NBER WORKING PAPER SERIES

CYBERSECURITY RISK

Chris Florackis
Christodoulos Louca
Roni Michaely
Michael Weber

Working Paper 28196
http://www.nber.org/papers/w28196

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020, Revised March 2022

We would like to thank Itay Goldstein (the editor), two anonymous referees, Kenneth Ahern (discussant), Scott Baker (discussant), Matia Bevilacqua, Emanuele Borgonovo, Leisen Dietmar (discussant), Alexandros Kostakis, Gabriele Lattanzio, Andreas Milidonis, Michael Minnis, Dewan Muktadir-Al-Mukit, Jaideep Oberoi (discussant), Panagiotis Stamatopoulos, Morad Zekhini (discussant), Guofu Zhou and conference/seminar participants at the University of Durham, University of Surrey, the 19th Annual Meeting of the Hellenic Finance and Accounting Association, the SFS Cavalcade North America 2021, the 2021 NBER Summer Institute on Big Data and High-Performance Computing for Financial Economics, the 2021 European Meeting of the Financial Management Association, the 2021 Annual Meeting of the European Financial Management Association, the Global Virtual Seminar Series on Fintech (Cybersecurity Day) organised by Center for Financial Markets and Policy (CFMP) at Georgetown University and the 2021 Annual Meeting of the American Finance Association. Weber gratefully acknowledges financial support from the University of Chicago Booth School of Business and the Fama Research Fund. Send correspondence to Michael Weber, Michael.Weber@chicagobooth.edu. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Cybersecurity Risk
Chris Florackis, Christodoulos Louca, Roni Michaely, and Michael Weber
NBER Working Paper No. 28196
December 2020, Revised March 2022
JEL No. G14,G31

ABSTRACT

Using textual analysis and comparing cybersecurity-risk disclosures of firms that were hacked to others that were not, we propose a novel firm-level measure of cybersecurity risk for all US-listed firms. We then examine whether cybersecurity risk is priced in the cross-section of stock returns. Portfolios of firms with high exposure to cybersecurity risk outperform other firms, on average, by up to 8.3%per year. At the same time, high-exposure firms perform poorly in periods of high cybersecurity risk. Reassuringly, the measure is higher in information-technology industries, correlates with characteristics linked to firms hit by cyberattacks, and predicts future cyberattacks.

Chris Florackis
University of Liverpool
Chatham Street, L69 7ZH
Liverpool L69 7ZH
United Kingdom
c.florackis@liverpool.ac.uk

Roni Michaely
University of Hong Kong
School of Business and Economics
Hong Kong
China
rm34@cornell.edu

Christodoulos Louca
Archiepiskopou Kyprianou 30
Limassol 3036
Cyprus
christodoulos.louca@cut.ac.cy

Michael Weber
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
michael.weber@chicagobooth.edu

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. Examples of cybersecurity risk include the risk of losing sensitive data, disruption in a firm's network, systems, and services, and physical electronic damage. Firm executives and market participants in advanced economies currently consider cybersecurity risk one of the top global concerns (WEF The Global Risks Report 2020), which is not surprising given the rapid increase in major cyberattacks in recent years. Despite substantial investments in information security systems, most firms remain highly exposed to cybersecurity risk.[1] In addition to being direct targets, many firms are indirectly affected or are collateral damage in a cyberattack. For example, the adverse effects of tactical cyber operations against SolarWinds, a major U.S. information technology firm, in 2020 went beyond the direct target and propagated to many of its client organizations, including several large U.S. federal agencies, in what was one of the largest and most sophisticated attacks ever. Considering the profound impact of cyberattacks on firms and economies around the world, it is important to have a deeper understanding of individual firms' exposure to cybersecurity risk, its quantification, and its effects on asset prices.

In this study, we propose a novel firm-level measure of cybersecurity risk for all listed firms in the U.S. and examine whether exposure to 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, on average, 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, standard return predictors do not subsume it in Fama-MacBeth regressions, and firms in specific industries do not drive the return premium. A cybersecurity-mimicking portfolio

[1] Gartner, a global research and advisory firm, for example, estimated a worldwide spending on information security products of $124 billion in 2019, representing an increase of 8.8% relative to 2018. In 2020, Steve Morgan, Founder of Cybersecurity Ventures, predicted that cybercrime damages could grow by 15 percent per year, to reach US$10.5 trillion annually by 2025. See "Cybercrime To Cost The World $10.5 Trillion Annually By 2025" (Cyber Magazine, Nov. 13th, 2020).

performs poorly in times of heightened cybersecurity risk and investors' concerns about data breaches, suggesting a risk-based explanation.

Our measure builds on two ideas: First, firms that are actually attacked are more vulnerable to cyberattacks ex-ante, and express this heightened risk ex-ante in their risk disclosures; and second, firms that have similar levels of cybersecurity risk use similar words to describe their risk exposure and exposure management. To construct our measure, we use firms that were subject to cyberattacks as a training sample and then compare the wording and language in the relevant parts of the risk-disclosure section in annual reports of the attacked firms with those of all other firms. Specifically, over the period 2007-2018, we first extract the discussion on cybersecurity risk in the "Item 1A. Risk Factors" section from firms' 10-K, which contains information about the most significant risk factors for each firm. Second, we identify a sample of firms that have been subject to a major cyberattack in any given year. These firms, which likely had a high prior exposure to cybersecurity risk, given the realization of a hack, 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). The higher the measured similarity in cybersecurity-risk disclosure for our sample firms and firms in the training sample, the greater is the exposure to cybersecurity risk.[2]

We validate our measure in several ways. First, firms that score high on the measure emphasize cybersecurity risk in their 10-K filings more than firms with low scores. For instance, firms with high measured exposure typically mention that the increasing sophistication of hackers makes defending against cybersecurity attacks difficult, despite investments in preventive systems. Firms with low measured exposure instead tend to

[2] Other studies that use document similarity to extract meaning from text collections include Hoberg and Phillips (2010; 2016), Brown and Tucker (2011), Hoberg and Maksimovic (2015), Lang and Stice-Lawrence (2015), and Lowry, Michaely, and Volkova (2020).

emphasize they can adequately deal with cybersecurity risk through preventive measures. Moreover, these low-exposure firms typically do not devote a separate section to cybersecurity risk in their 10-Ks.

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 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 insurance to partially protect against claims that may arise due to cyberattacks.

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. This trend is consistent with the recent growth in the number and significance of successful cyberattacks against major organizations, as well as firms' increasing vulnerability to cyberattacks.

Fifth, our measure is particularly high in industries that rely heavily on information technology systems to perform their operations, which makes them more vulnerable to cyberattacks (e.g., the Telephone & Television Transmission, Business Equipment, and Money Finance industries). According to our calculations, these industries exhibit a high cyberattack incident rate (see also Romanosky 2016).[3]

[3] Our measure also correlates with firm characteristics that previous research has linked to firms hit by cyberattacks. For example, in line with Kamiya et al. (2021), our measure relates cross-sectionally with firm characteristics such as size, age, profitability, growth opportunities and tangibility. It is also positively associated

Finally, and most directly, we show that firms with higher cybersecurity-risk scores are more likely to experience a future cyberattack. In economic terms, a one standardized unit increase in our cybersecurity-risk score increases the probability of a future cyberattack by 92.70%. Taken together, our firm-level measure of cybersecurity risk accurately reflects features that one would expect for firms exposed to the risk of cyberattacks.[4]

In the second part of the study, we use our measure to examine 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 exposure 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 sub-samples of stocks sorted by size, book-to-market, profitability, institutional ownership, illiquidity, idiosyncratic volatility, risk-section length, and 10-K readability. We further show that firms in industries that performed well during our sample period do not drive our results, in particular innovative firms and those with high R&D expenditure that have been shown to earn higher abnormal returns (see e.g., Li 2011; Hirshleifer, Hsu, and Li 2013, 2018). We also show the excess returns of high versus low cybersecurity-exposure stocks is larger when we exclude firms that partially insure against cyberattacks.

with other characteristics that likely indicate vulnerability to cyberattacks such as R&D expenditures and the presence of trade secrets.

[4] We discuss additional validation tests below.

We then 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 of the exposure to cybersecurity risk is not a short-term phenomenon.

Furthermore, we introduce a cybersecurity-risk factor and test its economic and statistical significance. If our measure captures cybersecurity risk and 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 when cybersecurity-risk concerns materialize. To perform the analysis, we resort to daily search volume index (SVI) data and identify days of increasing (abnormal) attention to cybersecurity risk. Consistent with the view that cybersecurity risk has a significant systematic component, 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.

Finally, we exploit a recent large-scale cyberattack providing *out-of-sample* evidence on both the validity of our measure and the effect of cybersecurity risk on stock prices. Specifically, we focus on the supply-chain cyberattack against SolarWinds, which was disclosed in an SEC filing on December 14th, 2020.[5] Firms with higher *ex-ante* cybersecurity-risk scores based on our measure exhibit negative cumulative abnormal returns around the SolarWinds hack. We also exploit the nature of the attack (supply-chain attack) and distinguish between SolarWinds' customers (affected firms) and non-customers (non-affected firms). We

[5] A supply-chain attack is a cyberattack that damages the target as well as other organizations in the target's supply chain. See Crosignani, Macchiavelli, and Silva (2021) for a recent study on the propagation of cyberattacks through firms' supply chains.

find that our *ex-ante* cybersecurity risk measure is positively associated with the probability of being in the group of affected firms.

1. Related Literature and Contribution

This study is related to several strands of the literature. First, it is related to a growing literature extracting important economic information utilizing text as data (e.g., Loughran and McDonald 2016; Neuhierl and Weber 2019; Lowry, Michaely, and Volkova 2020). Hassan et al. (2019) and Sautner et al. (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 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) use the management discussion and analysis section to obtain measures of financial constraints. In a similar spirit, 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. Likewise, Israelsen (2014) extracts a set of risk factors from risk-factor disclosures and shows that these factors are informative about stock return volatility and factor loadings. Lopez-Lira (2020) uses topic modelling of risk-factor disclosures to elicit risk factors, and evaluates which ones are systematic and priced in the cross-section of stock returns. Finally, Hanley and Hoberg (2019) develop a new approach that crowdsources signals about emerging risks in the financial sector from both investors and banks' risk-factor disclosures. We add to this literature by focussing on cyber-related risk

disclosures and examine whether these convey useful information about firms' exposure to cyber threats and the associated costs, rather than focusing on overall risk exposure.

Second, we add to the asset-pricing literature by showing that cybersecurity risk is priced in the cross-section of stocks (Gu, Kelly, and Xiu 2020; Freyberger, Neuhierl, and Weber 2020). Stocks of firms highly exposed to cybersecurity risk earn higher expected returns and co-move with other high-exposure firms. This finding is consistent with the view that at least a portion of cybersecurity risk is priced as a systematic risk factor and investors require a premium to hold stocks exposed to high cybersecurity risk. This result also alleviates the concern that we simply capture differences in realized returns rather than expected returns given our relatively short sample period.

Finally, we directly add to the literature focusing on the implications of cyberattacks on the attacked firms. For example, several studies focus on the impact of cyberattacks on firm valuation (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 2021); other studies focus on how firms adjust their financial, investment, governance, and risk-management policies following costly cyberattacks (see, e.g., Akey, Lewellen, and Liskovich 2020; Ashraf forthcoming; Boasiako and Keefe 2021; Kamiya et al. 2021). Rather than focusing only on attacked firms, we analyse cyber-related disclosures for the population of U.S. traded firms and assess their cybersecurity-risk exposures.

Most closely related are contemporaneous studies by Jiang, Khanna and Yang (2020) and Jamilov, Rey and Tahoun (2021). Jiang, Khanna and Yang (2020) apply a variety of machine learning techniques including logistic LASSO regressions to estimate the ex-ante likelihood that a firm will experience a cyberattack. In a similar spirit to our study, they then examine and find that their measure is related to stock returns. In addition, they find that institutional investors tend to sell stocks with high cybersecurity risk and buy those with low cybersecurity

risk. Methodologically, we differ in that we focus on measuring the similarity of cybersecurity-risk disclosure with past disclosures of firms in a training sample. The use of a training sample, which by construction includes firms with ex-ante high cybersecurity risk, enables us to capture systematic exposure to cybersecurity risk. Our measure is able to address concerns regarding firms' tendency to borrow disclosure language from their peers, and, more generally, with factors that influence firms' disclosure practices. Jamilov, Rey and Tahoun (2021) focus on quarterly earnings calls to construct a text-based measure of cybersecurity risk. They find that exposure to cybersecurity risk has direct and contagion effects on stock returns. We focus on cybersecurity-risk disclosures in "Item 1A. Risk Factors" section of 10-Ks, which are mandated by SEC regulations, and hence require firms to provide an accurate description of the most significant risks they are exposed to. Importantly, our approach avoids potential biases arising from the idiosyncratic nature of the questions and answers during earnings conference calls, and the ambiguity of the language used in them.[6]

2. Data and Methods

2.1 Data Sources

We combine several databases to construct our sample. We use the Center for Research in Security Prices (CRSP) files to obtain stock returns, Standard and Poor's Compustat database 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) data to identify cyberattacks.[7] We use Factiva to cross-reference the information from PRC and distinguish cyberattacks that attracted the attention of global news outlets or were covered in major

[6] For example, Dzieliński, Wagner and Zeckhauser (2017) show that many CEOs and CFOs are "vague talkers" (i.e., they commonly use qualifying words indicating uncertainty), which might diminish the information content of earnings news.

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

newswires. In addition, we perform Bloomberg searches and follow the methodology of Ben-Rephael, Da, and Israelsen (2017) to construct a stock-level measure of abnormal institutional investor attention. We use Google Trends to obtain Search Volume Index (SVI) data and identify days of heightened attention to cybersecurity risk. We also use the FactSet Revere Supply Chain Relationships database to identify firms' customers. We finally use data on the stock and flow of patents from the Duke Innovation & Scientific Enterprises Research Network (DISCERN) database by Arora, Belenzon, and Sheer (2021a, b).

2.2 Cybersecurity-risk Disclosures

Given the growing dependence of firms on information technology to perform their operations, the risk associated with cybersecurity has increased over time. According to SEC Regulation S-K Item 305, firms must provide information on how cybersecurity risk affects their operations in the "Item 1A. Risk Factors" section in their 10-Ks. Regarding material cybersecurity risks and incidents in particular, the SEC issued specific guidelines in 2011 and 2018, instructing public companies to inform their investors 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 attacked.

We use a web-crawling algorithm to download all "10-K," "10-K405," and "10-KSB40" filings, excluding amended documents, from SEC Edgar and extract the fiscal year and the central index key (CIK) from each filing. In addition, we extract the cybersecurity-risk disclosures from the "Item 1A. Risk Factors" section. To do so, we first compile and use a list of keywords/phrases, such as "unauthorized access", "attack" and "hacker", which directly describe cybersecurity risk. Then, to alleviate potential noise arising from the use of

keywords/phrases in parts of the disclosures other than those relating to cybersecurity risk, we require the presence/absence of additional relevant/irrelevant hits 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". This helps us to extract sentences that contain a direct description of cybersecurity risk.

Firms, however, may also use indirect descriptions that relate to cybersecurity risk. For instance, they may describe their business, security measures, and the potential consequences of a cyberattack; we categorize this description into internal, legal, and economic consequences and compile another list of indirect keywords/phrases to retrieve the relevant sentences. To ensure our algorithm retrieves only sentences related to cybersecurity risk, we first require the presence of a sentence with a direct cybersecurity-risk discussion. Then, because the discussion often clusters in paragraphs, we search the subsequent 10 sentences to find indirect keywords/phrases. Appendix A provides detailed information on how we extract cybersecurity-risk disclosures and several examples.

Note that some firms do not have an "Item 1A. Risk Factors" section, and thus, are excluded from the sample; 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 Factors" section by reference.[8] Finally, we link each firm's cybersecurity-risk disclosures with Compustat using the fiscal year, the CIK, and the mapping table from the WRDS SEC Analytics suite.

Table IA.1 of the Internet Appendix provides descriptive statistics of the evolution of cybersecurity-risk disclosures. Starting in 2007, the first year of our sample, a non-negligible

[8] One example of such a 10-K is the following:
https://www.sec.gov/Archives/edgar/data/72971/000007297114000337/wfc10k_20131231.htm

number of firms have cybersecurity-risk disclosures (28.75% of the firms in our sample). We note a modest increase in the percentage of firms with cyber-related disclosures over the period 2007 to 2010, which is followed by a significant increase in years 2011 and 2012, driven by the SEC's 2011 disclosure guidance on cybersecurity matters and cyber incidents. By 2012, more than 66% of U.S. firms (up from 39% in 2010) discuss their exposure to cybersecurity risk in their 10-Ks. Therefore, the process of disclosing cybersecurity risk largely reflects increased risk and disclosure obligations driven by regulators. By 2018, about 90% of U.S. firms include cybersecurity risk as part of their discussion in "Item 1A. Risk Factors" section.[9]

2.3 Training Sample

Developing a firm-level measure of cybersecurity risk through cybersecurity-risk disclosures in "Item 1A. Risk Factors" is challenging, because cyber-related disclosures usually consist of discussions of the firm's business operations, its exposure to cybersecurity risk as well as explicit measures and remedies to "manage" (i.e., reduce) cybersecurity risk and the resulting litigation risk from potential cyberattacks. Hence, purely focusing on the length of the cybersecurity disclosures is unlikely to accurately capture firms' exposure to this type of risk. Moreover, purely relying on individual firms' disclosures might capture firm-specific, non-systematic risk that might not be priced by financial markets. To alleviate such concerns, we utilize a training sample, which isolates firms that have been subject to a major cyberattack. Ex-ante, firms in the training sample likely had a high exposure to cybersecurity risk before actually being attacked; therefore, by estimating the similarity of each firm's cybersecurity-risk disclosure with past cybersecurity-risk disclosures of firms in the training sample, we aim to capture exposure to cybersecurity risk. In addition, utilizing a training sample allows us to

[9] In addition to the increase in the number of firms with cyber-related disclosures, the volume of disclosures and overall cybersecurity awareness has been increasing over time (see e.g., Berkman et al. 2018).

distil common elements of the discussions of cybersecurity risk and therefore to focus on a more systematic component of risk.

To construct the 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 affected. We exclude incidents involving governments, educational institutions, and non-profit organizations, and analyse only cyberattacks that involve lost personal information by hacking or malware-electronic entry by an outside party. We collect information on all recorded cyberattacks, and manually search news articles from Factiva to cross-reference the information and to identify which cyberattacks 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 cyberattacks "major" and use them as our training sample. Using only such "major" attacks ensures that we use information that is widely disseminated and available to investors (nevertheless, we repeat our measurement using all incidents of cyberattacks as the training sample, and the results are unchanged). Out of a total of 175 cyberattacks we identify during the period 2005-2018 with available cybersecurity-risk disclosures in "Item 1A. Risk Factors" section, we classify 69 as "major" cyberattacks; these cyberattacks correspond to 54 firm-year cyberattacks (a firm may experience more than one cyberattack in a given year). The first cyberattack occurred in 2006, which explains the start of our sample in 2007. Finally, we manually link the names of these firm-year cyberattacks in the PRC database with firm names in CRSP and Compustat.

2.4 Cybersecurity-risk Measure

We use the textual information on cybersecurity risk in "Item 1A. Risk Factors" section 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 in our training

sample. The idea behind the measure is that firms that use similar words to describe their risk exposure and exposure management, exhibit similar levels of cybersecurity risk. This approach has a long history in information processing and has become popular in finance and economics. For instance, Hoberg and Phillips (2016) estimate product-market language similarity between firms to develop a novel definition of industry, and Hoberg and Maksimovic (2015) estimate the similarity of firms' liquidity and capitalization resources relative to a training set of financially constrained firms.

We construct our measure as follows. 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 using 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 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 were subject to cyberattacks during the one-year period ending at the firm's filing date (training sample).[10] 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 to a cyberattack (i.e., for each firm and year, we have N_{t-1} such similarities). Cosine similarity is

[10] If no cyberattack occurs in the previous one-year period, we look for cyberattacks in the previous two-year period.

the cosine angle between two text vectors, whereas Jaccard similarity is 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 between (0, 1), and greater values imply a larger overlap between the two vectors of words (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_{t-1} similarities:

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

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

3. Validation

We now discuss the resulting measures and their properties to verify that they likely capture exposure to cybersecurity risk.

3.1 Excerpts from Cybersecurity-risk Disclosures

Table 1 compares excerpts from cybersecurity-risk disclosures from firms with high and low cybersecurity-risk scores to gain some intuitive understanding of our measure and how firms differ in their description of cybersecurity risk. Firms with high scores extensively discuss risk in their 10-Ks (Panel A); for instance, Walgreens Boots Alliance Inc, the firm with the highest score, acknowledges that 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.

Instead, the discussions of firms with low scores are quite different (Panel B). For example, Weyerhaeuser Co., the firm with the lowest score, mentions that its service providers and the firm itself employ *adequate* security measures, whereas other firms simply discuss cyberattacks in conjunction with other risks. Overall, firms with low scores believe that, through preventive measures, they can adequately deal with cybersecurity risk, and that cybersecurity risk is not important enough to warrant explicit and separate discussions.

In summary, these findings suggests firms with high values of our measure indeed discuss cybersecurity-risk threats extensively in their risk disclosures, whereas firms with low scores manage these risks and threats 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-disclosures. Intuitively, we would expect firms facing a higher threat of cybersecurity risk to devote more space discussing these risks than 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 total 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 being a more important source of risk than other types of risks.

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. Managers with higher exposure to cybersecurity risk will likely communicate their concerns to shareholders, to lower their 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.[11]

Further, we expect firms with higher cybersecurity-risk exposure to actively manage their exposure through real actions. One such real action is to purchase 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 above median cybersecurity-risk scores. We read all disclosures in which the word "insurance" appears and find that almost all these firms mention that their insurance policy only partially protects them against claims that may arise due to cyberattacks. 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".

Overall, these associations provide additional evidence that our measure likely 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 cyberattacks per year. The figure shows a positive time trend, especially after 2011, when the SEC issued the first cybersecurity-disclosure requirements. In addition, whereas 49.03% of the firm-years exhibit zero cybersecurity risk in 2011, only 10.59% do so in 2018. Overall, this

[11] The variables Litigious words, Precise words, and Negative words are defined as relative word counts; that is, the ratio of the count of certain words to total words in the cybersecurity-risk disclosures.

period is characterized by increasing cybersecurity risk, originating from the large number of successful cyberattacks 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 cyberattacks per year is 0.72, suggesting the time-series properties of our measure align well with the number of cyberattacks.

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, which makes them more vulnerable to cyberattacks. Indeed, 125 cyberattacks (71.4% of the total) occur in these industries. Firms in more "traditional" industries, such as Energy, Oil and Gas, and Manufacturing exhibit lower cybersecurity risk and fewer cyberattacks.

Taken together, both the time-series and industry variations of our measure intuitively relate to cybersecurity risk: the average exposure and the number of firms exposed to cybersecurity risk increase 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 define all variables in Appendix B. We winsorize the continuous variables in the sample at the 1st and 99th percentiles (by year) to mitigate the impact of outliers. The sample covers the period 2007-2018 and consists of 5,534 firms with 35,308 firm-year observations. 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

issuing such disclosures only after the SEC specific guidelines in 2011, the 25th percentile is zero.

We then run linear regressions to examine how our measure relates to firm, industry, 10-K, and corporate-governance characteristics in Table 4. 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 of possible boilerplate or generic cybersecurity-risk disclosures that could lead to highly sticky scores across time for the same firm. We cluster standard errors 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 development (*R&D Expenditures*). Naturally, such firms are likely more vulnerable to cyberattacks. Our measure also relates to 10-K and corporate-governance characteristics. Firms with higher scores have lengthier "Item 1A. Risk Factors" sections (*Risk Section Length (ln)*) and less readable 10-Ks (*Readability (ln)*). 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 that firms with better governance are also pre-emptively more active in attempting to understand, report, and manage their risks, including the risk of litigation and cyberattacks.

Overall, these results indicate our measure is related to characteristics of firms that were attacked (see Kamiya et al. 2021).

3.5 Firm Outcomes

Finally, we investigate whether our measure 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 an actual attack, resulting in negative stock returns. A negative stock market reaction may occur even in the absence of a cyberattack; that is, negative returns may occur for high cybersecurity-risk firms in times of heightened concerns over data breaches for various reasons (e.g., regulatory disclosure changes etc.). Accordingly, we expect firms with high cybersecurity risk should have negative asymmetries in stock returns. To explore this possibility, we estimate a linear regression in which 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. We indeed find our measure positively correlates with *NCSKEW* (Model 1 of Table 5). As a robustness test, we use extreme sigma (*EXTR_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 (Andreou, Louca, and Petrou 2017). We find a positive and statistically significant association between our measure and *EXTR_SIGMA* (Model 2).

Our next test focuses on whether our measure forecasts future cyberattacks. We estimate a logit regression in which the dependent variable equals 1 if a firm experiences a cyberattack 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 in Table 6. In Panel A, we focus on all cyberattacks reported in the PRC database for which we have complete data. Model 1 adds only year and industry fixed effects and Model 2 adds controls. Furthermore, it includes an indicator variable of whether a firm was attacked before (*Previous Attack Dummy*), which controls for the fact that 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 cyberattack. In terms of economic importance, a one standardized unit increase in our measure increases the probability of a cyberattack by 92.70%. This predictability is reassuring and provides direct evidence that our measure reliably captures firms' exposure to cybersecurity risk. As a robustness test, we repeat the analysis focusing on major attacks in Panel B and non-major attacks in Panel C of Table 6 and confirm the predictability. Notably, firms with no major attacks are not part of the training sample. Therefore, this analysis provides "out-of-sample" evidence of the predictability of future cyberattacks.

4. Cybersecurity Risk and Stock Returns

In the previous section, we show our measure is correlated with both language and real actions that likely relate with exposure to cybersecurity risk. In addition, the measure displays intuitive time-series, industry, and firm characteristics that are associated with the probability of cyberattacks. Consistent with these results, our measure is significantly associated with negative asymmetries in stock returns and predicts future cyberattacks. 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 investors may require higher expected returns from firms with high exposure to cybersecurity risk. We first use univariate portfolio sorts to examine the return difference of firms with high and low exposure. Second, we conduct bivariate portfolio sorts to better understand whether exposure to cybersecurity risk is prevalent across certain subsamples of stocks. Third, we run Fama-MacBeth (1973) cross-sectional regressions to ensure we are not simply capturing exposure to other well-known risk factors and characteristics predicting returns in the cross-section. Finally, we investigate the time-series

variation of a cybersecurity-risk factor and provide further out-of-sample evidence from the SolarWinds hack.

4.1 Univariate Portfolio-level Analysis

We implement the portfolio analysis as follows. We first assign firms into terciles 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 consists 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. Our objective is to test whether stocks in Portfolio 3 earn higher expected returns than those in Portfolio 1. For our baseline analysis, we start in December 2007 and construct portfolios at the end of each quarter. We then track the performance of the three portfolios and compute monthly returns in excess of the risk-free rate over the period March 2008 - March 2019. We calculate both equal-weighted (ew) and value-weighted (vw) monthly portfolio returns. We report average excess returns as well as alphas adjusted for market risk (CAPM alphas), for market, size (SMB), value (HML), and momentum (MOM) (FFC alphas), as well as alphas adjusted for market, size, value, profitability (RMW) and investment (CMA) (five-factor alphas).

The results are presented in Table 7. The average excess 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 0.68% between the two extreme portfolios. The difference is statistically significant at the 1% level with a Newey-West *t*-statistic of 4.56.[12] The 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 market factors, Fama-French-Carhart (FFC) factors, and Fama-French (2015) five factors does not

[12] We use 12 lags for the calculation of standard errors. Our results are stronger when we use fewer lags, such as six or four.

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.[13]

Panel B of Table 7 reports the average portfolio characteristics in each cybersecurity-risk portfolio. Specifically, we present information on 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 et al. 2006), and 10-K characteristics such as the length of the "Item 1A. Risk Factors" section and the complexity of 10-K disclosures (You and Zhang 2009; Lehavy, Li, and Merkley 2011). Portfolio 1 includes on average a larger number of stocks than Portfolio 3, because a non-negligible number of firms in the earlier period of our sample have no cybersecurity-risk disclosures in their "Item 1A. Risk Factors" section in 10-Ks . Note that firms may not report cybersecurity-risk disclosures, because (i) they simply have no such risk concerns, (ii) of low awareness of cybersecurity risk, or (iii) of poor disclosure practices. In section 6, we explicitly address these possibilities through several robustness tests and show firms with no cybersecurity-risk disclosures do not drive our results.

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

[13] We also examine the relation between cybersecurity risk and traditional risk factors. We estimate the average monthly correlations of our measure with the factor loadings on each of the Fama and French (2015) factors, constructed using rolling firm-level regressions of monthly returns over the previous 60 months. The average monthly correlations of our measure with the factor loadings are quite low; the correlations with market, size, value, profitability and investment factor loadings are -0.034, -0.017, 0.026, 0.048 and 0.033, respectively, suggesting that traditional factors cannot fully capture the premium that high cybersecurity stocks earn.

firm in Portfolio 3 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. These differences motivate us to conduct bivariate portfolio sorts 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

We now perform double sorts. 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 in 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 10-K readability. The intersection of the two independent sorts yields several double-sorted portfolios. We track the performance of these portfolios over the following quarter and report results in Table 8. Specifically, we directly report the excess returns of high- versus low-cybersecurity-risk portfolios within each subsample. The higher returns of high-cybersecurity-risk stocks exist in all subsamples of stocks and remains statistically significant in the vast majority of cases. These results ensure our findings are not confined to a small subsample of stocks, and alleviate concerns that exposure to cybersecurity risk largely captures other well-known risk proxies.

4.3 Cross-sectional Regressions with Individual Stocks

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 returns on lagged cybersecurity-risk exposures and additional 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 et al. 2006), R&D expenditures (Hirshleifer, Hsu, and Li 2013), 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" section and the readability of the 10-K filings. 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 include only our measure of cybersecurity risk in the regressions and find a time-series average of the cross-sectional slope of 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 firm-level characteristics, and, in Model 3, we additionally control for the length of "Item 1A. Risk Factors" section and the readability of 10-K filings. The results show the coefficient on cybersecurity risk remains positive and significant, although the magnitude of the 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 measure. 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 merely a short-term phenomenon.[14]

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

So far, we have documented that stocks more exposed 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 the higher returns of stocks with high exposure to cybersecurity risk is due to a compensation for that risk, we should find that high cybersecurity-risk stocks perform poorly and significantly worse than low cybersecurity-risk stocks on days of increased attention toward, and concerns about cybersecurity risk. To test this conjecture, we first form a simple cybersecurity-risk factor following 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). We then independently sort all stocks into several groups based on our cybersecurity-risk measure.[15] The cybersecurity-risk factor 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.

For the analysis of the time-series variation of the factor, we calculate daily returns of the factor over the period March 2008 to March 2019. We are interested in examining the performance of the cybersecurity-risk factor, especially during days of increased attention toward cybersecurity risk. We identify these days based on abnormal SVI search volume in Google Trends. SVI measures the intensity on "search terms" or "search topics" during a time

[14] The main text presents results based on cosine similarity as a proxy for cybersecurity risk. For robustness purposes, we rerun the main analyses of Sections 3 and 4 using Jaccard similarity and find qualitatively similar results (see Tables IA.2, IA.3, IA.4, IA.5 and IA.6 in the Internet Appendix). Hence, our findings are not sensitive to the method we use to measure the degree of similarity in cybersecurity-risk disclosures.

[15] For robustness purposes, we independently sort stocks into three, five and ten groups based on our measure. The benchmark results are based on five groups. The results are weaker (stronger) when we use three (ten) groups.

period and is 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 these potentially capture attention more comprehensively. We identify the following relevant topics: "hacker", "data breach", "cyberattack", "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, we use the topics "hacker" and "data breach".

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.

Given that Google Search Trends provide daily data only for a query period shorter than 9 months, we construct the variable *"High_Google_SVI_dummy"* as follows: First, we download a series of daily SVI data that overlap with each other for 100 days. We then rescale the datasets using the data of the overlapping window; the rescaling enables us to create a dataset of daily SVI that covers our sample period. We estimate daily abnormal SVI by scaling each daily SVI with the median SVI estimated during the past 2 weeks to adjust for seasonality. We then define extreme attention days as days when the daily abnormal SVI is greater than the mean abnormal SVI plus *n* standard deviations, both estimated during the past 2 weeks (on a rolling basis each

day).[16] For robustness purposes, we report results for both n=1.5 (benchmark results) and n=2 (robustness check). We also report results by changing the look-back window from 2 weeks to 4 weeks. We present results for the *"High_Google_SVI_dummy"* using both "hacker" and "data breach" topics jointly. However, the results are similar if we examine independently these topics.

We present the results in Table 10. Panel A presents results from our benchmark specification in which the cybersecurity-risk factor is based on five cybersecurity-risk portfolios. In Panel A, we set *"High_Google_SVI_dummy"* to 1 on days on which the daily abnormal SVI is greater than the mean abnormal SVI plus 1.5 standard deviations, both estimated during the past 2 weeks, and zero otherwise. Model 1 includes *High_Google_SVI_dummy* as the only explanatory variable. Model 2 controls for market risk (CAPM specification); in Model 3, we add the size, value and momentum factors (FFC specification), and Model 4 controls for the five Fama-French risk factors (Fama-French five-factor specification). The cybersecurity-risk factor exhibits positive returns, on average, over the sample period; the daily estimate for the constant term α is positive (at 0.0002, which implies an annualized return of about 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 about cybersecurity risk.

For robustness purposes, we also report in Panel B results from a cybersecurity-risk factor that is based on ten cyber risk portfolios. In Panel C, we re-estimate daily abnormal SVI by scaling each daily SVI with the median SVI estimated during the past 4 rather than 2 weeks. In Panel D, we re-define the *High_Google_SVI_dummy* to equal 1 on days on which the daily

[16] We thank an anonymous reviewer for recommending this procedure to construct the variable "High_Google_SVI_dummy".

abnormal SVI is greater than the mean abnormal SVI plus 2 (instead of 1.5) standard deviations. In all cases, the results remain similar. Finally, in Panels E and F, we re-estimate equation [3] after replacing the variable *High_Google_SVI_dummy* with the variable *High_Google_SVI_dummy + 5 days* and *High_Google_SVI_dummy + 1 month*, which moves the event window a trading week and month, respectively, 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.

Overall, these results suggest firms with high exposure to cybersecurity risk earn high returns on average, but they perform poorly on days with heightened concerns about cybersecurity. We interpret these results as evidence that cybersecurity risk has a significant systematic component; accordingly, the premium that high-cybersecurity-risk stocks earn compensates investors for holding high-cybersecurity-risk stocks, which significantly underperform in times of heightened cybersecurity risk and investors' concerns about data breaches.

5. SolarWinds Hack: Out-of-Sample Evidence

In this section, we exploit one of the largest and most sophisticated attacks ever, to provide out-of-sample evidence on the effect of cybersecurity risk on stock prices.[17] Specifically, we focus on SolarWinds, a U.S. information technology firm, which experienced a hack that leveraged cloud-based services to compromise the company itself as well as many organisations in its supply-chain (including several large U.S. federal agencies).[18] Figure 3

[17] We thank an anonymous referee for suggesting this test. The hack occurred after the initial submission of the manuscript.

[18] Beyond supply-chain attacks, there are other types of cyberattacks whose impact damages not only the target firm, but also many other related or, in some cases, unrelated organizations including suppliers, clients, banks and insurers. These include attacks that take down the web, attacks that hit power grids and others that simultaneously affect multiple industries. Examples of cyberattacks with "systematic" effects include the Heartland Payment Systems data breach in 2009, the Target Corporation data breach in 2013, the distributed denial-of-service (DDoS) attack on Dyn in 2016, the WannaCry and NotPetya attacks in 2017 and the ransomware attack against Colonial Pipeline in 2021.

presents key facts and the timeline of the attack between December 13th and December 22nd, 2020. SolarWinds disclosed the breach in an SEC filing on December 14th, 2020, which caused its stock price to collapse.

We identify December 14th as our event date and perform several tests around it. We first calculate cumulative abnormal returns (CAR[-1,+1] and CAR[-1,+3]) around the event date for all firms in our sample. We then check whether our cybersecurity risk measure is correlated with these CARs. Before discussing the results, we want to stress that this out-of-sample test is quite challenging empirically, given that, for many indirectly affected firms, the market did not know of their actual exposure to the hack until much later. We find in Panel A of Table 11 that firms with different ex-ante cybersecurity scores (measured in 2018 or 2017, if unavailable in 2018) perform differently around the SolarWinds hack. More specifically, while the least exposed firms exhibit positive CARs, firms highly exposed to cybersecurity risk exhibit negative CARs. The differences in CARs across the two portfolios are negative and statistically significant (i.e., at -1.5% [p=0.00] for the case of CAR[-1,+1]).

In Panel B of Table 11, we run regressions to test whether companies with *ex-ante* high exposures to cybersecurity risk experience more negative returns around the time the SolarWinds hack was announced through an SEC filing. As dependent variables, we use CAR[-1,+1] and CAR[-1,+3]. The main explanatory variable is our *ex-ante* measure of cybersecurity risk. In addition to a continuous variable, we use two dummy variables to identify firms highly exposed to cybersecurity risk. These are: *High Cyber Risk Dummy 1*, which takes the value of 1 for firms in the top tercile of the distribution, and 0 otherwise; and *High Cyber Risk Dummy 2*, which takes the value of 1 for firms in the top decile of the distribution, and 0 otherwise. The results show highly exposed firms exhibit negative CARs around the SolarWinds hack and the negative association is more pronounced for firms that are more exposed to cybersecurity risk (e.g., see the results with the dummy variable defined based on deciles) and becomes weaker

for longer event windows. In economic terms, the underperformance of high-exposure firms relative to low-exposure firms around the disclosure of the attack (CAR[-1,+1]) ranges between -0.7% and -1.2%, depending on the definition of high-exposure firms.

We then move to the question whether our cybersecurity risk measure can predict which companies, *ex post*, were most impacted by the SolarWinds hack. We considered several ways to identify affected firms. For example, we checked all 8-Ks filed in the period following the SolarWinds incident but could not identify any firm disclosing the breach in an SEC filing. Hence, we identify affected firms as follows: Given the SolarWinds hack was a supply-chain hack originating from its Orion network management software, the hack plausibly affected SolarWinds' client organizations. We therefore conjecture that SolarWinds' customers are more likely to be among the list of affected firms. We collected the relevant data from *FactSet Revere Supply Chain Relationships*, which provides information on companies' networks and in particular their key customers, suppliers, competitors, and strategic partners, collected from annual filings, investor presentations, and press releases. To ensure we include all of SolarWinds's key customers, we look at both direct relationships (i.e., relationships disclosed by the reporting company) and reverse relationships (i.e., relationships not disclosed by the reporting company but by companies doing business with the reporting company). We managed to identify 38 U.S. listed companies for which we have complete data.

We exploit this dataset as follows: First, we examine how affected firms (i.e., SolarWinds' customers) differ from non-affected firms (i.e., non-customers) in terms of several characteristics. Panel A of Table 12 shows 64% of affected firms exhibited abnormal institutional investor attention (AIA), in any of the five trading days following SolarWinds's disclosure of the breach on Dec 14th. AIA is measured using the methodology of Ben-Rephael,

Da and Israelsen (2017) and Bloomberg searches.[19] The corresponding proportion for non-affected firms is only 37.13% and the difference is statistically significant. Affected firms also have negative CARs around the SolarWinds hack while non-affected firms have positive CARs. The difference in means is statistically significant for all CARs we consider. We also find that, based on our measure, affected firms have a higher ex-ante exposure to cybersecurity risk than non-affected firms.

Motivated by these findings, we also examine whether our ex-ante measure of cybersecurity risk predicts which firms were affected by the SolarWinds hack. In Panel B of Table 12, we estimate a logit model for the probability of being affected. The results of Model 1 show a positive and statistically significant association between our ex-ante cybersecurity-risk score and the probability of being in the group of affected firms. In Models 2 and 3, we repeat the analysis using the dummies High Cyber Risk Dummy 1 and High Cyber Risk Dummy 2 as explanatory variables. The results show a positive association between cybersecurity risk and the probability of being affected by the SolarWinds hack.

Overall, the results in this section provide out-of-sample evidence that our proposed measure captures exposure to cybersecurity risk and provide additional evidence in support of the view that cybersecurity risk is priced as a systematic source of risk.

6. Further Evidence and Robustness Tests

In this section we examine whether our results are driven by firm-level disclosure practices and check for confounding effects. We also perform a series of robustness tests.

6.1 Disclosure Practices

[19] Bloomberg's numerical attention score can take values between 0 and 4. As in Ben-Rephael, Da and Israelsen (2017), we are interested in abnormal attention, and not just the level of attention, and hence we define AIA as a dummy variable that takes a value of one if Bloomberg's daily maximum score is 3 or 4, and zero otherwise.

First, we deal with the fact that a non-negligible number of firms in the sample have no cybersecurity-risk related disclosures in their "Item 1A. Risk Factors" section in 10-Ks. This feature of the data is concentrated in the early years of the sample (i.e., 2008-2011) and results in 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, no cybersecurity-risk disclosure may also be driven by (i) low awareness of cybersecurity risk and/or (ii) poor disclosure practices.

This problem is less severe after the SEC's 2011 guidance for public-disclosure obligations with respect to cybersecurity risk and cyber incidents. Therefore, we check whether the outperformance of stocks highly exposed to cybersecurity risk remains qualitatively similar during the later period of our sample. The results in Panel A of Table IA.7 of the Internet Appendix show the excess returns and five-factor alphas for the long-short portfolio are higher than those reported in Table 7 (e.g., the value weighted five-factor alpha increases to 0.65% - up from 0.57%). As an additional test, we replace all firm-year observations 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 are qualitatively similar (see Panel B of Table IA.7 of the Internet Appendix). Furthermore, a firm's exposure to cybersecurity risk is likely persistent over time, and hence we 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 over time; therefore, given non-disclosures are concentrated in the earlier 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, we find in Panel C of Table IA.7 of the Internet Appendix that the results remain largely unaffected. 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, suggest that firm-years with no cybersecurity-related disclosures in 10-Ks do not drive our baseline findings.

Second, another potential concern with disclosure practices relates to the tendency of firms to borrow disclosure language from their peers. This potentially minimizes the cost of disclosure activity, improves judicial and regulatory assessments of risk factor disclosures, and leads to the use of language that is more likely to satisfy the external auditor (see e.g., McMullin 2016; Cazier, McMullin, and Treu 2020). Thus, one could argue firms may use similar language in their cybersecurity-risk disclosure not because they have a similar exposure to cybersecurity risk but simply because they operate under the same environment (e.g., industry) or because they have the same auditor.

We conduct additional analyses to examine whether common disclosure language affects our similarity measure. First, under the assumption that borrowing language is more common within rather than across industries (see Cazier, McMullin, and Treu 2020), we identify all firms that belong to the same Fama-French 48 industry as firms in the training sample (i.e., peer firms), exclude them from the analysis, and rerun the portfolio tests using only scores of firms that do *not* belong to the same industry as firms in the training firms.[20] The results in

[20] For example, Cazier, McMullin, and Treu (2020) argue that the regulatory review process may favor the use of standardized language that is common across industry peers, because the SEC's filing review process is conducted by staff with specialized industry expertise who review filings for multiple firms within the same industry. Hence, the use of words similar to those of industry peers may be less likely to be considered inadequate and attract further scrutiny.

Panel D of Table IA.7 of the Internet Appendix show the premium that high cybersecurity-risk stocks earn remains robust.[21]

Second, we examine whether a common external auditor is an important source of common disclosure language. Prior literature (e.g., McMullin 2016) suggests firms may borrow language from 10-Ks of firms audited by the same external auditor not only to reduce the cost of the disclosure activity, but also to increase the likelihood of auditor approval. We therefore reconstruct our measure after estimating the similarity of firm *i*'s cybersecurity-risk disclosure with past cybersecurity-risk disclosures of firms in a new training sample that *excludes* firms that have the same auditor as firm *i*. The new measure "Cybersecurity Risk Index (Excluding Peers-Auditor)" exhibits a high correlation with the original measure (at 0.953) and the results in Panel E of Table IA.7 of the Internet Appendix remain qualitatively unchanged.[22] Overall, these results rule out the possibility that our measure and results are driven by disclosures shared between peer firms, and therefore, peer-disclosure bias is unlikely to drive the main findings.

6.2 Alternative Explanations and Confounding Effects

We now examine the extent to which our cybersecurity risk measure captures industry effects. For instance, technology-intensive industries outperformed during our sample period and at the same time, they tend to have higher scores based on our measure; likewise, the Energy and Durables industries underperformed during our sample period and they tend to have lower scores based on our measure. Hence, our results might simply capture industries that performed poorly or well during our sample period rather than reflecting differences in expected returns due to a compensation for risk.

[21] We also use this sample to re-estimate our benchmark logistic regression predicting future cyberattacks. While the filter above reduces the average number of firm-years by about 18%, the results of Model 2 of Table IA.8 of the Internet Appendix show that cybersecurity risk remains a strong predictor of future cyberattacks.

[22] As shown in Model 3 of Table IA.8 of the Internet Appendix, the new measure "Cybersecurity Risk Index (Excluding Peers-Auditor)" predicts future cyberattacks and behaves in a similar manner as the original measure.

We address this alternative explanation in several ways. First, we repeat our portfolio analysis using *industry-adjusted returns.* To adjust returns by industry, we focus on the Fama-French 12 industry portfolios and their monthly average returns. The results in Panel F of Table IA.7 of the Internet Appendix remain qualitatively similar to our benchmark portfolio results when using both equal-weighted and value-weighted industry-level returns for the industry adjustment. Second, 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 G of Table IA.7 of the Internet Appendix presents the estimates and in all cases we find a positive and statistically significant premium of high cybersecurity-risk stocks. Therefore, the results are not driven by out- or underperformance of any particular industry. Third, we explore the possibility that multiple industries drive the results. Specifically, we estimate the relationship between our measure and Fama and French 12 industry alphas and we find a positive relationship that is driven by two industries with negative alphas, namely Energy and Durables. Therefore, we repeat our portfolio analysis after excluding firms that belong to these industries. As shown in Panel H of Table IA.7 of the Internet Appendix, the results remain robust. Finally, industries with positive alphas could reflect the type of technology firms invest in, or more broadly the general innovation activity firms engage in. Therefore, as an additional test, we extend our bivariate portfolio-level analysis by directly considering R&D expenditures and firm-level innovation activity (as proxied by patent flow and patent stock).[23] As shown in Table IA.9 of the Internet Appendix, the premium of high cybersecurity-risk stocks, as measured by the 5-factor alpha, remains robust across all sub-samples of stocks sorted by R&D, patent flow, and patent stock. Taken together, these results suggest our cybersecurity risk measure and the return spread associated with it, are not driven by firms in industries that

[23] As Lattanzio and Ma (2021) show, firms' exposure to cybersecurity risk is significantly associated with patent activity.

performed well during our sample or by innovative firms (i.e., those with high R&D expenditure) that have been shown to earn higher abnormal returns (see e.g., Li 2011; Hirshleifer, Hsu, and Li 2013, 2018).

We also analyse the extent to which cybersecurity-risk disclosures relate to risks other than cybersecurity risk, because our results could be an artefact of a broader risk effect. First, we explore this idea by performing a placebo test.[24] Specifically, we extract all the non-cybersecurity-risk disclosures in the "Item 1.A Risk Factors" section. Then, following a similar approach to the one we use to construct our cybersecurity risk measure, we exclude certain type of words (e.g., pronouns, conjunctions etc.) and store the text in separate word root vectors. The universe of all words in the sample is 15,452. For each firm, we populate this vector with the number of times each word appears in a firm's risk factor disclosures and estimate the cosine similarity between any two firms' disclosures. Note that, we retain the same training sample, which comprises firms that have experienced a major cyberattack. Therefore, the new cosine similarity captures a broader similarity in risk disclosures. The new measure exhibits a moderate correlation with our cybersecurity risk measure (0.41). Importantly, though, unlike our measure, it is not a consistent predictor of stock returns and future cyberattacks (see results in Table IA.10 of the Internet Appendix).

Second, we check whether other measures of risk, such as political, non-political and overall risk (Hassan et al. 2019) and exposure to climate risk (Sautner et al. 2020) affect the results. These risk measures use textual analysis of earnings conference calls. The results in Table IA.11 of the Internet Appendix show that cybersecurity risk remains a robust return predictor even after controlling for these additional types of risk.

Next, we explore the decline in the predictive power of our cybersecurity risk measure when we include additional covariates (see results in Table 9). We find that among the

[24] We thank an anonymous referee for suggesting this test.

covariates, idiosyncratic volatility (IVOL) and profitability (ROA) are the two key variables responsible for partially subsuming the explanatory power of cybersecurity risk. The results presented in Table IA.12 of the Internet Appendix show that the coefficient of our measure drops from 0.298 to 0.143 when controlling only for IVOL and from 0.298 to 0.202 when controlling only for ROA. When we orthogonalize cybersecurity risk with respect to IVOL and also with respect to ROA we find that cybersecurity risk remains strongly positively related to subsequent returns (see Panel C-Table IA.12 of the Internet Appendix).

We then adopt a broader perspective and examine the extent to which other variables omitted from the regression could further reduce the coefficient estimate of our cybersecurity risk measure. Specifically, we test for omitted variable bias using the approach proposed by Oster (2019). We evaluate the change in coefficients and R-squared values after including control variables in the model and making a plausible assumption about the importance of included variables compared to the omitted ones. The results support the view that omitted variables are unlikely to subsume the effect of cybersecurity risk on returns. More specifically, we find that even if the impact of unobservables on returns were twice as large as the influence of the observables (that is, $\delta=2$), the coefficient on cybersecurity risk would remain positive and significant. In fact, the hypothetical δ estimate that eliminates the observed effect of our measure on returns is 2.56, which is substantially larger than 1 (cut-off suggested in Oster (2019) as a critical value).

6.3 Additional Robustness Tests

First, 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. We identify these firms by searching for the word "insurance" in their cybersecurity-risk disclosures. By construction, these firms are more likely to be classified as high cybersecurity-risk firms (P3) than low cybersecurity-risk firms (P1) (note that P1 includes only 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 cyberattacks, which suggests investors should be less concerned about their exposure to cybersecurity risk. Panel I of Table IA.7 of the Internet Appendix shows, consistent with this reasoning, 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).

Second, 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). The results presented in Panel J of Table IA.7 of the Internet Appendix show a decline in the spread of the long-short portfolio, especially for the case of value-weighted returns. Nevertheless, the documented premium 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 firms that experienced major cyberattacks in the past do not drive our findings.

Third, we examine the extent to which our sample period and/or the way we rebalance our portfolios drive our results. Data requirements make 2007 the earliest year for which we can estimate cybersecurity risk, and given our sample period ends in 2018, our panel data is relatively short in its time dimension. We therefore conduct an additional robustness test. We assume that a firm's exposure to cybersecurity risk is persistent over time, and we forward-fill our measure up to 2020 (i.e., we replace all missing values for years 2019 and 2020 with the last available non-missing observation from 2017 or 2018). While this exercise does not fully resolve the data issue discussed above, it enables us to extend our sample by several months

and perform the analysis from March 2008 to December 2020. The results, in Panel K of Table IA.7 of the Internet Appendix, continue to support the existence of a positive and statistically significant premium of high cybersecurity-risk stocks. In Panel L of Table IA.7 of the Internet Appendix, 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.

Fourth, to supplement the daily analysis of the cybersecurity-risk factor (Section 4.4) and the evidence from SolarWinds hack (Section 5), which both support a *risk-based explanation* for our findings, we extend our cross-sectional tests and provide further evidence that cybersecurity risk has a significant systematic component. To do so, we focus on exposure to the cybersecurity-risk factor rather than the cybersecurity risk index itself. Specifically, using rolling firm-level regressions of monthly returns over the previous 60 months, we first estimate *betas* on our cybersecurity-risk factor (*Cyber Beta*). We then examine the cross-sectional relation between Cyber Beta and stock returns by running stock-level Fama-MacBeth (1973) regressions. As shown in the results presented in Table IA.13 of the Internet Appendix, a strong positive relation exists between Cyber Beta and stock returns (the coefficient on Cyber Beta is positive and statistically significant at the 1% level). These results confirm the conclusions drawn from the cross-sectional tests in Section 4.3.

Finally, we examine the possibility that other, simpler measures of cybersecurity risk can also be used to proxy for cybersecurity risk. A natural alternative is the number of sentences of cybersecurity-related disclosures in each 10-K document (*Cyber-related Disclosures*). We compare our measure with the variable "Cyber-related Disclosures" and find that while the latter has some ability to predict cyberattacks, our cybersecurity risk measure has a substantially stronger predictive power (see results in Table IA.14 of the Internet Appendix). These results are not surprising, because the number of sentences of cyber-related disclosures overlooks how firms manage cybersecurity risk and hence it is a noisy proxy for cybersecurity

risk. Another alternative is to estimate the probability of a cyberattack using accounting variables. We use a logistic regression and consider all firm-level variables used in Table 6. More specifically, we use fiscal year *t-1* variables to predict year *t* cyberattacks and use the coefficients of this regression to construct the cyberattack probability for *t+1*. This procedure enables us to construct an alternative *ex-ante* measure of cybersecurity risk (*Cyberattack Probability*). We then assess the ability of this measure to predict future cyberattacks and find the coefficient on *Cyberattack Probability* is statistically insignificant (see Table IA.14 of the Internet Appendix). Overall, we show that our measure exhibits superior ability in predicting future cyberattacks compared with these measures and hence it is more successful in capturing cybersecurity risk.

7. Conclusion

We construct a novel firm-level measure of cybersecurity risk using textual analysis of cybersecurity-risk disclosures in "Item 1A. Risk Factors" section of 10-K statements and use it to examine whether cybersecurity risk is priced in the cross-section of stock returns. The measure successfully identifies firms that discuss risk extensively and it displays intuitive relations with quantitative measures based on cybersecurity-risk disclosure language. In addition, the measure displays plausible 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 to perform their operations. We also find the measure correlates with several characteristics linked to firms hit by cyberattacks, such as size, age, growth opportunities, asset tangibility, R&D expenditures, and the presence of trade secrets. Finally, we find the measure also predicts the probability of experiencing a future cyberattack. 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 going long on firms with high cybersecurity-risk and short on low-cybersecurity-risk stocks earns a statistically significant 66-69 basis points per month - up to 8.3% - in equal-weighted returns over the following year. Fama-MacBeth cross-sectional regressions confirm a positive and statistically significant association between future 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. Overall, these results support the idea that investors require compensation for bearing cybersecurity risk.

Although we document that the return premium high-exposure firms earn is a robust feature of the data, the short sample period limits our analysis. For instance, sorting on a firm-specific measure might artificially result in high long-short portfolio returns purely due to luck, or due to certain industries performing well over a specific time period (Harvey, Liu and Zhu 2016; Freyberger, Neuhierl, and Weber 2020). For several reasons we believe these concerns are immaterial in our setting: First, our portfolio analysis is not based on any arbitrary measure, but a measure that is related to an actual risk firms are exposed to, and investors and firms frequently discuss it. Moreover, we find higher returns in the second half of our sample period, during which the actual exposure to cybersecurity risk had increased. Second, we directly show our results are not driven by the exceptionally good performance of certain industries such as Tech and R&D-intensive industries. Third, controlling for other risk factors and studying industry-adjusted returns do not materially affect our results. Finally, we use an out-of-sample test that exploits the recent large-scale hack of SolarWinds and show that firms with higher *ex-ante* cybersecurity-risk scores on our measure exhibit negative CARs around the hack.

Yet, the losses even around the hack of SolarWinds might not be large enough to justify the documented risk premium. One potential explanation for the magnitude of the premium is that investors require compensation to hedge against consequences arising from a truly disastrous hack, which so far has not materialized. In his speech at the Office for the Director of National Intelligence, President Biden alludes to such a possibility when warning that cyberattacks could even lead to a "real shooting war". Another potential explanation is that the documented premium also captures risks, other than cybersecurity. Our results and extensive robustness tests, however, make this alternative less likely.

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

References

Akey, P., S. Lewellen, I. Liskovich, and C. Schiller. 2020. Hacking corporate reputations. Working Paper, University of Toronto.

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

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

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

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

Arora, A., S. Belenzon, and L. Sheer. 2021a. Knowledge spillovers and corporate investment in scientific research. *American Economic Review* 111:871-98.

———. 2021b. Matching patents to Compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures. *Research Policy* 50: 104217.

Ashraf, M. Forthcoming. The role of peer events in corporate governance: Evidence from data breaches. *Accounting Review*.

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

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

Ben-Rephael, A., Z. Da, and R.D. Israelsen. 2017. It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies* 30:3009-3047.

Berkman, H., J. Jona, G. Lee, and N. Soderstrom. 2018. Cybersecurity awareness and market valuations. *Journal of Accounting and Public Policy* 37:508-526.

Boasiako, K.A., and M.O.C. Keefe. 2021. Data breaches and corporate liquidity management. *European Financial Management* 27:528-551.

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

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

Campbell, J.Y., and L. Hentschel. 1992. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31:281-318.

Campbell, J.L., H. Chen, D.S. Dhaliwal, H-S. Lu, and L.B. Steele. 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.

Cazier, R. A., J.L. McMullin, and J.S. Treu. 2021. Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language. *Accounting Review* 96:131-155.

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

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

Crosignani, M., M. Macchiavelli, and A.F. Silva. 2021. Pirates without borders: The propagation of cyberattacks through firms' supply chains. FRB of New York Staff Report.

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

deHaan, E., L. Alastair, and R. Litjens. 2021. Measurement error in Google ticker search. Working Paper. University of Washington.

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

Dzieliński, M., A.F. Wagner, and R.J. Zeckhauser. 2017. Straight talkers and vague talkers: The effects of managerial style in earnings conference calls. Working Paper. Stockholm University.

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

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

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

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

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33:2326-2377.

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

Gentzkow, M., B. Kelly, and M. Taddy. 2019. Text as data. *Journal of Economic Literature* 57:535-74.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *Review of Financial Studies* 33:2223-2273.

Hanley, K.W., and G. Hoberg. 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23:2821-2864.

———. 2019. Dynamic interpretation of emerging risks in the financial sector. *Review of Financial Studies* 32:4543-4603.

Harvey, C.R., Y. Liu, and H. Zhu. 2016. … and the cross-section of expected returns. *Review of Financial Studies* 29:5-68.

Harvey, C.R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263-1295.

Hassan, T.A., S. Hollander, L. van Lent, and A. Tahoun. 2019. Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics* 134:2135-2202.

Hilary, G., B. Segal, and M. Zhang. 2016. Cyber-risk disclosure: Who cares? Georgetown University.

Hirshleifer, D., P.H. Hsu, and D. Li. 2013. Innovative efficiency and stock returns. *Journal of Financial Economics* 107:632-654.

———. 2018. Innovative originality, profitability, and stock returns. *Review of Financial Studies* 31:2553-2605.

Hoberg, G., and V. Maksimovic. 2015. Redefining financial constraints: A text-based analysis. *Review of Financial Studies* 28:1312-1352.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23:3773-3811.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423-1465.

Israelsen, R.D. 2014. Tell it like it is: Disclosed risks and factor portfolios. Working Paper. Michigan State University.

Jamilov, R., H. Rey, and A. Tahoun. 2021. The anatomy of cyber risk. Working Paper. University of Oxford.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *Journal of Finance* 45:881-898.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65-91.

Jiang, H., N. Khanna, and Q. Yang. 2020. The cyber risk premium. Working Paper. Michigan State University.

Johnson, M.S., M.J. Kang, and T. Lawson. 2017. Stock price reaction to data breaches. *Journal of Finance Issues* 16:1-13.

Kamiya, S., J.K. Kang, J. Kim, A. Milidonis, and R. Stulz. 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics* 139:719-749.

Lang, M., and L. Stice-Lawrence. 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60:110-135.

Lattanzio, G., and Y. Ma. 2021. Corporate innovation in the cyber age. Nazarbayev University.

Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review* 86:1087-1115.

Lending, C., C. Minnick, and P.J. Schorno. 2018. Corporate governance, social responsibility and data breaches. *Financial Review* 53:413-455.

Lettau, M., M. Maggiori, and M. Weber. 2014. Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics* 114:197-225.

Li, D. 2011. Financial constraints, R&D investment, and stock returns. *Review of Financial Studies* 24:2974-3007.

Lopez-Lira, A. 2020. Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. Working Paper. University of Florida.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66:35-65.

———. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54:1187-1230.

Lowry, M., R. Michaely, and E. Volkova. 2020. Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO. *Review of Financial Studies* 33:5510-5554.

McMullin, J.L. 2016. Can I borrow your footnotes? Footnote boilerplate's learning externality. Working Paper. Indiana University.

Neuhierl, A. and M. Weber. 2020. Monetary policy communication, policy slope, and the stock market. *Journal of Monetary Economics* 108:140-155.

Oster, E. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37:187-204.

Romanosky, S. 2016. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity* 2:121-135.

Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang. 2020. Firm-level climate change exposure. Frankfurt School of Finance & Management.

Securities and Exchange Commission. 2011. CF disclosure guidance: Topic no. 2: Cybersecurity.

———. 2018. Commission statement and guidance on public company cybersecurity disclosures.

Tosun, O.K. 2021. Cyberattacks and stock market activity. *International Review of Financial Analysis* 76:101795.

Weber, M. 2018. Cash flow duration and the term structure of equity returns. *Journal of Financial Economics* 188:486-503.

You, H., and X.J. Zhang. 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 selected 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 cyberattacks; 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.[25] 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".

[25] 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.[26] 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).

[26] 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, cyberattack, cyberattacks, cyber risk, cyber security, cybersecurity, cyber intrusions, unauthorized access, 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 governmental authorities regarding the Company's compliance with these obligations.	Yes	Indirect: Description of Company Business

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 cyberattack 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 cyberattack, 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 cyberattacks.	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	Cyberattacks impacting our networks or systems could have an adverse effect on our business.	Yes	Direct: Description of Cybersecurity Risk
2	Cyberattacks, 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	Cyberattacks 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	Cyberattacks 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 cyberattacks 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	Cyberattacks 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 cyberattacks 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 cyberattacks,	Yes	Direct: Description of Cybersecurity Risk

	including protection of our systems and networks, may be insufficient to repel or mitigate the effects of a major cyberattack in the future.		
9	The inability to operate our networks and systems or those of our suppliers, vendors and other service providers as a result of cyberattacks, 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 cyberattack 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 cyberattack.	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 cyberattacks 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
AIA	Abnormal institutional investor attention, as measured from Bloomberg searches using the methodology of Ben-Rephael, Da and Israelsen (2017), in any of the five trading days following SolarWinds's disclosure of the data breach on Dec 14th, 2020. Given that we are interested in abnormal attention, and not just the level of attention, we define AIA as a dummy variable that takes a value of one if Bloomberg's daily maximum is 3 or 4 in any of the five trading days of interest, and zero otherwise.	Bloomberg
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
CAR[-1,1]	Cumulative abnormal return from day -1 to day +1 around SolarWinds's disclosure of the data breach on Dec 14th 2020 (event date), as calculated using the market model.	CRSP
CAR[-1,3]	Cumulative abnormal return from day -1 to day +3 around SolarWinds's disclosure of the data breach on Dec 14th, 2020 (event date), as calculated using the market model.	CRSP
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 cyberattacks, and 0 otherwise.	10-K
Cyber-related Disclosures	The length (number of sentences) of cyber-related disclosures in Item 1A. Risk Factors section.	10-K
Cyberattack Probability	An ex-ante measure of cybersecurity risk calculated after using fiscal year t-1 variables to predict year *t* cyberattacks and then using the coefficients of this regression to construct the cyberattack probability for *t+1*.	Compustat
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 cyberattack 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 cyberattack 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 ratio of "litigious" words to total 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 ratio of "negative" words to total 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)
Patent Flow	Patent flow is the number of patents a firm produces in a given year.	Duke Innovation & Scientific Enterprises Research Network
Patent Stock	The summing up the number of patents a firm owns prior to, and up to, a given year.	Duke Innovation & Scientific Enterprises Research Network

Precise Words	The ratio of "precise" words to total 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 cyberattacks, 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 cyberattacks 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 cyberattacks), 2007 is the earliest year for which we get an estimate of cybersecurity risk.

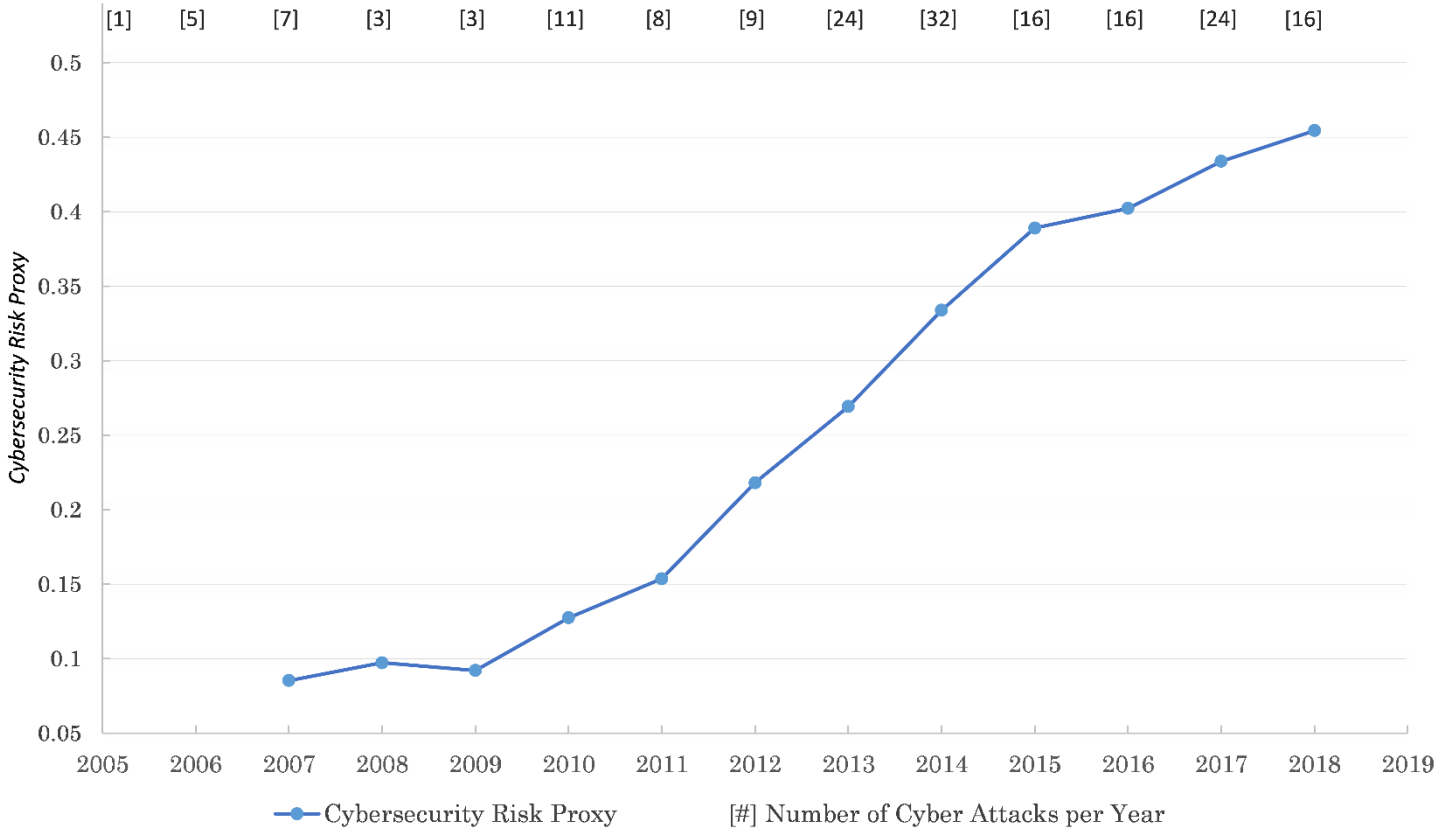

Figure 2
Cybersecurity Risk across Industries

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

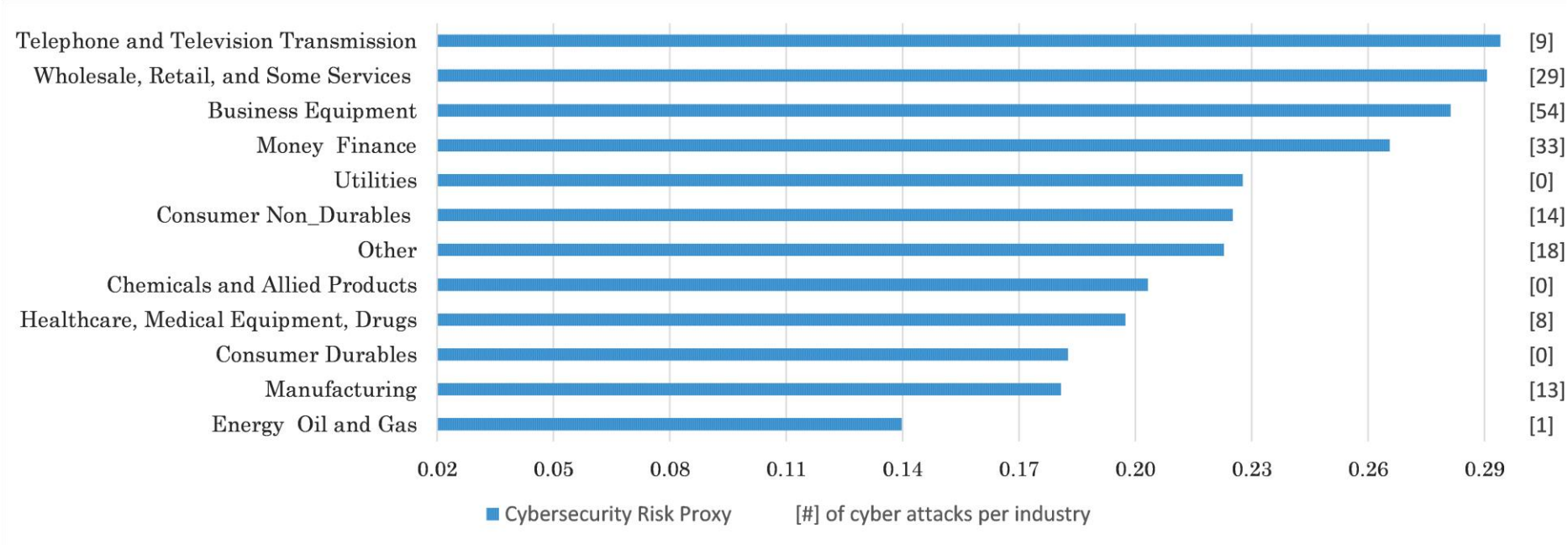

Figure 3
SolarWinds Hack: Key Facts and Timeline

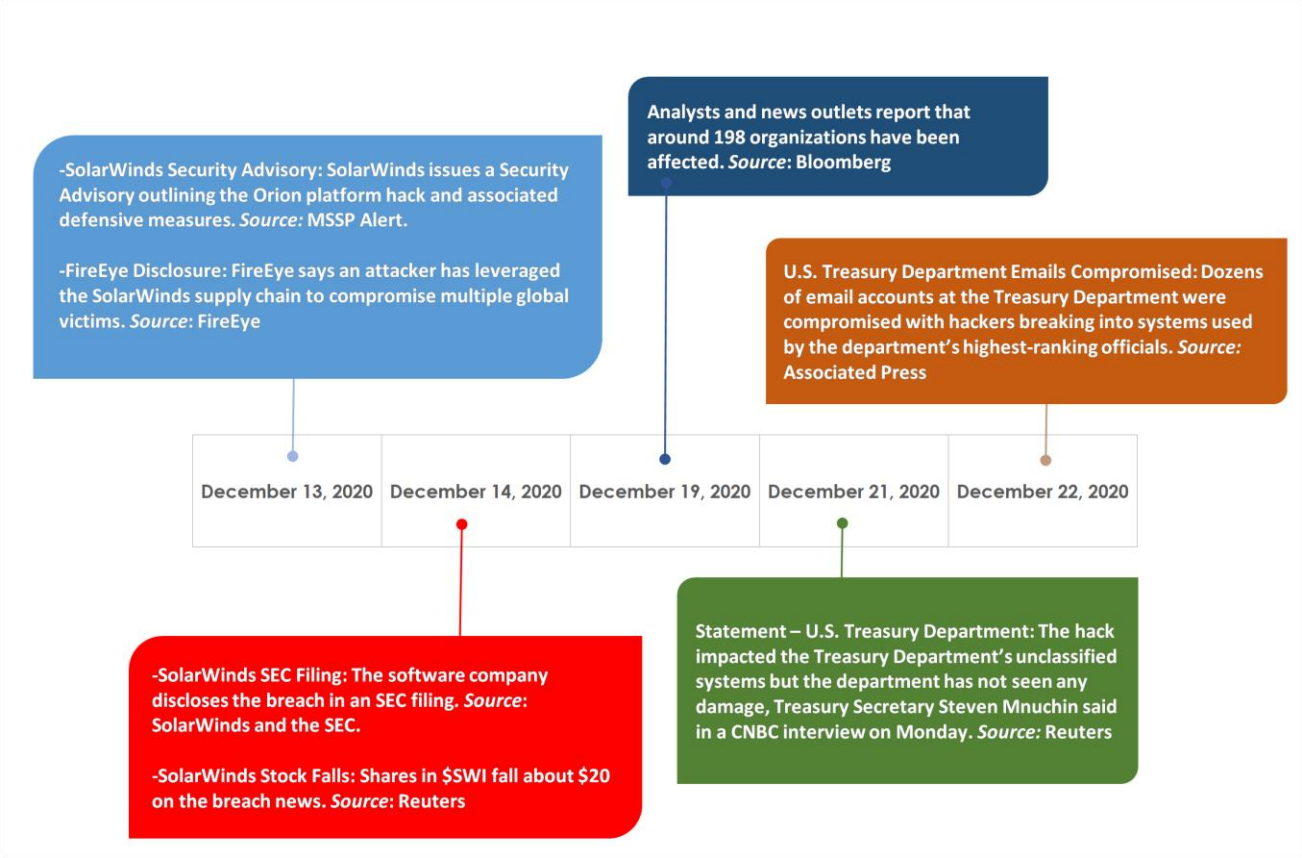

Table 1
Excerpts from Cybersecurity-risk Disclosures

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

Company Name	Fiscal Year	Cybersecurity Score	Text from Cybersecurity Risk Disclosures
Walgreens Boots Alliance Inc	2018	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	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	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	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	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	Cybersecurity Score	Text from Cybersecurity Risk Disclosures
Weyerhaeuser Co	2015	0.036	We and our service providers employ what we believe are adequate security measures.
Hess Corp	2012	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 cyberattacks.
Wayside Technology Group Inc	2013	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	0.109	Disruptions could be caused by a variety of factors, such as catastrophic events or weather, power outages, or cyberattacks on our systems by outside parties.
Dover Corp	2012	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 (ratio) is the ratio of "negative" words to total words in cybersecurity-risk disclosures. Precise Words (ratio) is the ratio of "precise" words to total words in cybersecurity-risk disclosures. Litigious words (ratio) is the ratio of "litigious" words to total 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 cyberattacks, 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 (ratio)*	0.033 ***	-0.215 ***	-0.133 ***	1.000			
(v)	*Precise Words (ratio)*	0.084 ***	0.071 ***	0.016 ***	-0.145 ***	1.000		
(vi)	*Litigious Words (ratio)*	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 by cosine similarity (see Section 2.4 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 Cyberattacks

This table reports the results of logit regressions of cybersecurity risk (cosine similarity) on future cyberattacks. Panel A includes all cyberattacks reported in PRC database for which we have complete risk disclosure and financial data. In Panel B we restrict our attention to major cyberattacks 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 cyberattacks (those that did not attract attention from major Newswires). Future cyberattacks 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 by 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	**0.167**	**0.710** *	**0.845** **	**0.678** ***
		[0.37]	[1.70]	[2.17]	[4.56]
	vw	**0.508**	**0.830** **	**1.117** ***	**0.609** ***
		[1.25]	[2.39]	[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		1235	955	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 10%, 5% and 1% 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					
Illiquidity	LOW	**0.271 ***	**0.365 *****	**0.167**	**0.268 *****
		[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]
Idiosyncratic Volatility	LOW	**0.103**	**0.087**	**0.583 ****	**0.551 *****
		[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]
Risk Section Length (ln)	LOW	**0.348 ****	**0.348 *****	**0.540 ****	**0.559 *****
		[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]
Readability (ln)	LOW	**0.861 *****	**0.786 *****	**0.826 *****	**0.750 *****
		[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, R&D Expenditures, 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 10%, 5% and 1% 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 *** **	**0.102 ** **	**0.124 *** **	**0.117 *** **	**0.111 *** **	**0.150 *** **	**0.122 *** **	**0.115 *** **
	[6.28]	[2.64]	[2.80]	[2.86]	[2.82]	[3.11]	[2.98]	[2.69]
Beta	-	**0.088**	**0.088**	**0.086**	**0.032**	**0.026**	**0.021**	**0.014**
		[0.81]	[0.84]	[0.76]	[0.29]	[0.26]	[0.22]	[0.15]
Market Value	-	**-0.081**	**-0.056**	**-0.067**	**-0.011**	**-0.018**	**0.042**	**-0.011**
		[-1.22]	[-0.77]	[-0.95]	[-0.15]	[-0.27]	[0.58]	[-0.14]
Book-to-Market	-	**0.067**	**0.067**	**0.063**	**0.056**	**0.070**	**0.078**	**0.078 ***
		[1.25]	[1.31]	[1.24]	[1.09]	[1.44]	[1.61]	[1.79]
Momentum	-	**0.153 ***	**0.148 ***	**0.103**	**0.113**	**0.079**	**0.164 *** **	**0.147 *** **
		[1.76]	[1.70]	[1.28]	[1.51]	[1.21]	[4.02]	[3.21]
Reversal	-	**-0.117 ** **	**-0.123 ** **	**0.207 *** **	**0.139 ** **	**0.139 *** **	**0.115 ***	**0.067**
		[-2.08]	[-2.22]	[2.87]	[2.47]	[2.97]	[1.89]	[1.03]
Illiquidity	-	**-0.013**	**-0.015**	**0.023**	**0.029**	**0.052 ***	**0.027**	**-0.005**
		[-0.35]	[-0.41]	[0.77]	[0.94]	[1.65]	[0.86]	[-0.14]
CoSkew	-	**-0.029**	**-0.026**	**-0.008**	**-0.022**	**-0.005**	**-0.026**	**0.034**
		[-0.95]	[-0.85]	[-0.29]	[-0.63]	[-0.15]	[-0.78]	[1.03]
Indiosyncratic Volatility	-	**-0.474 *** **	**-0.467 *** **	**-0.385 *** **	**-0.487 *** **	**-0.495 *** **	**-0.405 *** **	**-0.458 *** **
		[-5.76]	[-5.79]	[-4.19]	[-5.53]	[-6.01]	[-4.76]	[-5.67]
Asset Growth	-	**-0.108 *** **	**-0.099 *** **	**-0.070 ***	**-0.042**	**-0.066**	**0.013**	**0.008**
		[-2.75]	[-2.66]	[-1.95]	[-1.20]	[-1.46]	[0.27]	[0.21]
ROA	-	**0.518 *** **	**0.504 *** **	**0.477 *** **	**0.444 *** **	**0.493 *** **	**0.550 *** **	**0.593 *** **
		[7.59]	[8.10]	[7.90]	[6.83]	[7.79]	[11.21]	[11.04]
R&D Expenditures	-	**0.322 *** **	**0.318 *** **	**0.288 *** **	**0.281 *** **	**0.283 *** **	**0.335 *** **	**0.276 *** **
		[4.45]	[5.09]	[4.81]	[4.54]	[4.59]	[5.86]	[4.90]
Max	-	**-0.422 *** **	**-0.415 *** **	**-0.396 *** **	**-0.210 *** **	**-0.187 ** **	**-0.116**	**-0.124**
		[-4.57]	[-4.58]	[-4.06]	[-2.69]	[-2.38]	[-1.46]	[-1.57]
Risk Section Length (ln)	-	-	**-0.055**	**-0.065 ***	**-0.059 ***	**-0.060 ***	**-0.021**	**-0.027**
			[-1.42]	[-1.84]	[-1.76]	[-1.72]	[-0.63]	[-0.83]
Readability (ln)	-	-	**-0.032**	**-0.024**	**-0.014**	**0.003**	**0.004**	**0.021**
			[-0.63]	[-0.48]	[-0.28]	[0.07]	[0.08]	[0.43]
Constant	**0.515**	**0.512**	**0.509**	**0.475**	**0.496**	**0.509**	**0.799 ** **	**0.915 ** **
	[1.14]	[1.12]	[1.12]	[1.03]	[1.08]	[1.13]	[2.24]	[2.43]
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 (see Section 4.4 for details); "High_Google_SVI_dummy" is a dummy variable that takes the value of 1 on days with high Google SVI index (greater than the mean SVI index plus 1.5 standard deviations, both estimated during the past 2 weeks) of the search topics "Data Breach" and "Hacker", and 0 otherwise. We estimate daily abnormal SVI by scaling each daily SVI with the median SVI estimated during the past 2 weeks.; 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 controls only for the market risk factor (CAPM specification), Model 3 controls for market, value and momentum factors (FFC specification), while Model 4 control for all five risk factors proposed by Fama and French (2015) (FF-5 specification). Panel A presents the main results, Panels B-D present robustness results, using (i) a cybersecurity risk factor based on 10 portfolios rather than 5 portfolios, (ii) by scaling SVI index with the median SVI estimated during the past 4 weeks, and (iii) by focusing on more extreme high Google SVI index (greater than the mean SVI index plus 2 standard deviations, both estimated during the past 2 weeks), respectively. Panel E (F) replaces 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. ** and *** denote statistical significance at the 5% and 1% levels, respectively.

	Cybersecurity Risk Factor $_t$			
	CONTROLS			
	NONE	*CAPM*	*FFC*	*FF-5*
	[1]	[2]	[3]	[4]
Panel A: Benchmark				
Constant	**0.0002 *** **	**0.0002 *** **	**0.0002 *** **	**0.0002 *** **
	[3.39]	[3.55]	[3.48]	[3.63]
High Google SVI Dummy	**-0.0004 ** **	**-0.0004 ** **	**-0.0004 ** **	**-0.0004 *** **
	[-2.43]	[-2.41]	[-2.46]	[-2.64]
Panel B: Robustness (Alternative Factor)				
Constant	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **	**0.0002 *** **
	[2.50]	[2.46]	[2.48]	[2.70]
High Google SVI Dummy	**-0.0005 *** **	**-0.0005 *** **	**-0.0005 *** **	**-0.0005 *** **
	[-2.69]	[-2.70]	[-2.64]	[-2.76]
Panel C: Robustness (Alternative Shocks 1)				
Constant	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **	**0.0002 *** **
	[2.33]	[2.30]	[2.34]	[2.58]
High Google SVI Dummy	**-0.0005 ** **	**-0.0005 ** **	**-0.0005 ** **	**-0.0006 *** **
	[-2.41]	[-2.43]	[-2.42]	[-2.58]
Panel D: Robustness (Alternative Shocks 2)				
Constant	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **
	[2.40]	[2.35]	[2.38]	[2.57]
High Google SVI Dummy	**-0.0006 *** **	**-0.0006 *** **	**-0.0006 *** **	**-0.0007 *** **
	[-2.78]	[-2.80]	[-2.76]	[-2.86]
Panel E: Placebo Tests (1 week after the peak of SVI)				
Constant	**0.0001 ** **	**0.0001 ** **	**0.0001 ** **	**0.0001 ** **
	[2.00]	[2.15]	[2.06]	[2.04]
Placebo High Google SVI Dummy + 1 week	**0.0001**	**0.0001**	**0.0001**	**0.0001**
	[0.25]	[0.29]	[0.35]	[0.44]
Panel F: Placebo Tests (1 month after the peak of SVI)				
Constant	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **	**0.0002 ** **
	[2.38]	[2.55]	[2.46]	[2.44]
Placebo High Google SVI Dummy + 1 month	**-0.0001**	**-0.0001**	**-0.0001**	**-0.0001**
	[-0.63]	[-0.66]	[-0.59]	[-0.56]

Table 11
Cybersecurity Risk and Stock Returns: Evidence from the SolarWinds Hack

This table reports results from an event study analysis. For all firms in the sample, we use the market model to estimate cumulative abnormal returns CAR[-1, +1] and CAR[-1, +3] around December 14th, 2020, which is the date when SolarWinds disclosed a cyberattack in an SEC filing. Panel A reports the average CARs for the top and bottom decile portfolios formed based on Cybersecurity Risk Index (measured in 2018 or 2017, if unavailable in 2018). Panel B reports regression results. The dependent variables are cumulative abnormal returns CAR[-1, +1] and CAR[-1, +3]. The main independent variable is the Cybersecurity Risk Index. We alternatively use (i) a dummy variable that equals 1 for firms with Cybersecurity Risk Index in the top tercile of the distribution (High Cyber Risk Dummy 1), and 0 otherwise, and (ii) a dummy variable that equals 1 for firms with Cybersecurity Risk Index in the top decile of the distribution (High Cyber Risk Dummy 2), and 0 otherwise. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Panel A: Average CARs per portfolio	CAR[-1,+1]			CAR[-1,+3]		
Low Cybersecurity Risk Portfolio [P1]	0.006			0.004		
High Cybersecurity Risk Portfolio [P10]	-0.009			-0.008		
High-Low [P10-P1]	-0.015			-0.012		
t-test [p-value]	[0.00] ***			[0.01] **		
Panel B: Regression Analysis	CAR[-1,+1]			CAR[-1,+3]		
Cybersecurity Risk Index	**-0.011** **	-	-	**-0.011** *	-	-
	[-2.11]	-	-	[-1.66]	-	-
High Cyber Risk Dummy 1	-	**-0.007** ***	-	-	**-0.004**	-
	-	[-3.01]	-	-	[-1.48]	-
High Cyber Risk Dummy 2	-	-	**-0.012** ***	-	-	**-0.010** ***
	-	-	[-3.99]	-	-	[-2.57]
Number of Observations	3,289	3,289	3,289	3,289	3,289	3,289

Table 12
Solarwinds Customers (Affected Firms) vs. Non-costumers (Non-Affected Firms)

This table reports results from an event study analysis. The event date is the December 14th, 2020, which is the date when SolarWinds disclosed a cyberattack in an SEC filing. Panel A reports averages of characteristics for affected and non-affected firms. Affected firms include SolarWinds's key customers whereas non-affected firms include all other firms in our sample. The characteristics include abnormal institutional investor attention (AIA), as measured from Bloomberg searches using the methodology of Ben-Rephael, Da and Israelsen (2017) in any of the five trading days following SolarWinds's disclosure of the breach, CAR[-1, +1], CAR[-1, +3] and the Cybersecurity Risk Index. Panel B reports logit regression results. The dependent variable is a dummy variable that equals 1 when the firm is among SolarWinds's key customers, and zero otherwise. The main independent variable is the Cybersecurity Risk Index. We alternatively use (i) a dummy variable that equals 1 for firms with Cybersecurity Risk Index in the top tercile of the distribution (High Cyber Risk Dummy 1), and 0 otherwise, and (ii) a dummy variable that equals 1 for firms with Cybersecurity Risk Index in the top decile of the distribution (High Cyber Risk Dummy 2), and 0 otherwise. The continuous variable in Panel B is standardized to have a mean of 0 and standard deviation of 1. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A: Differences in Means	Affected Firms	Non-Affected Firms	*t-test* [p-value]
% of Firms with AIA	64.00	37.13	[0.01] ***
CAR[-1,+1]	-0.012	0.002	[0.01] ***
CAR[-1,+3]	-0.017	0.001	[0.01] ***
Cybersecurity Risk Index	0.491	0.442	[0.03] **
Panel B: Logistic Regression	Prob (1=Affected Firm / 0=Non-Affected Firm)		
Cybersecurity Risk Index	**0.860** ***	-	-
	[2.77]	-	-
High Cyber Risk Dummy 1	-	**0.815** **	-
	-	[2.40]	-
High Cyber Risk Dummy 2	-	-	**0.490**
	-	-	[1.10]
Observations	3,289	3,289	3,289

Internet Appendix

for

Cybersecurity Risk

Table IA.1
Percentage of Firms with Cyber-related Disclosures by Industry and Year

This table presents the percentage of firms with cyber-related disclosures by (i) year and (ii) Fama and French 12 industry, where: 1- Consumer Non Durables; 2 -Consumer Durables; 3-Manufacturing; 4-Energy Oil and Gas; 5-Chemicals and Allied Products; 6-Business Equipment; 7-Telephone and Television Transmission; 8-Utilities; 9-Wholesale, Retail, and Some Services; 10-Healthcare, Medical Equipment, Drugs; 11-Money Finance and 12-Other.

	Full	*Fama-French Industry Group =*											
Year	Sample	1	2	3	4	5	6	7	8	9	10	11	12
2007	0.29	0.27	0.13	0.12	0.07	0.15	0.41	0.46	0.09	0.36	0.23	0.36	0.25
2008	0.31	0.26	0.11	0.13	0.07	0.22	0.42	0.54	0.10	0.39	0.23	0.41	0.27
2009	0.35	0.27	0.17	0.11	0.07	0.28	0.46	0.55	0.19	0.46	0.26	0.48	0.33
2010	0.39	0.36	0.17	0.15	0.08	0.32	0.49	0.57	0.26	0.53	0.28	0.53	0.36
2011	0.55	0.50	0.35	0.34	0.28	0.54	0.64	0.77	0.70	0.68	0.43	0.65	0.51
2012	0.66	0.60	0.49	0.49	0.42	0.60	0.75	0.81	0.83	0.78	0.52	0.74	0.61
2013	0.73	0.71	0.68	0.57	0.57	0.72	0.79	0.89	0.91	0.86	0.61	0.81	0.68
2014	0.80	0.79	0.80	0.71	0.65	0.75	0.84	0.93	0.94	0.90	0.70	0.86	0.78
2015	0.85	0.83	0.85	0.81	0.72	0.77	0.90	0.96	0.97	0.91	0.78	0.90	0.81
2016	0.88	0.86	0.89	0.85	0.78	0.83	0.91	0.96	0.97	0.94	0.84	0.89	0.86
2017	0.90	0.90	0.93	0.88	0.85	0.88	0.93	0.97	0.95	0.94	0.88	0.91	0.88
2018	0.89	0.95	0.92	0.87	0.60	1.00	0.95	1.00	1.00	0.90	0.82	0.86	0.82

Table IA.2
Cybersecurity Risk and Firm Characteristics-Jaccard Similarity

This table reports the results of linear regressions of firm characteristics on cybersecurity risk as measured by 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.3

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.4
Cybersecurity Risk and Future Cyberattacks-Jaccard Similarity

This table reports the results of logit regressions of cybersecurity risk (Jaccard similarity) on future cyberattacks. Panel A includes all cyberattacks reported in PRC database for which we have complete risk disclosure and financial data. In Panel B we restrict our attention to major cyberattacks 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 cyberattacks (those that did not attract attention from major Newswires). Future cyberattacks 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** ***
	[-10.12]	[-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.5
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 by 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 10%, 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.6
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, R&E expenditures, 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 10%, 5% and 1% levels, respectively.

	$Returns_{t+1}$		
	[1]	[2]	[2]
Cybersecurity Risk Index (Jaccard)	**0.297** ***	**0.111** ***	**0.136** ***
	[6.34]	[2.83]	[3.03]
Beta	-	**0.089**	**0.091**
		[0.82]	[0.86]
Market Value	-	**-0.086**	**-0.061**
		[-1.28]	[-0.83]
Book-to-Market	-	**0.068**	**0.068**
		[1.27]	[1.34]
Momentum	-	**0.153** *	**0.147** *
		[1.75]	[1.69]
Reversal	-	**-0.116** **	**-0.122** **
		[-2.07]	[-2.21]
Illiquidity	-	**-0.013**	**-0.015**
		[-0.35]	[-0.42]
CoSkew	-	**-0.029**	**-0.026**
		[-0.95]	[-0.85]
Indiosyncratic Volatility	-	**-0.476** ***	**-0.468** ***
		[-5.78]	[-5.80]
Asset Growth	-	**-0.109** ***	**-0.100** **
		[-2.76]	[-2.65]
ROA	-	**0.519** ***	**0.504** ***
		[7.61]	[8.10]
R&D Expenditures	-	**0.321** ***	**0.318** ***
		[4.44]	[5.08]
Max	-	**-0.423** ***	**-0.416** ***
		[-4.58]	[-4.57]
Risk Section Length (ln)	-	-	**-0.062**
			[-1.62]
Readability (ln)	-	-	**-0.033**
			[-0.64]
Constant	**0.515**	**0.512**	**0.509**
	[1.14]	[1.12]	[1.12]
Observations	409,016	342,573	342,573

Table IA.7
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, 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 C, we form our portfolios based on a revised cybersecurity risk measure, which replaces all zeros with the next non-zero observation for each firm. In Panel D, we exclude from the portfolio analysis all firms that belong to the same Fama-French 48 industry as firms in the training sample (i.e., peer firms). In Panel E, we form the portfolios based on a new cybersecurity risk measure, which is constructed after estimating the similarity of firm *i*'s cybersecurity-risk disclosure with past cybersecurity-risk disclosures of firms in a new training sample that *excludes* firms that have the same auditor as firm *i*. In Panel F, we repeat our portfolio analysis using industry-adjusted returns. To adjust returns by industry, we focus on the Fama-French 12 industry portfolios and their monthly average returns. For robustness purposes, we use both equal-weighted and value-weighted industry-level returns for the industry adjustment. In Panel G, 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 H, we report portfolio results after excluding firms from the Energy Oil and Gas and Consumer Durables industries. In Panel I (Panel J) we exclude from the analysis all firms with cyber insurance (firms in the training sample). In Panel K, we report results after forward-filling our measure and extending the sample to December 2020. Finally, in In Panel L we present results based on monthly and yearly rebalancing. 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 10%, 5% and 1% 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 Period				
January 2012 to March 2019	**0.916 ***** [5.88]	**0.870 ***** [5.56]	**0.770 ***** [3.00]	**0.652 ***** [3.45]
Panel B: 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 C: 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]
Panel D: Cybersecurity Risk after:				
Excluding peer firms (those that belong to the same FF48 industry with training firms)	**0.709 ***** [4.57]	**0.676 ***** [4.19]	**0.638 ***** [3.12]	**0.631 ***** [3.69]
Panel E: Cybersecurity Risk after:				
Excluding peer firms (same Auditor) from training sample	**0.757 ***** [5.54]	**0.748 ***** [5.40]	**0.553 ***** [2.64]	**0.540 ***** [3.16]

Table Continued Overleaf

Table IA.7 (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: Industry Adjustment based on				
Equal-Weighted Industry-Level Returns	**0.640** ***	**0.653** ***	**0.541** ***	**0.569** ***
	[5.82]	[5.53]	[2.60]	[3.14]
Value-Weighted Industry-Level Returns	**0.618** ***	**0.601** ***	**0.489** ***	**0.484** ***
	[4.65]	[4.50]	[2.57]	[2.83]
Panel G: 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]
Panel H: All Firms Excluding:				
Consumer Durables & Energy Oil and Gas	**0.601** ***	**0.574** ***	**0.551** ***	**0.499** ***
	[4.85]	[2.81]	[4.09]	[2.89]
Panel I: All Firms Excluding:				
Firms with Cyber Insurance	**0.660** ***	**0.634** ***	**0.676** ***	**0.649** ***
	[4.40]	[4.27]	[3.34]	[4.05]
Panel J: All Firms Excluding:				
Firms in Training Sample	**0.669** ***	**0.682** ***	**0.522** **	**0.507** **
	[4.42]	[4.67]	[2.20]	[2.42]
Panel K: Extended Sample				
March 2008 to December 2020	**0.737** ***	**0.679** ***	**0.747** ***	**0.623** ***
	[5.10]	[4.72]	[3.27]	[3.99]
Panel L: 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 IA.8
Cybersecurity Risk and Future Cyberattacks: Dealing with Peer Effects in Disclosure Language

This table reports the results of logit regressions of cybersecurity risk (cosine similarity) on future cyberattacks. The dependent variable is a dummy variable that equals 1 if the firm experiences a cyberattack at time *t+1*, and zero otherwise. The key explanatory variable in Model 1 is our cybersecurity risk measure (benchmark specification-see section 3.5). In Model 2, we replace cybersecurity risk measure with the variable Cybersecurity Risk Index (Excluding Peers-Industry), which is simply the cosine similarity of firms that do not belong to the same Fama-French 48 industry as firms in the training sample (i.e., peer firms). In Model 3, we use the variable Cybersecurity Risk Index (Excluding Peers-Auditor) as our key explanatory variable. This is constructed after estimating the cosine similarity of firm's *i* cybersecurity-risk disclosure with past cybersecurity-risk disclosures of firms in a new training sample, that excludes firms that have the same auditor with firm *i*. All independent variables are measured at time *t*. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. *** denotes statistical significance at the 1% level.

	Model 1	Model 2	Model 3
Cybersecurity Risk Index	**0.656 *****	-	-
	[4.60]	-	-
Cybersecurity Risk Index (Excluding Peers-Industry)		**0.660 *****	
		[4.21]	
Cybersecurity Risk Index (Excluding Peers-Auditor)	-	-	**0.631 *****
	-	-	[4.77]
Clustered SE	Firm	Firm	Firm
Industry fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes
Number of Observations	30,830	25,280	30,059
Pseudo-R-squared	0.223	0.244	0.215

Table IA.9
Double-Sorting: R&D and Innovation Activity

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) R&D Expenditures, defined as the ratio of R&D expenditures to total assets, (ii) Patent Flow, defined as the number of patents a firm produces in a given year and (iii) Patent Stock, defined as the number of patents a firm owns prior to, and up to, a given year. Data on patent stock and patent flow are drawn from the Duke Innovation & Scientific Enterprises Research Network (DISCERN*;* *https://doi.org/10.5281/zenodo.3594743*) database by Arora, Belenzon, and Sheer (2021a, 2021b). These data cover the period 1980 to 2015. For the purposes of our analysis, we forward-fill the missing data for years 2016-2018 with the 2015 values for the variables of interest. 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 10%, 5% and 1% 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*
R&D Expenditures	LOW	**0.452** ***	**0.438** ***	**0.437** ***	**0.497** ***
		[2.92]	[2.72]	[2.85]	[3.62]
	HIGH	**0.986** ***	**0.924** ***	**0.830** **	**0.681** **
		[5.92]	[5.56]	[1.98]	[2.05]
Patent Flow	LOW	**0.604** ***	**0.502** ***	**0.189**	**0.317** **
		[4.63]	[3.57]	[1.11]	[2.16]
	HIGH	**0.831** ***	**0.802** ***	**0.930** *	**0.762** **
		[4.33]	[4.64]	[1.93]	[2.08]
Patent Stock	LOW	**0.688** ***	**0.590** ***	**0.232**	**0.351** *
		[5.09]	[4.09]	[1.58]	[1.95]
	HIGH	**0.748** ***	**0.712** ***	**0.879** *	**0.725** **
		[4.22]	[5.26]	[1.94]	[2.10]

Table IA.10
Placebo Tests

This table reports results after using a placebo measure that is based on similarity in risk disclosures, other than cyber-related disclosures (see section 6.2 for details). Panel A reports the results of logit regressions. The dependent variable is a dummy variable that equals 1 if the firm experiences a cyberattack at time *t+1*, and zero otherwise. The key variable of interest is the placebo measure (Non-Cyber Disclosure Similarity). Model 1 (Model 2) excludes from (includes in) the model our cybersecurity risk measure. Both models include standard controls (as in Table 6 of the paper), which are measured at time *t*. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. Panel B reports alphas from the Fama and French's (2015) model for the spread strategy that is long the portfolio of stocks with the highest similarity in overall risk disclosures (Non-Cyber Disclosure Similarity) and short the portfolio of stocks with the lowest similarity in overfall risk disclosures (Non-Cyber Disclosure Similarity). Results are reported both for equally-weighted and value-weighted tercile, quartile, quintile and decile portfolios. Average alphas are bolded and their associated Newey-West t-statistics are reported in square brackets. * and *** denote statistical significance at 10% and 1% levels, respectively.

Panel A: Future Attacks	Model 1	Model 2
Cybersecurity Risk Index	-	**0.727 *** **
	-	[4.72]
Non-Cyber Disclosure Similarity	**-0.058**	**-0.268**
	[-0.31]	[-1.40]
Clustered SE	Firm	Firm
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Control Variables	Yes	Yes
Number of Observations	30,665	30,665
Pseudo-R-squared	0.201	0.225

Panel B: Portfolio Analysis	5-Factor Alpha of Spread Portfolios sorted on Non-Cyber Disclosure Similarity [High-Low]	
	Equal-weighted	*Value-weighted*
Tercile Portfolios (P3-P1)	**0.099**	**0.257**
	[0.92]	[1.63]
Quartile Portfolios (P4-P1)	**0.104**	**0.341 ***
	[0.84]	[1.73]
Quintile Portfolios (P5-P1)	**0.111**	**0.329 ***
	[0.86]	[1.71]
Decile Portfolios (P10-P1)	**-0.074**	**0.222**
	[-0.53]	[1.53]

Table IA.11
Fama-MacBeth Regressions: Controlling for Extra Risks

This table reports the results from Fama-MacBeth regressions on the relation between our Cybersecurity Risk Index and subsequent monthly 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 variables that capture other types of risk such as political, non-political and overall risk (Hassan et al., 2019) and climate risk (Sautner et al., 2020). These extra risk measures have been developed using textual analysis of earnings conference calls and the data are available online (see https://www.firmlevelrisk.com/ and https://osf.io/fd6jq/). 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. *** denotes statistical significance at the 1%.

	$Returns_{t+1}$			
	[1]	[2]	[3]	[4]
Cybersecurity Risk Index	**0.217** ***	**0.216** ***	**0.216** ***	**0.226** ***
	[4.08]	[4.10]	[4.10]	[4.37]
Political Risk (Hassan et al., 2019)	**-0.007**	-	-	-
	[-0.23]			
Non-political Risk (Hassan et al., 2019)	-	**0.017**	-	-
		[0.58]		
Overall Risk (Hassan et al., 2019)	-	-	**0.002**	-
			[0.05]	
Climate Risk (Sautner et al., 2020)	-	-	-	**-0.048**
				[-1.41]
Observations	291,625	291,625	291,625	308,375

Table IA.12
Fama-MacBeth Regressions: Further Results

This table presents further evidence from Fama-MacBeth regressions on the relation between our Cybersecurity Risk Index and subsequent monthly stock returns (1-month). Panel A presents the correlation coefficients and corresponding levels of statistical significance of the relationship between our cybersecurity risk measure and the variables idiosyncratic volatility (IVOL) and ROA. In Panel B, we present results from Fama-Macbeth regressions that do not include (Model 1) and do include IVOL/ROA as controls (Model 2/ Model 3). In Panel C, we present results from Fama-Macbeth regressions after orthogonalizing our cybersecurity risk with respect to IVOL and also with respect to ROA. To do so, we regress our measure on IVOL and obtain the residuals. The residual provides us with a new cybersecurity risk measure that is orthogonal to IVOL, namely Cybersecurity Risk (Orthogonal to IVOL). We repeat this exercise by regressing cybersecurity risk on ROA to obtain another cybersecurity risk measure that is orthogonal to ROA, namely Cybersecurity Risk (Orthogonal to ROA). All variables are defined in Appendix B. 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 5% and 1%, respectively.

Panel A: Correlations	(i)	(ii)	(iii)
(i) Cybersecurity Risk Index	**1.000**		
(ii) IVOL	**-0.214 *****	**1.000**	
(iii) ROA	**0.081 *****	**-0.419 *****	**1.000**
Panel B: The Effect of IVOL and ROA	[1]	[2]	[3]
Cybersecurity Risk Index	**0.298 *****	**0.143 *****	**0.202 *****
	[6.28]	[3.83]	[4.85]
IVOL	-	**-0.847 *****	-
		[-12.29]	
ROA	-	-	**0.606 *****
			[10.35]
Observations	409,016	406,850	407,700
Panel C: Orthogonalized Variables	[1]	[2]	[3]
Cybersecurity Risk Index	**0.124 *****	-	-
	[2.80]		
Cybersecurity Risk Index	-	**0.175 ****	-
(Orthogonal to IVOL)		[2.33]	
Cybersecurity Risk Index	-	-	**0.164 *****
(Orthogonal to ROA)			[3.36]
Controls	Yes	Yes	Yes
Observations	342,573	342,573	342,573

Table IA.13
Fama MacBeth Regressions: Exposure to Cybersecurity Risk Factor

This table reports the results from Fama-MacBeth regressions on the relation between Cyber Beta (calculated using rolling firm-level regressions of monthly returns on our cybersecurity risk factor over the previous 60 months) and subsequent monthly stock returns (1-month). For each month of our sample, we run cross-sectional regressions of excess stock returns on lagged Cyber Beta 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, R&D, 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 10% and 1% levels, respectively.

	$Returns_{t+1}$
Cyber Beta	**0.099 *****
	[3.42]
Beta	**-0.049**
	[-0.59]
Market Value	**-0.047**
	[-0.42]
Book-to-Market	**0.031**
	[0.66]
Momentum	**0.164 ***
	[1.74]
Reversal	**-0.045**
	[-0.58]
Illiquidity	**0.030**
	[0.67]
CoSkew	**-0.014**
	[-0.31]
Indiosyncratic Volatility	**-0.469 *****
	[-5.62]
Asset Growth	**-0.034**
	[-0.59]
ROA	**0.630 *****
	[8.89]
R&D	**0.386 *****
	[4.16]
Max	**-0.354 *****
	[-3.18]
Risk Section Length (ln)	**-0.032**
	[-0.95]
Readability (ln)	**0.064**
	[0.92]
Constant	**0.519**
	[1.37]
Observations	171,945

Table IA.14
Cybersecurity Risk and Future Cyberattacks: Comparison across Measures

This table reports the results of logit regressions of cybersecurity risk, as measured in three different ways, on future cyberattacks. The dependent variable is a dummy variable that equals 1 if the firm experiences a cyberattack at time *t+1*, and zero otherwise. To measure cybersecurity risk, we firstly use our own measure (Cybersecurity Risk Index) in Model 1 (benchmark specification). In Model 2, we alternatively use the length (number of sentences) of cyber-related disclosures (Cyber-related Disclosures). In Model 3, we use an *ex-ante* measure of cybersecurity risk calculated after using fiscal year *t-1* variables to predict year *t* cyberattacks and then using the coefficients of this regression to construct the cyberattack probability for *t+1* (*Cyberattack Probability*). Model 4 simultaneously controls for all three cybersecurity risk measures. All models include the controls used in Model 2 of Table 6, which are lagged. The continuous variables are standardized to have a mean of 0 and standard deviation of 1. Standard errors are clustered at the firm level. *** indicate statistical significance at the 1% level.

	Model 1	Model 2	Model 3	Model 4
Cybersecurity Risk Index	**0.656 *** **	-	-	**0.438 *** **
	[4.60]	-	-	[2.82]
Cyber-related Disclosures	-	**0.353 *** **	-	**0.248 *** **
	-	[5.88]	-	[3.17]
Cyberattack Probability	-	-	**-0.022**	**-0.010**
	-	-	[-0.91]	[-0.29]
Clustered SE	Firm	Firm	Firm	Firm
Industry fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Number of Observations	30,830	37,820	31,294	30,830
Pseudo-R-squared	0.223	0.223	0.213	0.229