0.961** ***	**0.656** ***	**0.749** ***	**0.461** **	**1.129** ***	**0.813** **
	[7.10]	[4.60]	[3.85]	[2.27]	[7.06]	[4.17]
Previous Attack Dummy	-	**1.503** ***	-	**1.694** **	-	**1.122** **
	-	[3.79]	-	[3.07]	-	[2.26]
Firm Size (ln)	-	**1.510** ***	-	**1.867** ***	-	**1.221** ***
	-	[10.53]	-	[8.41]	-	[7.72]
Firm Age (ln)	-	**-0.143**	-	**-0.244**	-	**-0.051**
	-	[-1.30]	-	[-1.53]	-	[-0.38]
Tobin's Q	-	**0.197**	-	**0.321**	-	**0.078**
	-	[1.31]	-	[1.63]	-	[0.39]
ROA	-	**0.483**	-	**0.400**	-	**0.563**
	-	[1.48]	-	[1.02]	-	[1.27]
Tanginility	-	**-0.042**	-	**-0.199**	-	**0.074**
	-	[-0.29]	-	[-0.89]	-	[0.42]
R&D Expenditures	-	**-0.031**	-	**-0.227**	-	**0.078**
	-	[-0.08]	-	[-0.46]	-	[0.15]
Secrets	-	**0.288** ***	-	**0.116**	-	**0.395** ***
	-	[2.95]	-	[0.83]	-	[3.10]
Cash Flow Volatility (Industry)	-	**-0.167**	-	**-0.507**	-	**-0.004**
	-	[-0.72]	-	[-1.36]	-	[-0.01]
Risk Section Length (ln)	-	**-0.235**	-	**-0.338** *	-	**-0.094**
	-	[-1.62]	-	[-1.77]	-	[-0.57]
Readability (ln)	-	**-0.006**	-	**-0.149**	-	**0.111**
	-	[-0.04]	-	[-0.71]	-	[0.55]
Institutional Ownership	-	**0.135**	-	**0.440** ***	-	**-0.088**
	-	[1.12]	-	[2.67]	-	[-0.62]
Independent Directors	-	**-0.065**	-	**-0.140**	-	**0.003**
	-	[-0.58]	-	[-0.98]	-	[0.02]
Risk Committee	-	**-0.182**	-	**-0.416**	-	**-0.029**
	-	[-0.45]	-	[-0.95]	-	[-0.05]
Constant	**-8.261** ***	**-8.790** ***	**-8.568** ***	**-9.860** ***	**-9.303** ***	**-9.408** ***
	[-9.87]	[-10.42]	[-7.63]	[-9.00]	[-7.74]	[-7.65]
Clustered SE	Firm	Firm	Firm	Firm	Firm	Firm
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	41,140	30,830	38,934	30,830	41,140	30,830
Pseudo-R-squared	0.093	0.223	0.074	0.235	0.099	0.204

Table 7
Cybersecurity Risk Portfolios

This table reports average excess returns, CAPM alphas, four-factor alphas from Carhart's (1997) FFC model (FFC alphas) and five-factor alphas from Fama and French's (2015) model (Five-Factor alphas) for portfolios constructed on the basis of our Cybersecurity Risk Index, which is measured through cosine similarity. Starting from December 2007, we sort stocks at the end of each quarter in ascending order on the basis of their Cybersecurity Risk and allocate them into three groups (Low Cyber-Risk Stocks, Middle Group and High Cyber-Risk Stocks). We track the performance of the three portfolios over the following quarter until these are rebalanced. We form the spread strategy P3-P1 that is long the portfolio with the highest cybersecurity risk stocks (P3) and short the portfolio with the lowest cybersecurity risk stocks (P1). Panel A reports returns for equally-weighted (ew) and value-weighted (vw) portfolios over the period March 2008- March 2019. Average (monthly) excess portfolio returns and alphas are bolded; their associated Newey-West *t*-statistics are reported in square brackets. We exclude from the analysis firms that appear in a sample for a period less than 3 years and have zero disclosures on cyber-related issues throughout that period. Panel B reports the (equally-weighted) average number of firms per portfolio, average exposure to cybersecurity risk and average value for a series of firm/stock and 10-K characteristics. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Panel A: Future (1-month) portfolio returns sorted by our Cybersecurity Risk Index					
		Portfolios			
		Low Cyber-Risk [P1]	*Middle Group* [P2]	*High Cyber-Risk* [P3]	[P3]-[P1]
Excess return	ew	**0.169**	**0.710** *	**0.843** **	**0.674** ***
		[0.38]	[1.70]	[2.17]	[4.54]
	vw	**0.508**	**0.831** **	**1.117** ***	**0.609** ***
		[1.25]	[2.40]	[3.32]	[3.02]
CAPM alpha	ew	**-0.727** **	**-0.219**	**-0.054**	**0.673** ***
		[-3.32]	[-0.97]	[-0.37]	[4.69]
	vw	**-0.339** *	**-0.010**	**0.321** ***	**0.660** ***
		[-1.90]	[-0.09]	[4.01]	[3.41]
FFC alpha	ew	**-0.675** ***	**-0.169**	**0.011**	**0.686** ***
		[-4.87]	[-1.54]	[0.13]	[4.80]
	vw	**-0.277** *	**0.020**	**0.282** ***	**0.559** ***
		[-1.87]	[0.18]	[3.43]	[3.30]
Five-factor alpha	ew	**-0.602** ***	**-0.108**	**0.055**	**0.657** ***
		[-3.80]	[-0.72]	[0.74]	[4.38]
	vw	**-0.306** **	**0.016**	**0.268** ***	**0.574** ***
		[-2.30]	[0.12]	[3.23]	[3.58]
Panel B: Firm/Stock/10-K characteristics					
Number of firms		1233	960	966	-
Cybersecurity Risk Index		0.000	0.310	0.465	-
Market Value (ln)		12.375	13.483	13.717	-
Book-to-Market		0.717	0.596	0.615	-
ROA		0.023	0.024	0.069	-
Institutional Ownership		0.169	0.212	0.215	-
Illiquidity		1.971	0.881	0.842	-
Idiosyncratic Volatility		3.085	2.561	2.258	-
Risk Section Length (ln)		4.679	5.491	5.546	-
Readability (ln)		15.586	15.849	15.927	-

Table 8
Double-Sorted Portfolios

This table reports average returns and 5-factor alphas from the Fama and French's (2015) model for double-sorted portfolios on the basis of the cybersecurity risk index and each of the following firm characteristics: (i) Market Value, which is the natural logarithm of market value, (ii) Book-to-Market, is the book value of common equity divided by the market value of common equity; (iii) ROA, a measure of profitability, proxied by return on assets; (iv) Institutional Ownership, defined as the number of shares held by institutional shareholders that own more than 5% of a firm's equity to the total number of shares outstanding (v) Illiquidity, the ratio of the daily absolute stock return to the daily dollar trading volume averaged within the month, (vi) Idiosyncratic Volatility, defined as the standard deviation of the residuals estimated from the Fama and French (1993) three-factor model on monthly data within the prior 5 years; (vii) Risk Section Length, which is the number of sentences in Item 1A. Risk Factors of the Form 10-K; and (viii) Readability, which is the file size in megabytes of the SEC "complete submission text file" for the 10-K filing. Starting from December 2007, we sort stocks at the end of each quarter in ascending order on the basis of their Cybersecurity Risk and allocate them into three groups (Low Cyber-Risk Stocks, Middle Group and High Cyber-Risk Stocks), and we also independently sort stocks into ascending order according to the value of each characteristic mentioned above and allocate them into two portfolios (LOW and HIGH) based on median values for each quarter. The intersection of these two classifications yields the double-sorted portfolios. We track the performance of the intersection portfolios over the following quarter until these are rebalanced. We report both equal-weighted and value-weighted average returns and five-factor alphas for the spread strategy High-Low Cyber Risk Stocks within each HIGH and LOW classification. Newey-West *t*-statistics are reported in square brackets. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

		Equal-weighted portfolios High - Low Cyber Risk Stocks		Value-weighted portfolios High - Low Cyber Risk Stocks	
		Avg. Return	*5-Factor alpha*	*Avg. Return*	*5-Factor alpha*
Panel A: Firm Characteristics					
	LOW	**0.681** ***	**0.668** ***	**0.418** ***	**0.451** ***
Market Value		[4.39]	[3.47]	[2.63]	[2.74]
	HIGH	**0.195** *	**0.284** ***	**0.577** ***	**0.547** ***
		[1.91]	[2.60]	[2.65]	[3.12]
	LOW	**0.818** ***	**0.758** ***	**0.755** **	**0.725** ***
Book-to-Market		[5.82]	[5.71]	[2.51]	[3.01]
	HIGH	**0.463** ***	**0.519** ***	**0.280**	**0.328** *
		[2.71]	[2.87]	[1.49]	[1.88]
	LOW	**0.918** ***	**0.959** ***	**0.589** *	**0.537** *
ROA		[5.01]	[4.72]	[1.90]	[1.68]
	HIGH	**0.287** **	**0.216** *	**0.411** **	**0.412** **
		[2.04]	[1.66]	[2.27]	[2.41]
	LOW	**0.770** ***	**0.755** ***	**0.664** ***	**0.589** ***
Institutional Ownership		[5.22]	[4.86]	[2.62]	[2.99]
	HIGH	**0.285** **	**0.277** ***	**0.170**	**0.272** *
		[2.48]	[2.63]	[1.14]	[1.71]

Table Continued Overleaf

Table 8 (Continued)

		Equal-weighted portfolios High - Low Cyber Risk Stocks		Value-weighted portfolios High - Low Cyber Risk Stocks	
		Avg. Return	*5-Factor alpha*	*Avg. Return*	*5-Factor alpha*
Panel B: Stock & 10-K Characteristics					
	LOW	**0.271 ***	**0.365 *****	**0.167**	**0.268 *****
Illiquidity		[1.91]	[3.14]	[1.33]	[2.70]
	HIGH	**0.702 *****	**0.710 *****	**0.262 ***	**0.309 ***
		[4.38]	[3.74]	[1.68]	[1.89]
	LOW	**0.103**	**0.087**	**0.583 ****	**0.551 *****
Idiosyncratic Volatility		[1.06]	[0.82]	[2.51]	[2.77]
	HIGH	**0.791 *****	**0.759 *****	**0.416**	**0.468 ***
		[5.72]	[5.07]	[1.24]	[1.79]
	LOW	**0.348 ****	**0.348 *****	**0.540 ****	**0.559 *****
Risk Section Length (ln)		[2.56]	[2.61]	[2.15]	[2.87]
	HIGH	**1.193 *****	**1.225 *****	**0.530 *****	**0.488 *****
		[4.97]	[4.91]	[2.81]	[3.03]
	LOW	**0.861 *****	**0.786 *****	**0.826 *****	**0.750 *****
Readability (ln)		[6.22]	[4.84]	[3.60]	[4.24]
	HIGH	**0.273 ****	**0.303 *****	**0.445 ****	**0.441 ****
		[2.02]	[2.62]	[2.03]	[2.47]

Table 9
Cross sectional Fama-MacBeth Regressions

This table reports the results from Fama-MacBeth regressions on the relation between our Cybersecurity Risk Index and subsequent monthly stock returns (1-month to 12-month). For each month of our sample we run cross sectional regressions of excess stock returns on lagged cybersecurity risk and a set of firm characteristics that are also lagged. These include beta, size, book-to-market, momentum, short-term reversal, illiquidity, coskewness, idiosyncratic volatility, asset growth, profitability and demand for lottery-like stocks (max) , length of Item 1A. Risk Factors of the Form 10-K and 10-K readability. All variables are defined in Appendix B. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. The coefficients are reported as time-series averages of the estimates from the cross sectional regressions. The *t*-statistics, which are reported in brackets, are based on the Newey-West heteroskedasticity and autocorrelation consistent standard errors. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

	$Returns_{t+1}$			$Returns_{t+2}$	$Returns_{t+3}$	$Returns_{t+6}$	$Returns_{t+9}$	$Returns_{t+12}$
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Cybersecurity Risk Index	**0.298 *** **	**0.097 ** **	**0.117 *** **	**0.110 *** **	**0.104 *** **	**0.145 *** **	**0.112 *** **	**0.108 ** **
	[6.28]	[2.53]	[2.67]	[2.72]	[2.66]	[3.01]	[2.76]	[2.52]
Beta	-	**0.097**	**0.094**	**0.092**	**0.038**	**0.030**	**0.027**	**0.017**
		[0.89]	[0.90]	[0.82]	[0.34]	[0.30]	[0.28]	[0.18]
Market Value	-	**-0.074**	**-0.028**	**-0.043**	**0.013**	**0.006**	**0.067**	**0.012**
		[-1.09]	[-0.37]	[-0.60]	[0.17]	[0.09]	[0.92]	[0.15]
Book-to-Market	-	**-0.001**	**0.007**	**0.009**	**0.002**	**0.015**	**0.010**	**0.022**
		[-0.02]	[0.14]	[0.16]	[0.03]	[0.28]	[0.20]	[0.49]
Momentum	-	**0.163 ***	**0.154 ***	**0.108**	**0.116**	**0.084**	**0.172 *** **	**0.153 *** **
		[1.86]	[1.75]	[1.34]	[1.56]	[1.28]	[4.15]	[3.32]
Reversal	-	**-0.114 ** **	**-0.121 ** **	**0.207 *** **	**0.136 ** **	**0.145 *** **	**0.113 ***	**0.071**
		[-2.02]	[-2.19]	[2.85]	[2.41]	[3.08]	[1.86]	[1.10]
Illiquidity	-	**-0.019**	**-0.018**	**0.020**	**0.025**	**0.048**	**0.021**	**-0.009**
		[-0.53]	[-0.51]	[0.67]	[0.82]	[1.54]	[0.68]	[-0.27]
CoSkew	-	**-0.027**	**-0.025**	**-0.007**	**-0.021**	**-0.006**	**-0.025**	**0.035**
		[-0.88]	[-0.79]	[-0.23]	[-0.59]	[-0.18]	[-0.72]	[1.02]
Indiosyncratic Volatility	-	**-0.454 *** **	**-0.454 *** **	**-0.369 *** **	**-0.469 *** **	**-0.475 *** **	**-0.381 *** **	**-0.436 *** **
		[-5.40]	[-5.58]	[-3.97]	[-5.31]	[-5.66]	[-4.36]	[-5.33]
Asset Growth	-	**-0.121 *** **	**-0.114 ** **	**-0.081 ** **	**-0.054**	**-0.078 ***	**0.000**	**-0.003**
		[-3.14]	[-3.11]	[-2.27]	[-1.56]	[-1.70]	[-0.01]	[-0.07]
ROA	-	**0.295 *** **	**0.285 *** **	**0.282 *** **	**0.256 *** **	**0.303 *** **	**0.329 *** **	**0.407 *** **
		[5.29]	[5.17]	[5.26]	[4.62]	[5.45]	[7.14]	[7.70]
Max	-	**-0.420 *** **	**-0.413 *** **	**-0.402 *** **	**-0.215 *** **	**-0.193 ** **	**-0.118**	**-0.131 ***
		[-4.42]	[-4.46]	[-4.11]	[-2.72]	[-2.43]	[-1.48]	[-1.67]
Risk Section Length (ln)	-	-	**-0.034**	**-0.046**	**-0.040**	**-0.041**	**0.003**	**-0.009**
			[-0.81]	[-1.23]	[-1.11]	[-1.10]	[0.08]	[-0.26]
Readability (ln)	-	-	**-0.081**	**-0.066**	**-0.055**	**-0.040**	**-0.042**	**-0.019**
			[-1.46]	[-1.23]	[-1.03]	[-0.80]	[-0.82]	[-0.36]
Constant	**0.515**	**0.514**	**0.512**	**0.479**	**0.500**	**0.512**	**0.804 ** **	**0.917 ** **
	[1.14]	[1.13]	[1.12]	[1.04]	[1.09]	[1.14]	[2.25]	[2.44]
Observations	409,016	342,573	342,573	334,847	333,325	328,887	324,633	314,506

Table 10
Cybersecurity Risk Factor: Time Series Variation

This table presents the results of the regression $CRF_t = a + \beta \times High\_Google\_SVI\_dummy_t + \gamma_i \times X_t + error$, where CRF is our cybersecurity risk factor, "High_Google_SVI_dummy" is a dummy variable that takes the value of 1 on days with high Googler SVI index of the search topics "Data Breach" and "Hacker", and 0 otherwise, X is a vector of the (daily) risk factors proposed by Carhart (1997) and Fama and French (2015), namely market, size, value, momentum, operating profitability and investment factors. Model 1 does not control for any risk factors. Model 2 only controls for the market risk factor (CAPM specification), Model 3 controls for market, value and momentum factors (FFC specification), while Models 4 control for all five risk factors proposed by Fama and French (2015) (FF-5 specification). In Panels B (C) we replace the variable *High_Google_SVI_dummy* with the variable *High_Google_SVI_dummy + 5 days (+ 1 month)*, which takes the value of 1 on days a week (a month after) after the actual peak of the SVI index, and zero otherwise. For the estimation we use daily data over the period March 2008-March 2019. The *t*-statistics, which are reported in brackets, are based on the Newey-West heteroskedasticity and autocorrelation consistent standard errors. *** denotes statistical significance at the 1% level.

	Cybersecurity Risk Factor $_t$			
	CONTROLS			
	NONE	*CAPM*	*FFC*	*FF-5*
	[1]	[2]	[3]	[4]
Panel A: Days with High SVI (Google Trends)				
Constant	**0.0002 *****	**0.0002 *****	**0.0002 *****	**0.0002 *****
	[3.47]	[3.64]	[3.50]	[3.54]
High Google SVI Dummy	**-0.0015 *****	**-0.0014 *****	**-0.0014 *****	**-0.0013 *****
	[-4.39]	[-4.34]	[-4.18]	[-4.30]
Observations (days)	2,789	2,789	2,789	2,789
Panel B: Placebo Tests (5 days after the the peak of SVI)				
Constant	**0.0002 *****	**0.0002 *****	**0.0002 *****	**0.0002 *****
	[2.94]	[3.13]	[2.99]	[3.03]
Placebo High Google SVI Dummy + 5 days	**0.0003**	**0.0003**	**0.0003**	**0.0003**
	[0.62]	[0.59]	[0.68]	[0.77]
Observations (days)	2,789	2,789	2,789	2,789
Panel C: Placebo Tests (1 month after the the peak of SVI)				
Constant	**0.0002 *****	**0.0002 *****	**0.0002 *****	**0.0002 *****
	[2.91]	[3.08]	[2.97]	[3.04]
Placebo High Google SVI Dummy + 1 month	**0.0002**	**0.0002**	**0.0002**	**0.0002**
	[0.55]	[0.55]	[0.48]	[0.47]
Observations (days)	2,789	2,789	2,789	2,789

Table 11
Cybersecurity Risk Portfolios-Robustness Tests

This table reports average excess returns and alphas from the Fama and French's (2015) model (Five-Factor alphas) for the spread strategy that is long the portfolio with the highest cybersecurity risk stocks and short the portfolio with the lowest cybersecurity risk stocks. Results are reported both for equally-weighted and value-weighted portfolios. We exclude from the analysis firms that appear in a sample for a period less than 3 years and have zero disclosures on cyber-related issues throughout that period. In Panel A we repeat our portfolio analysis for the period January 2012-March 2019 (Post SEC's guidance on cybersecurity). In Panel B (Panel C) we exclude firms with cyber insurance (in the training sample). In Panel D we present results based on monthly and yearly rebalancing. In Panel E, we repeat the analysis 12 times after excluding each of the Fama-French 12 industries in turn to flush out abnormal impact of any particular industry group. In Panel F, we form our portfolios based on another revised cybersecurity risk measure, which replaces all zeros with the industry/sector median value in any given year. In Panel G, we form our portfolios based on a revised cybersecurity risk measure, which replaces all zeros with the next non-zero observation for each firm. Average (monthly) excess portfolio returns and alphas are bolded; their associated Newey-West t-statistics are reported in square brackets. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

	Equal-weighted portfolios High - Low Cyber Risk Stocks		Value-weighted portfolios High - Low Cyber Risk Stocks	
	Avg. Return	*5-Factor alpha*	*Avg. Return*	*5-Factor alpha*
Panel A: Post SEC's Guidance on Cybersecurity				
January 2012 to March 2019	**0.916 *****	**0.870 *****	**0.770 *****	**0.652 *****
	[5.88]	[5.56]	[3.00]	[3.45]
Panel B: All Firms Excluding:				
Firms with Cyber Insurance	**0.660 *****	**0.634 *****	**0.676 *****	**0.649 *****
	[4.40]	[4.27]	[3.34]	[4.05]
Panel C: All Firms Excluding:				
Firms in Training Sample	**0.669 *****	**0.682 *****	**0.522 ****	**0.507 ****
	[4.42]	[4.67]	[2.20]	[2.42]
Panel D: Alternative Rebalancing				
Monthly Rebalancing	**0.667 *****	**0.646 *****	**0.599 *****	**0.561 *****
	[4.48]	[4.32]	[2.95]	[3.49]
Yearly Rebalancing	**0.691 *****	**0.669 *****	**0.586 *****	**0.559 *****
	[4.73]	[4.52]	[2.84]	[3.40]

Table Continued Overleaf

Table 11 (Continued)

	Equal-weighted portfolios High - Low Cyber Risk Stocks		Value-weighted portfolios High - Low Cyber Risk Stocks	
	Avg. Return	*5-Factor alpha*	*Avg. Return*	*5-Factor alpha*
Panel E: All Firms Excluding:				
Consumer Non_Durables	**0.704** ***	**0.675** ***	**0.698** ***	**0.644** ***
	[4.45]	[3.24]	[4.44]	[3.89]
Consumer Durables	**0.674** ***	**0.619** ***	**0.646** ***	**0.577** ***
	[4.63]	[3.07]	[4.42]	[3.57]
Manufacturing	**0.690** ***	**0.633** ***	**0.636** ***	**0.564** ***
	[5.01]	[3.22]	[4.28]	[3.44]
Energy Oil and Gas	**0.614** ***	**0.566** ***	**0.570** ***	**0.495** ***
	[4.60]	[2.80]	[3.99]	[2.88]
Chemicals and Allied Products	**0.666** ***	**0.638** ***	**0.647** ***	**0.592** ***
	[4.58]	[3.06]	[4.42]	[3.54]
Business Equipment	**0.575** ***	**0.519** **	**0.570** ***	**0.580** ***
	[3.56]	[2.46]	[3.52]	[3.27]
Telephone and Television Transmission	**0.685** ***	**0.609** ***	**0.666** ***	**0.571** ***
	[4.59]	[2.96]	[4.45]	[3.41]
Utilities	**0.688** ***	**0.620** ***	**0.672** ***	**0.585** ***
	[4.58]	[3.03]	[4.46]	[3.67]
Wholesale, Retail, and Some Services	**0.760** ***	**0.573** ***	**0.775** ***	**0.541** ***
	[4.54]	[2.91]	[4.92]	[3.44]
Healthcare, Medical Equipment, Drugs	**0.659** ***	**0.681** ***	**0.660** ***	**0.650** ***
	[4.31]	[3.17]	[4.47]	[3.75]
Money Finance	**0.746** ***	**0.711** ***	**0.671** ***	**0.692** ***
	[4.93]	[2.74]	[4.41]	[3.53]
Other	**0.665** ***	**0.450** ***	**0.661** ***	**0.409** ***
	[4.14]	[3.08]	[3.97]	[3.36]

Table Continued Overleaf

Table 11 (Continued)

	Equal-weighted portfolios High - Low Cyber Risk Stocks		Value-weighted portfolios High - Low Cyber Risk Stocks	
	Avg. Return	*5-Factor alpha*	*Avg. Return*	*5-Factor alpha*
Panel F: Replacing zeros with industry medians				
Tercile Portfolios (P3-P1)	**0.572 ***** [3.87]	**0.660 ***** [5.09]	**0.369 **** [2.07]	**0.470 ***** [2.62]
Quartile Portfolios (P4-P1)	**0.601 ***** [3.27]	**0.723 ***** [4.56]	**0.393 *** [1.91]	**0.569 ***** [3.40]
Quintile Portfolios (P5-P1)	**0.602 ***** [3.04]	**0.704 ***** [4.10]	**0.459 **** [2.07]	**0.584 ***** [3.27]
Decile Portfolios (P10-P1)	**0.527 **** [2.43]	**0.673 ***** [3.80]	**0.387** [1.49]	**0.487 **** [2.36]
Panel G: Replacing zeros with next non-zero obs.				
Tercile Portfolios (P3-P1)	**0.235 **** [2.08]	**0.300 ***** [2.79]	**0.289 ***** [3.11]	**0.289 ***** [2.73]
Quartile Portfolios (P4-P1)	**0.306 **** [2.27]	**0.386 ***** [3.17]	**0.283 **** [2.49]	**0.308 **** [2.23]
Quintile Portfolios (P5-P1)	**0.384 ***** [2.66]	**0.471 ***** [3.61]	**0.314 **** [2.16]	**0.363 **** [2.35]
Decile Portfolios (P10-P1)	**0.515 ***** [2.75]	**0.423 **** [2.24]	**0.631 ***** [3.51]	**0.529 **** [2.46]

Internet Appendix

for

Cybersecurity Risk

Chris Florackis, Christodoulos Louca, Roni Michaely, Michael Weber

Table IA.1
Cybersecurity Risk and Firm Characteristics-Jaccard Similarity

This table reports the results of linear regressions of firm characteristics on cybersecurity risk as measured through Jaccard similarity. All variables are defined in Appendix B. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

	Model 1		Model 2	
Firm Size (ln)	**0.006**	***	**0.007**	***
	[14.44]		[4.94]	
Firm Age (ln)	**-0.003**	***	**-0.016**	***
	[-4.72]		[-5.73]	
Tobin's Q	**0.004**	***	**0.001**	*
	[7.84]		[1.89]	
ROA	**0.021**	***	**0.012**	***
	[5.55]		[3.25]	
Tanginility	**-0.034**	***	**-0.010**	
	[-8.78]		[-1.28]	
R&D Expenditures	**-0.008**		**0.037**	***
	[-0.99]		[4.72]	
Secrets	**0.010**	***	**0.013**	***
	[5.91]		[4.70]	
Cash Flow Volatility (Industry)	**-0.095**	***	**0.006**	
	[-6.70]		[0.42]	
Risk Section Length (ln)	**0.022**	***	**0.020**	***
	[39.70]		[20.15]	
Readability (ln)	**0.002**	**	**0.002**	*
	[2.36]		[1.73]	
Institutional Ownership	**0.010**	***	**0.006**	
	[2.66]		[1.57]	
Independent Directors	**0.037**	***	**0.015**	*
	[5.04]		[1.68]	
Risk Committee	**0.007**	**	**0.000**	
	[2.51]		[-0.10]	
Constant	**-0.175**	***	**-0.112**	***
	[-12.03]		[-6.33]	
No of Observations	35,308		35,308	
Clustered SE	Firm		Firm	
Firm fixed effects	No		Yes	
Industry fixed effects	Yes		No	
Year fixed effects	Yes		Yes	
R-Squared	0.510		0.782	

Table IA.2

Cybersecurity Risk and Negative Asymmetries in Stock Returns- Jaccard Similarity

This table reports the results of regressions of cybersecurity risk on two different proxies for negative asymmetries in stock returns. In Model 1, we use *NCSKEW*, which equals the negative of the third moment of firm-specific weekly returns for each firm in a year divided by the standard deviation of firm-specific weekly returns raised to the third power. In Model 2, we use *EXTR_SIGMA*, which is the negative of the worst deviation of firm-specific weekly returns from the average firm specific weekly return divided by the standard deviation of firm-specific weekly returns. Cybersecurity risk is measured at the beginning of each year using Jaccard similarity. All variables are defined in Appendix B. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	NCSKEW	*EXTR_SIGMA*
	Model 1	Model 2
Cybersecurity Risk Index (Jaccard)	**0.312** ***	**0.280** ***
	[3.53]	[3.47]
Firm Size (ln)	**0.047** ***	**0.025** ***
	[5.03]	[2.98]
Firm Age (ln)	**-0.030** ***	**-0.023** ***
	[-3.94]	[-3.25]
Tobin's Q	**-0.086** ***	**-0.075** ***
	[-9.65]	[-10.11]
ROA	**0.022**	**0.036** ***
	[1.49]	[2.70]
Tanginility	**-0.020** **	**-0.032** ***
	[-2.24]	[-4.00]
R&D Expenditures	**0.045** ***	**0.053** ***
	[2.98]	[3.78]
Secrets	**0.019** ***	**0.022** ***
	[2.63]	[3.26]
Cash Flow Volatility (Industry)	**0.004**	**0.002**
	[0.40]	[0.23]
Risk Section Length (ln)	**0.016** ***	**0.009**
	[2.61]	[1.59]
Readability (ln)	**-0.015** *	**-0.010**
	[-1.88]	[-1.38]
Institutional Ownership	**0.043** ***	**0.029** ***
	[6.85]	[5.07]
Independent Directors	**-0.013** *	**-0.007**
	[-1.95]	[-1.08]
Risk Committee	**-0.033**	**-0.050** **
	[-1.58]	[-2.38]
Constant	**0.211** ***	**2.714** ***
	[9.05]	[116.4]
Clustered SE	Firm	Firm
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of Observations	24,657	24,657
R-squared	0.025	0.029

Table IA.3
Cybersecurity Risk and Future Cyber attacks-Jaccard Similarity

This table reports the results of logit regressions of cybersecurity risk (Jaccard similarity) on future cyber attacks. Panel A includes all cyber attacks reported in PRC database for which we have complete risk disclosure and financial data. In Panel B we restrict our attention to major cyber attacks and in particular those that attracted attention by global news outlets (e.g. CNBC, Financial Times and the Wall Street Journal) and covered in major Newswires (e.g. AP, Bloomberg, Reuters). In Panel C we restrict our attention to non-major cyber attacks (those that did not attract attention from major Newswires). Future cyber attacks are measured at time *t+1* while all independent variables are measured at time *t*. All variables are defined in Appendix B. Standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	Panel A: All Cyber Attacks		*Panel B: Major Cyber Attacks*		*Panel C: Non-major Cyber Attacks*	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Cybersecurity Risk Index (Jaccard)	**0.749 *****	**0.504 *****	**0.743 *****	**0.562 *****	**0.721 *****	**0.440 *****
	[8.67]	[5.37]	[5.45]	[4.75]	[8.20]	[3.65]
Previous Attack Dummy	-	**1.435 *****	-	**1.622 *****	-	**1.061 ****
	-	[3.76]	-	[2.99]	-	[2.14]
Firm Size (ln)	-	**1.496 *****	-	**1.814 *****	-	**1.236 *****
	-	[10.55]	-	[8.36]	-	[7.61]
Firm Age (ln)	-	**-0.121**	-	**-0.221**	-	**-0.031**
	-	[-1.10]	-	[-1.38]	-	[-0.22]
Tobin's Q	-	**0.190**	-	**0.300**	-	**0.087**
	-	[1.28]	-	[1.57]	-	[0.43]
ROA	-	**0.505**	-	**0.416**	-	**0.597**
	-	[1.59]	-	[1.11]	-	[1.39]
Tanginility	-	**-0.038**	-	**-0.192**	-	**0.065**
	-	[-0.26]	-	[-0.86]	-	[0.36]
R&D Expenditures	-	**-0.016**	-	**-0.149**	-	**0.068**
	-	[-0.04]	-	[-0.31]	-	[0.13]
Secrets	-	**0.280 *****	-	**0.105**	-	**0.393 *****
	-	[2.87]	-	[0.75]	-	[3.08]
Cash Flow Volatility (Industry)	-	**-0.177**	-	**-0.531**	-	**-0.029**
	-	[-0.79]	-	[-1.43]	-	[-0.10]
Risk Section Length (ln)	-	**-0.185**	-	**-0.389 ****	-	**0.017**
	-	[-1.42]	-	[-2.24]	-	[0.12]
Readability (ln)	-	**0.017**	-	**-0.102**	-	**0.121**
	-	[0.12]	-	[-0.49]	-	[0.61]
Institutional Ownership	-	**0.139**	-	**0.444 *****	-	**-0.082**
	-	[1.17]	-	[2.70]	-	[-0.57]
Independent Directors	-	**-0.067**	-	**-0.158**	-	**0.012**
	-	[-0.59]	-	[-1.11]	-	[0.07]
Risk Committee	-	**-0.203**	-	**-0.448**	-	**-0.037**
	-	[-0.49]	-	[-1.02]	-	[-0.06]
Constant	**-8.035 *****	**-8.607 *****	**-8.594 *****	**-9.867 *****	**-8.783 *****	**-9.027 *****
	1[-0.12]	[-10.71]	[-7.80]	[-9.13]	[-8.04]	[-8.06]
Clustered SE	Firm	Firm	Firm	Firm	Firm	Firm
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	41,140	30,830	38,934	30,830	41,140	30,830
Pseudo-R-squared	0.094	0.223	0.090	0.244	0.086	0.196

Table IA.4
Cybersecurity Risk Portfolios-Jaccard Similarity

This table reports average excess returns, CAPM alphas, four-factor alphas from Carhart's (1997) FFC model (FFC alphas) and five-factor alphas from Fama and French's (2015) model (Five-Factor alphas) for portfolios constructed on the basis of our Cybersecurity Risk Index, as measured through Jaccard similarity. Starting from December 2007, we sort stocks at the end of each quarter in ascending order on the basis of their Cybersecurity Risk and allocate them into three groups (Low Cyber-Risk Stocks, Middle Group and High Cyber-Risk Stocks). We track the performance of the three portfolios over the following quarter until these are rebalanced. We form the spread strategy P3-P1 that is long the portfolio with the highest cybersecurity risk stocks (P3) and short the portfolio with the lowest cybersecurity risk stocks (P1). Returns are reported for equally-weighted (ew) and value-weighted (vw) portfolios over the period March 2008- March 2019. Average (monthly) excess portfolio returns and alphas are bolded; their associated Newey-West *t*-statistics are reported in square brackets. We exclude from the analysis firms that appear in a sample for a period less than 3 years and have zero disclosures on cyber-related issues throughout that period. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

Future (1-month) portfolio returns sorted by our Cybersecurity Risk Index					
		Portfolios			
		Low Cyber-Risk	*Middle Group*	*High Cyber-Risk*	
		[P1]	[P2]	[P3]	[P3]-[P1]
Excess return	ew	**0.169**	**0.701** *	**0.852** **	**0.683** ***
		[0.38]	[1.71]	[2.15]	[4.70]
	vw	**0.508**	**0.883** ***	**1.025** ***	**0.517** **
		[1.25]	[2.62]	[2.95]	[2.33]
CAPM alpha	ew	**-0.727** **	**-0.214**	**-0.059**	**0.668** ***
		[-3.32]	[-0.98]	[-0.39]	[4.71]
	vw	**-0.339** *	**0.090**	**0.181** *	**0.520** **
		[-1.90]	[0.80]	[1.85]	[2.37]
FFC alpha	ew	**-0.675** ***	**-0.154**	**-0.005**	**0.670** ***
		[-4.87]	[-1.41]	[-0.06]	[4.70]
	vw	**-0.277** *	**0.103**	**0.166**	**0.443** **
		[-1.87]	[1.04]	[1.61]	[2.26]
Five-factor alpha	ew	**-0.602** ***	**-0.111**	**0.058**	**0.660** ***
		[-3.80]	[-0.79]	[0.69]	[4.40]
	vw	**-0.306** **	**0.053**	**0.186**	**0.492** ***
		[-2.30]	[0.49]	[1.55]	[2.58]

Table IA.5
Cross sectional Fama-MacBeth Regressions-Jaccard Similarity

This table reports the results from Fama-MacBeth regressions on the relation between our Cybersecurity Risk Index, as measured Jaccard similarity and subsequent stock returns (1-month). For each month of our sample we run cross sectional regressions of excess stock returns on lagged cybersecurity risk and a set of firm characteristics that are also lagged. These include beta, size, book-to-market, momentum, short-term reversal, illiquidity, coskewness, idiosyncratic volatility, asset growth, profitability and demand for lottery-like stocks (max) , length of Item 1A. Risk Factors of the Form 10-K and 10-K readability. All variables are defined in Appendix B. To facilitate the assessment of the economic significance of our findings, all explanatory variables are standardized. The coefficients are reported as time-series averages of the estimates from the cross sectional regressions. The *t*-statistics, which are reported in brackets, are based on the Newey-West heteroskedasticity and autocorrelation consistent standard errors. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

	$Returns_{t+1}$		
	[1]	[2]	[2]
Cybersecurity Risk Index (Jaccard)	**0.297 *** **	**0.107 *** **	**0.130 *** **
	[6.34]	[2.74]	[2.90]
Beta	-	**0.098**	**0.096**
		[0.90]	[0.92]
Market Value	-	**-0.079**	**-0.032**
		[-1.16]	[-0.43]
Book-to-Market	-	**0.006**	**0.008**
		[0.16]	[0.16]
Momentum	-	**0.163 ***	**0.153 ***
		[1.85]	[1.74]
Reversal	-	**-0.113 ****	**-0.120 ****
		[-2.01]	[-2.17]
Illiquidity	-	**-0.019**	**-0.019**
		[-0.53]	[-0.52]
CoSkew	-	**-0.027**	**-0.025**
		[-0.88]	[-0.79]
Indiosyncratic Volatility	-	**-0.456 *** **	**-0.456 *** **
		[-5.42]	[-5.59]
Asset Growth	-	**-0.122 *** **	**-0.115 ****
		[-3.14]	[-3.10]
ROA	-	**0.296 *** **	**0.285 *** **
		[5.30]	[5.14]
Max	-	**-0.421 *** **	**-0.414 *** **
		[-4.43]	[-4.45]
Risk Section Length (ln)	-	-	**-0.040**
			[-0.99]
Readability (ln)	-	-	**-0.082**
			[-1.47]
Constant	**0.515**	**0.515**	**0.513**
	[1.14]	[1.13]	[1.12]
Observations	409,016	342,573	342,573