Institute for New Economic Thinking

Working Papers
ineteconomics.org/workingpapers

The Anatomy of Cyber Risk

Rustam Jamilov,* Hélène Rey† and Ahmed Tahoun‡

Working Paper No. 206

May 10th, 2023

Abstract

This paper employs computational linguistics to introduce a novel text-based measure of firm-level cyber risk exposure based on quarterly earnings conference calls of listed firms. Our quarterly measures are available for more than 13,000 firms from 85 countries over 2002-2021. We document that cyber risk exposure predicts cyber attacks, affects stock returns and profits, and is priced in the equity option market.

* All Souls College, University of Oxford, rustam.jamilov@allsouls.ox.ac.uk.
† London Business School, CEPR, and NBER, hrey@london.edu. Corresponding Author.
‡ London Business School, atahoun@london.edu.

For useful comments and feedback we thank our discussants Jinfei Sheng, Ishita Sen, and Christodoulos Louca as well as Marie Briere, Matteo Crosignani, Thomas Ferguson, Chris Florackis, Jason Healey, Ralph Koijen, Anna Kovner, Marco Macchiavelli, Emanuel Moench, Patricia Mosser, Richard Portes, Elias Papaioannou, Zacharias Sautner, Andre Silva, Laurence van Lent, Grigory Vilkov and seminar participants at numerous venues. Aakash Kalyani, Markus Schwedeler, and Marco Bellifemine provided excellent research assistance. We thank Chris Florackis for sharing their factor data. Jamilov gratefully acknowledges financial support from the Wheeler Institute for Business and Development. Tahoun sincerely appreciates support from the Institute for New Economic Thinking (INET). The paper does not represent the views of the French Macroprudential Authority.

The cost of option protection against price, variance, and tail risks is greater for more cyber-exposed firms. Cyber risks spill over across firms and persist at the sectoral level. The geography of cyber risk exposure is well approximated by a gravity model extended with cross-border portfolio flows. Back-of-the-envelope calculations suggest that the global cost of cyber risk is over $200 billion per year.

https://doi.org/10.36687/inetwp206

JEL Codes: D22, L10

Keywords: Cyber risk, textual analysis, earnings calls

1 Introduction

The World Economic Forum identifies systemic cyber risk as one of the most likely and impactful risks for firms (WEF, 2016). The European Systemic Risk Board has characterized cyber security as a systemic risk to the European financial system (ESRB, 2020). Systemic risk surveys of financial market participants cite cyber security as the second most challenging risk for managing a firm, falling behind only political risk (BoE, 2020). Major institutions have lost more than $500 billion from operational risk events over the decade of 2011-2020, predominantly due to cyberattacks (ORX, 2020). According to the Center for Strategic and International Studies, cybercrime had caused economic losses of up to 1% of global economic output in 2014 (CSIS, 2014). During the COVID-19 pandemic the world saw an unprecedented rise in cybercrime, to the point that multiple unique cyberattacks were being reported each day (Lallie et al., 2021). An International Monetary Fund survey warns that cybersecurity is a real threat to financial stability and that the majority of national supervisory authorities do not have a clear cyber strategy or a dedicated cyber incident reporting protocol (Adrian and Ferreira, 2023). As the frequency of realized cyberattacks is growing and the uncertainty about potential future events intensifies, measurement and quantification of cyber risk and uncertainty are transforming into first-order issues for scholars and policy-makers alike.

This paper constructs novel, comprehensive text-based measures of firm-level exposure to cyber risk by leveraging quarterly earnings calls of listed firms and natural language processing techniques in the spirit of Hassan et al. (2019)[1]. Conference calls usually take place concurrently with an earnings release and grant a chance for management to describe the overall business position of their company (Hollander et al., 2010). Earnings calls are forward-looking since many interesting dialogues take place during post-announcement Q&A sessions when analysts ask questions about various pressing issues and future plans (Huang et al., 2018). Call participants are skilled experts, are arguably among the most knowledgeable of the firm's business model, and are thus likely to initiate any relevant conversation that may potentially affect future revenue and profits.

Using these earnings calls we measure cyber risk exposure faced by each firm in a given quarter through the means of counting the number of times cybersecurity-related terms get mentioned. Our universe of terms is comprised of cyber lexicon libraries from three reputable authorities on the subject: Financial Stability Board (FSB), National Cyber Security Centre (NCSC), and Cybersecurity and Infrastructure Security Agency (CISA). Our validation

[1] This approach has been applied to the cases of climate change risk (Sautner et al., 2023), epidemic diseases like COVID-19 (Hassan et al., 2023), the Brexit vote in the United Kingdom (Hassan et al., 2022b), and country-level risk (Hassan et al., 2022a).

approach purges out uninformative terms by running, term-by-term, predictive regressions on actual reported cyberattacks. The surviving terms then constitute our primary, validated firm-level quarterly measure $\mathrm{CyberRisk}_{\mathrm{i,t}}$. In addition, we bifurcate $\mathrm{CyberRisk}_{\mathrm{i,t}}$ into discussions that adhere to certain predefined topics. Using extensive text libraries from various sources, we construct four novel and relevant topics: insurance, law, cryptocurrencies, and social media. We also complement our analysis with two existing topics from Hassan et al. (2019): uncertainty and sentiment (the latter coming in positive and negative tones (Loughran and McDonald, 2011)), and the epidemic disease topic from Hassan et al. (2023).

We summarize briefly the main contributions of our paper before giving more details on the structure of the analysis and discussing the literature. Because of our long time series (2002q1-2021q3) and the quarterly frequency of our data, we are able to substantially increase our understanding of cybersecurity risk and its effect on firms and the economy. Our first step is to establish the validity of our measures. We then uncover some novel stylised facts about cyber risk, its time variation and its global and multisectorial dimensions. Importantly, we analyse the financial market footprints of cyber risk with a special attention to option markets. We first show a sizable effect of cyber exposure on the return on assets (RoA) of firms, cash flows and valuations. We then demonstrate using option data that exposure to cyber risk has a significant and large effect on implied volatility, variance risk premium, and downside risk as proxied by the implied volatility slope. We also show that there are important financial spillovers from affected firms to non-cyber-exposed peer firms belonging to the same industry and country. This is an important result indicating the possibility that cyber risk can become more systemic. Exploiting the rich geographic dimension of our data, we fit an extended gravity model to explain the international distribution of cyber risk exposure. Finally, we perform a large number of robustness checks and show some intriguing correlations between cyber risk and other variables such as cryptocurrencies.

To establish the validity of our new measures and understand their properties, we run a series of statistical exercises. First, we document stylized facts on the extent of variation of cyber risk across time, regions, and industries. Aggregate $\mathrm{CyberRisk}_{\mathrm{i,t}}$ has increased considerably after 2013 with the SEC mandate for listed firms to begin to report material cybersecurity incidents and exposure and after 2015 when several high profiles cyber-attacks made headlines. Following the COVID-19 pandemic, the exposure index is currently at its historical peak. Second, $\mathrm{CyberRisk}_{\mathrm{i,t}}$ is concentrated in the United States and in the IT and Services sectors. Interestingly, regional composition has been systematically shifting away from the U.S. and towards the rest of the world over time. Industrial composition, particularly over the past decade, has shifted towards the financial sector. Third, we show that $\mathrm{CyberRisk}_{\mathrm{i,t}}$ can predict realized cyberattacks within 1, 4 or 8 quarters. Fourth, we study balance sheet and income statement characteristics of most affected firms. Cyber-exposed firms are likely to be large with a high share of intangible assets, high liquidity and cash flow ratios, and growth opportunities. Fifth, we conduct a series of case studies for some cyberattacked ("losers") and cybersecurity ("winners") firms. Known cyberattacks, such as the 2017 Equifax breach or the 2019 First American Financial data leak are as expected associated with large spikes in $\text{CyberRisk}_{\text{i,t}}$. Leading cybersecurity firms such as Cisco or CyberArk consistently record high levels of exposure. Finally, we provide detailed earnings call snippets from selected transcripts of heavily exposed firms. Snippets highlight a wide range of intensity and tone of dialogue, ranging from extensive discussions of insurance coverage to identification of foreign state actors as potential orchestrators of incidents. In addition, snippets showcase the importance of Q&A sessions as a significant fraction of cyber and topical words occurs in response to analysts' questions.

To quantify the economic implications of cyber risk, we document that exposure is negatively associated with firms' quarterly stock return performance and positively associated with firms' realized stock market volatility. We further demonstrate that high levels of $\text{CyberRisk}_{\text{i,t}}$ predict worse firm-level economic outcomes such as low cash flow, return on assets, and firm market value. A simple back-of-the-envelope calculation reveals that the global cost of cyber risk exposure amounts to $226 billion per year. This value is in the ballpark of estimates found in other contemporary studies. Our simple calculation does not account for indirect and second-order effects, and so the true financial cost of cyber risk could be substantially larger.

We go further: our main empirical question involves understanding whether cyber risk *exposure*, as opposed to actual incidents, has any effect on firm outcomes. A key advantage of our approach is that we can not only capture discussions surrounding cyberattacked firms at the moment of the incident, but also quantify concerns about *potential* future events that may or may not materialize. In other words, we are, to the best of our knowledge, the first to quantify *uncertainty* stemming from cyber risk exposure. Cyber risk uncertainty may affect investors' beliefs about operational capabilities, resilience of computer and network systems, likelihoods of future attacks or breaches, and thus potentially causes direct monetary or indirect reputational losses. As a result, uncertainty about future cyber risk vulnerabilities may affect asset prices today. In the cross section of firms, the immediate implication is that market-based costs of protection should "price in" greater cyber risk uncertainty emanating from a higher realization of $\text{CyberRisk}_{\text{i,t}}$.

We test this prediction by estimating firm-level and sector-level impacts of $\text{CyberRisk}_{\text{i,t}}$ on equity option market variables.[2] We specifically focus on three measures: implied volatil-

[2] Analysis of the behavior of option prices around newsworthy events has a long tradition in empirical

ity of an option (IV), the variance risk premium (VRP, defined as the difference between IV and realized volatility), and the slope of a linear function that relates implied volatility to moneyness (SlopeD). The implied volatility slope measure is a proxy of downside risk and originates from Kelly et al. (2016) who build on the theoretical framework of Pastor and Veronesi (2013). These three variables reflect the value of option protection against three aspects of risks associated with cyber risk and uncertainty: price risk, variance risk, and tail risk, respectively.

We find strong evidence that cyber risk uncertainty is priced in the option market. Our first main result is that, at the firm level, $\mathrm{CyberRisk}_{\mathrm{i,t}}$ is positively and significantly associated with firms' IV, VRP, and SlopeD. The result is robust to the inclusion of firm and time fixed effects and various controls. The finding is economically significant: switching from no cyber risk exposure to positive exposure increases firms' IV, VRP, and SlopeD by 3%, 1.5%, and 1.6% of the respective variable's standard deviation. To put these numbers in context, Hassan et al. (2019) find that an increase in firm-level political risk, a measure constructed from earnings calls with comparable techniques, raises firms' implied volatility by 1.3%-5.6% of the variable's standard deviation. In addition, Sautner et al. (2023) also employ earnings calls and estimate the impact of firm-level climate risk exposure on IV, VRP, and SlopeD to be in the range of 0.3%-2.43% of variables' standard deviations. These magnitudes are consistent with the view that cyber risk is among first-order sources of risk for firms.

We then move beyond firm-level analysis and ask whether idiosyncratic firm-level cyber risk can be regarded as a source of "systemic" risk for firms and markets. We conduct two exercises that address this question. First, and this is our second key empirical result, we document that $\mathrm{CyberRisk}_{\mathrm{i,t}}$ *spills over* from affected firms to their peers defined as firms that are in the same country and industry as the exposed firms but with zero cyber risk exposure of their own. Analysis of heterogeneous spillover effects reveals that this finding is not driven by a particular tail of the distribution of firm size - a key absorbing characteristic - and is fairly homogenous across the economy. Second, and this is our third important result, we show that cyber risk exposure and uncertainty persist at the sectoral level. We aggregate all variables to the level of an industry and test whether idiosyncratic $\mathrm{CyberRisk}_{\mathrm{i,t}}$ washes out in the aggregate. We find that sector-level effects on RoA and option market variables are strong and statistically significant at both 3 and 4 digit NAICS levels.

We run most of the empirical tests also for our topical measures. The Cyber $\mathrm{Insurance}_{\mathrm{i,t}}$ index stands out on several dimensions. First, it has the highest unconditional pairwise correlation with $\mathrm{CyberRisk}_{\mathrm{i,t}}$ across the whole sample. Second, analysis of earnings call snippets

accounting and finance research. See, for example, Beber and Brandt (2006) on the effects of macroeconomic news and Patell and Wolfson (1979) for corporate earnings announcements.

shows that insurance-related terms are flagged consistently in transcripts of heavily exposed firms. In particular, they frequently appear in the questions that investors pose to firm managers. Third, Cyber $\mathrm{Insurance}_{\mathrm{i,t}}$ has large and significant predictive power for realized cyberattacks. Finally, Cyber $\mathrm{Insurance}_{\mathrm{i,t}}$ is significantly positively associated with firm-level IV, VRP, and SlopeD measures. These findings suggest that insurance considerations are viewed by analysts, investors, and financial markets as especially important when it comes to cyber risk uncertainty.

We supplement our main analysis with additional findings and conclude with several robustness tests. Notably, we test whether our measures are statistically associated with the market price of crypto coins. We document strong contemporaneous, backward, and forward-looking association between the price of Bitcoin (the dominant crypto currency) and our crypto topical measure, suggesting that analyst attention - specifically in the context of cybersecurity discussions in earnings calls - is correlated with crypto price movements. With this auxiliary exercise we do not establish causal linkages but hope to encourage future research to conduct more comprehensive, targeted studies of this issue. Finally, we explain the international distribution of cyber risk exposure with a gravity model extended with measures of financial proximity to the world technological leader - the U.S. We document that our expanded gravity model can explain a large fraction of cross-country variation in cyber risk exposure.

Literature Our paper contributes to the growing literature on the impact of cyber risk on economic and financial performance of firms. Kamiya et al. (2021) employ the Privacy Rights Clearinghouse database and estimate the effects of reported cyberattacks on firm-level stock returns and subsequent economic outcomes. Eisenbach et al. (2022) study how cyberattacks get amplified through the U.S. financial system, with a focus on the wholesale payments network. Crosignani et al. (2022) show that cyberattacks can propagate through firms' supply chain networks by examining the 2017 NotPetya malware attack - one of the most damaging in history. Akey et al. (2021) find that data leaks and breaches cause deterioration in firm value and erase reputational capital, leading firms to rebuild that capital through activities such as corporate social responsibility. Other notable studies in this literature include Biener et al. (2015), Makridis and Dean (2018), Kashyap and Wetherilt (2019), Duffie and Younger (2019), Woods et al. (2019), Jiang et al. (2020), Healey et al. (2021), Lhuissier and Tripier (2021), Tosun (2021), Anhert et al. (2022), Kotidis and Schreft (2022), Anand et al. (2022), Adeney et al. (2022), Aldasoro et al. (2022).

The study closest to ours, Florackis et al. (2022) (FLMW henceworth), leverages textual analysis and information in annual 10-K filings of U.S. listed firms to construct cybersecurity risk proxies. Reassuringly, we find that our indicators and FLMW's measures are broadly similar in trend and cyclical behavior[3]. Our study differs from FLMW substantively along several dimensions. First, the quarterly frequency of earnings calls considerably increases the number of observations and allows for more robust cyberattack forecasting and asset pricing analyses. Second, earnings calls feature Q&A sessions which make cyber-related conversations richer, more unrehearsed, multi-dimensional, and timely. Third, while FLMW test whether cybersecurity risk is priced in the cross section of stock returns, our focus is primarily on the option market and the impact of cyber risk uncertainty on the premia for protection against price, variance, and tail risks.

Furthermore, to complement the literature on direct and firm-level effects of cyber risk, we provide evidence on contagion and systemic effects by establishing that firm-level cyber risk exposure and uncertainty spill over across firms and persist at the sectoral level. In addition, relative to the literature that relies on reported cyberattacks, our approach is robust to the critique that most cyberattacks go unreported and only the largest events get publicized (Amir et al., 2018). Our focus on cyber risk *exposure* is far less likely to suffer from such selection issues: our dataset spans all English-language transcripts of listed firms and during the Q&A sessions of earnings calls analysts pressure firm executives on issues that the latter could potentially ignore or postpone otherwise, rendering timely information disclosure much more probable. We present an explicit example of this during our analysis of transcript snippets.

The methodology of our paper builds on two streams of literature. First, we belong to the growing literature on the applications of textual analysis to "important-but-hard-to-measure" questions in accounting, economics, and finance (Loughran and McDonald, 2011; Baker et al., 2016; Koijen et al., 2016; Loughran and McDonald, 2016; Gentzkow et al., 2019; Neuhierl and Weber, 2020). Second, we borrow from the literature that employs forward-looking option-based risk measures. Option prices have been used for predicting future asset price dynamics (Chang et al., 2013), proxying investment opportunities (Vanden, 2008), and measuring the impact of inflation on public debt valuations (Hilscher et al., 2022). Bollerslev et al. (2009) show that the variance risk premium (VRP) predicts future excess returns. Kelly et al. (2016) show that political uncertainty is priced in the stock option market. That study also introduces the implied volatility slope (SlopeD) measure which we adopt as a proxy of tail risk. Ilhan et al. (2020) find that climate policy uncertainty matters in the cross section of firms and has significant effects on option market variables such as the VRP and SlopeD. Sautner et al. (2023) quantify the impact of firm-level climate change risk exposure on economic and financial outcomes, including option market variables like SlopeD.

[3]We thank the authors for sharing their indices.

The paper is structured as follows. In Section 2 we described the data and our measures, which we empirically validate in section 3. Section 4 describes novel stylised facts on cyber risk in time series and cross section. Section 5 analyses financial market implications with a specific focus on the option market to estimate the effect of cyber uncertainty on firm valuations as well as spillovers and industrywide effects. Section 6 presents additional results pertaining to the gravity model of international cyber exposure and the links between cyber risk and crypto assets, as well as multiple robustness checks. Section 7 concludes.

2 Data and Measurement

2.1 Data

Our primary data source for the construction of cyber risk measures is quarterly earnings conference calls of firms which are publicly listed in the United States from Thomson Reuters' StreetEvents. We have collected 348,393 English-language transcripts that cover 13,024 unique firms from 86 countries over 2002q1-2021q3. Firms normally host one earnings call per quarter, usually within 30 days of the start of each quarter. In our sample there are therefore roughly four observations per firm per year. The structure of each earnings call is typically the following: firm management starts by delivering a prepared speech on issues and topics that they wish to willfully disclose and highlight, followed by Q&A sessions with call participants (e.g. financial analysts). Each call usually lasts around 45 minutes and the average number of spoken words per transcript is less than 8,000. We run a search of cybersecurity-related terms, unigram (single words) or bigrams (two-word combinations) through each conference call in its entirety[4]. As is done typically, all non-alphabet characters are removed. For example, any term with a dash in between (e.g. cyber-risk) gets concatenated into a single word (i.e. cyberrisk). All capitalized letters are kept. Finally, the algorithm does not need the bigram if it already found the first or second word independently as a separate term.

The main source of our option data is the OptionMetrics' Ivy DB Volatility Surface File. We use three option market measures to identify the impact of cyber risk uncertainty: implied volatility, variance risk premium, and implied volatility slope. Uncertainty should be positively related to all three variables. Let $\mathrm{IV}_{\mathrm{t,m}}$ be the implied volatility at time t of an option maturing at $\mathrm{m} > \mathrm{t}$. Following Carr and Wu (2009) and Bollerslev et al. (2009) we compute the variance risk premium (VRP) for each firm as the daily difference between

[4] Combinations that include more than two words are not part of the algorithm due to computational constraints.

implied and realized variance: $VRP_{t,m} = IV^2_{t,m} - RV^2_{t,m}$. The realized variance is computed from daily log returns over the future window (i.e. from *t+1* to *t+m*) that corresponds to the maturity of the option used for implied variance.[5] The VRP captures the cost of protection against general variance risk (or "uncertainty", as pointed out by Bali and Zhou (2016)). We aggregate both the $IV_{t,m}$ and $VRP_{t,m}$ to the firm × quarterly level. Finally, following Kelly et al. (2016) we compute the implied volatility slope variable (SlopeD): this is the steepness of the function that relates IV to moneyness, as measured by the option's Black-Scholes delta[6]. Specifically, we run OLS regressions of $IV_{t,m}$ of Out of The Money (OTM) puts (defined as puts with deltas between -0.5 and -0.1) on deltas and a constant.[7] The resulting slope coefficient constitutes our firm × quarterly SlopeD measure. Higher SlopeD suggests that deeper OTM puts are more expensive, which in turn implies a relatively greater cost of protection against *downside tail risks*. For our baseline analysis, we use 91-day options as this is the maturity that closely corresponds to the quarterly release schedule of earnings calls. We provide robustness results for alternative maturities (30, 60, and 182) in the Online Appendix.[8]

To trace out the association of our exposure measure with realized cyberattacks, we manually merge earnings call announcement data with the Privacy Rights Clearinghouse (PRC) database on reported cyberattacks. Because there is no common firm identifier, we employ a variant of the fuzzy search algorithm. Specifically, we create a vector of integers for each firm name in the PRC and earnings datasets. Then, for each firm in PRC data, we take the cosine distance from each firm in the earnings call data and keep the closest match. To create the vector of integers for a firm name, we count all unique letters, adjacent two-letter, and adjacent three-letter combinations. Finally, we compute a measure of semantic distance (normalized to lie in the [0,1] interval, with 0 implying a perfect match) between firm names in the two datasets. We impose a cutoff (equal to the median distance) to throw out bad matches. We then confirm each surviving match with manual checks. In the end, 293 unique firm-cyberattack pairs are matched to the earnings call data.

[5] Our definition of realized variance follows Kelly et al. (2016) and Ilhan et al. (2020) and is the "ex post" as opposed to an "ex ante" VRP. While our main results do not change if we adopt the ex ante version, using the ex post VRP sharpens our results because the ex ante version is based only on expectations built prior to the actual observation date, which makes results noisier. Capturing the full information set from t to m is particularly important for the case of cyberattacks or exposure spikes, which are difficult to forecast.

[6] Delta measures the rate of change of option value with respect to changes in the underlying asset's price.

[7] We follow Kelly et al. (2016) and Sautner et al. (2023) and ignore the deepest OTM options due to measurement errors in option prices (Hentschel, 2003).

[8] As argued in Beber and Brandt (2006) among others, very short-maturity options' implied volatilities are typically inaccurate due to various sources of measurement error. We therefore do not analyze maturities shorter than 30 days.

Finally, we obtain information on stock prices from the Center for Research in Security Prices (CRSP) and, for each firm-quarter, basic balance sheet and income statement information from Standard and Poors' Compustat. Table 1 provides summary statistics on all main variables used throughout the paper and Appendix A gives details on variable construction and data cleaning steps.

2.2 Term Dictionaries

Our measurement approach follows Baker et al. (2016) and starts with a broad predefined dictionary of words related to cybersecurity risk. Rather than arbitrarily deciding on which specific words to search for by ourselves, we build our starting dictionary from three reputable institutional sources. This starting point is credible because these institutions act as information aggregators on all practical issues related to cyber risk that firms face on a daily basis. In other words, such term libraries include most if not all words that are commonly used in cyber-related discussions of private market participants across industries. These are not just the words that authorities *believe* to be relevant to the topic but an amalgamation of various private and public origins. For example, one of our sources - the National Initiative for Cybersecurity Careers and Studies (NICCS) - has collected terms from a variety of origins.

Our first source of cybersecurity-related words is the Financial Stability Board (FSB) "Cyber Lexicon"[9]. The lexicon comprises 50 terms which, according to the FSB, constitute some of the core terms related to cyber security and resilience. The list is designed to support the work of the FSB, authorities, and private sector agents. Some of the terms include "cyber alert", "malware", "patch management", "vulnerability assessment", etc. Our second source is the "NCSC Glossary" of common cybersecurity terms provided by the National Cyber Security Centre [10]. The list includes 61 terms such as "cyberattack", "botnet", "malvertising", "pharming", "virus", etc. Finally, our third source for the dictionary is the "Glossary of Common Cybersecurity Terms and Phrases" made available by the NICCS, an initiative managed by the Cybersecurity and Infrastructure Security Agency (CISA).[11] This is our most comprehensive source, totaling 164 entries, and including terms such as "spam", "security breach", "attack signature", "incident response", etc.

In total, our library consists of 275 terms which are detailed in full in Table A.1. As we discuss below, not all of them will eventually constitute our baseline firm-level measure because of the dictionary validation procedure.

[9] Available at https://www.fsb.org/2018/11/cyber-lexicon/
[10] Available at https://www.ncsc.gov.uk/information/ncsc-glossary
[11] Available at https://niccs.cisa.gov/cybersecurity-career-resources/glossary

2.3 Dictionary Validation

While our dictionary is very comprehensive, it is potentially problematic if some of its terms are not primarily associated with cyber risk but tend to capture alternative sources of risk and uncertainty. For example, it is not immediately obvious that terms like "hazard" from Table A.1 are necessarily cyber-related. What is an "objective" way to determine which cyber terms are not important or relevant?

Our dictionary validation procedure employs data on actual, realized cyberattacks from the Privacy Rights Clearinghouse (PRC) and preserves only those terms that are useful in predicting future attacks. This approach is agnostic, allowing us not to take an arbitrary stance on any particular sub-set of the library but instead be driven by observed events.[12] It is also arguably the most policy-relevant approach since our validated measure is designed to be potent at predicting future cyberattacks, which could be of particular interest to authorities.[13]

Specifically, suppose the set of bigrams[14] contained in a transcript of firm i in quarter t is $\mathbb{B}_{\mathrm{i,t}}$. Further assume that the set of all cybersecurity terms from our initial dictionary is $\mathbb{C}$. Then, for every c in $\mathbb{C}$, we build a firm-quarter binary variable which takes the value of 1 if c appears anywhere in $\mathbb{B}_{\mathrm{i,t}}$, and 0 otherwise:

$$\mathrm{TermInd}^{\mathrm{c}}_{\mathrm{i,t}} = 1[\mathrm{c} \in \mathbb{B}_{\mathrm{i,t}}], \quad \forall \mathrm{c} \in \mathbb{C} \tag{1}$$

where $1[\cdot]$ is an indicator function. We then estimate, for every c in $\mathbb{C}$, a logistic regression where the main regressor is $\mathrm{TermInd}^{\mathrm{c}}_{\mathrm{i,t}}$ and the outcome variable equals 1 if the same firm i gets cyberattacked within the next $k{=}4$ quarters, excluding the current quarter t, and 0 otherwise. The specification includes quarter and industry (2-digit NAICS codes) fixed effects as well as firm controls: (log) total assets, (log) age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets.[15]

For each term-specific regression we compute the odds ratio (OR), i.e. the ratio of the odds of an attack for firms with positive term-specific exposure divided by the odds of an attack for firms with no term-specific exposure. We then throw out all terms with an OR of

[12] One word of caution related to this approach, which we already mentioned in the Introduction, is that cyberattacks tend to be under-reported (Amir et al., 2018). This implies that our final, validated measure is more conservative than it could have been in the first-best reporting scenario.

[13] One feature of the PRC data is that data coverage is predominantly U.S. centered. However, our final exposure measure is available for firms in all regions. It is unlikely that firms that are cyberattacked in the rest of the world, especially in developed economies, have exposure that is fundamentally different from firms that have high exposure and get attacked in the U.S. In addition, our term libraries are sourced from institutions that are either international in nature or service market participants worldwide.

[14] Henceforth, we use the term bigrams to denote both unigrams and bigrams in order to ease exposition.

[15] Variable construction is detailed in Appendix A and summary statistics are reported in Table 1.

less than or equal to 1 and keep the rest. In other words, we are only interested in keeping terms that have a positive impact on the likelihood of future cyberattacks. In total, there are 117 unique terms that remain, meaning that 158 terms have been parsed out for one of the following reasons. First, combinations that include more than two words are not part of our algorithm due to computational constraints. For example, terms such as "access control mechanism" from CISA cannot get picked up. Second, any duplicates (lower or upper case) across libraries get ignored. Third, some terms have 0 counts across all transcripts and quarters and we treat them as "missing". The first three steps leave us with 229 unique working terms. 49 of the remaining terms yield an OR of exactly unity due to very low count frequency (e.g. the bigram "tabletop exercise" from CISA has a global count of 2). Finally, 63 terms yield an OR of strictly less than unity. We do not discard these terms since they may yet possess useful information. We will return to them in the robustness Section 6.

We label the set of all validated terms as $\tilde{\mathbb{C}}$. Table 2 lists all terms in $\tilde{\mathbb{C}}$ and sorts them by absolute frequency. We have concatenated any bi-grams into single words for readability. The 25 most frequent terms are "data", "software", "digital", "network", "accountability", "availability", "computer", "compromise", "disclosure", "spam", "router", "vulnerabilitymanagement", "domain", "encryption", "firewall", "antivirus", "confidentiality", "datasecurity", "bug", "app", "accessmanagement", "criticalinfrastructure", "vpn", "identitymanagement", and "ict". These include some potentially risk-related terms (e.g., "compromise", "vulnerabilitymanagement"), opportunity-related terms (e.g., "computer", "app"), but also more neutral business-related terms (e.g., "data", "availability").

2.4 Firm-Level Cyber Risk Exposure

We are now ready to construct our baseline measures of firm-level cyber risk exposure $\mathrm{CyberRisk}_{i,t}$. We define three variants of the same measure. First, absolute frequency ($\mathrm{CyberRisk}^{A}_{i,t}$) which is the number of times terms from $\tilde{\mathbb{C}}$ appear in each earnings-call transcript. Second, relative frequency ($\mathrm{CyberRisk}^{R}_{i,t}$) which is $\mathrm{CyberRisk}^{A}_{i,t}$ scaled by the total number of words in each transcript $\mathrm{B}_{i,t}$. Finally, a binary indicator ($\mathrm{CyberRisk}^{I}_{i,t}$) that takes the value of 1 if any of the terms in $\tilde{\mathbb{C}}$ appears in the transcript, and 0 otherwise:

$$\begin{aligned}\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}} &= \sum_{\mathrm{b}}^{\mathrm{B_{i,t}}} \left(1[\mathrm{b} \in \tilde{\mathbb{C}}]\right) \\ \mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}} &= \frac{\sum_{\mathrm{b}}^{\mathrm{B_{i,t}}} \left(1[\mathrm{b} \in \tilde{\mathbb{C}}]\right)}{\mathrm{B_{i,t}}} \\ \mathrm{CyberRisk}^{\mathrm{I}}_{\mathrm{i,t}} &= 1\left[\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}} > 0\right]\end{aligned} \tag{2}$$

where $1[\cdot]$ is an indicator function.

Our measurement approach, together with the dictionary validation step, can be viewed as a particular weighting scheme that weights terms based on their ability to predict future attacks. One can rewrite our definitions of $\mathrm{CyberRisk}_{\mathrm{i,t}}$ in terms of the un-validated dictionary set $\mathbb{C}$ but with a weighting scheme $\mathrm{w_b}$ that assigns a value of 0 for terms for which the predictive logistic regression Odd Ratio is ≤ 1, and a weight of 1 otherwise. Such representation is consistent with the canonical weighting scheme in the text classification literature where $1[\mathrm{b} \in \mathbb{C}]$ is the *term frequency* and $\mathrm{w_b}$ is the *binary* term weight (Salton and Buckley, 1988; Hassan et al., 2019; Engle et al., 2020). The resulting two terms $1[\mathrm{b} \in \mathbb{C}] \times \mathrm{w_b}$ - yielding a weighted sum of cyber-related bigrams - would then produce the same values for $\mathrm{CyberRisk}_{\mathrm{i,t}}$ as in Equations 2.[16]

2.5 Topical Analysis

In addition to our baseline cyber risk exposure measure we also construct a series of joint-search queries between cyber bigrams and other topics of special interest. Our goal is to construct topical indices that are related to cyber risk chatter and may also be useful for the literature. Instead of picking topical categories exogenously, we first establish common contexts to cyber risk conversations based on a detailed manual reading of 250 earnings-call transcripts (which include a sample of known cyberattacked firms, cybersecurity firms, and transcripts with a higher than median exposure that were selected at random). We narrow down the list of particularly cyber-relevant topics to four: Insurance, Law, Cryptocurrencies, and Social Media. These topics, in various circumstances and degrees of intensity, get dis-

[16] Though not shown in the paper, we have also constructed the *inverse transcript frequency* measure (Gentzkow et al., 2019; Sautner et al., 2023) as $\mathrm{CyberRisk}^{\mathrm{ITF}}_{\mathrm{i,t}} = \frac{\sum_{\mathrm{b}}^{\mathrm{B_{i,t}}}\left(1[\mathrm{b}\in\tilde{\mathbb{C}}]\times\log\left(\frac{\mathrm{N}_{\mathbb{T}}}{\mathrm{f_{b,\mathbb{T}}}}\right)\right)}{\mathrm{B_{i,t}}}$, where $\mathrm{N}_{\mathbb{T}}$ is the total number of transcripts and $\mathrm{f_{b,\mathbb{T}}}$ is the number of transcripts where the bigram b gets a positive count. This robustness exercise accounts for fluctuations in the importance of individual bigrams. The correlation between $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ and $\mathrm{CyberRisk}^{\mathrm{ITF}}_{\mathrm{i,t}}$ is 98.45%; results do not change and are omitted for brevity.

cussed regularly among highly-exposed firms. For example, the issue of cyber risk insurance (costs, breadth of coverage) gets mentioned consistently in the transcripts of affected firms. Another example is the well-known reliance of cybercrime activists on crypto coins as the currency of cyber-ransomware.

In order to build the four topical indices in a systematic manner, we construct topic-specific text libraries based on various publicly available sources. First, for the Insurance topic we source 227 bigrams from the "Glossary of Insurance Terms" by the National Association of Insurance Commissioners (NAIC).[17] The vocabulary is developed by NAIC researchers and is based on various insurance references. Second, for the Law topic we obtain 150 terms from the "Glossary of Legal Terms" from U.S. Courts. The library is maintained by the Administrative Office of the U.S. Courts on behalf of the Federal Judiciary.[18] Third, for the Crypto topic we collect 205 terms from the "Cryptopedia" which is powered by Gemini - a cryptocurrency exchange and custodian.[19]. Finally, for the Social Media topic we were unable to find a single institutional source that would cover all terms of interest and instead have included 78 unique terms from various sources. The full topical libraries are provided in Table A.2.

In addition to the four novel topics that we describe above, we also source four existing topics from other studies. First, the Political Risk and Uncertainty topics from Hassan et al. (2019). Second, the positive and negative sentiment dictionary of Loughran and McDonald (2011), which we label simply as Sentiment. And finally, the Diseases topic from Hassan et al. (2023), a library which includes COVID-19 (and other epidemic diseases) related vocabulary. For details on the composition of each library we refer the reader to the relevant respective paper.

We validate each topical library with a similar procedure as in Section 2.3. First, for every c in $\mathbb{C}$, we build a binary variable which takes the value of 1 if c appears anywhere in $\mathbb{B}_{\mathrm{i,t}}$ *and* occurs in proximity to any topic-specific term k, and 0 otherwise. We repeat this step for each of the eight topics:

$$\mathrm{TopicInd}_{\mathrm{i,t}}^{\mathrm{c,k}} = 1[\mathrm{c} \in \mathbb{B}_{\mathrm{i,t}}] \times 1[\mathrm{c} - \mathrm{k}] < \mathrm{Z}, \quad \forall \mathrm{c} \in \mathbb{C} \tag{3}$$

where $\mathrm{Z} = 50$ words for the four novel topics: Insurance, Law, Crypto, and Social Media. For consistency with the original studies, we keep $\mathrm{Z} = 10$ for the remaining four existing topics: Uncertainty, Sentiment, Politics, and Diseases.

Next, we run the same logit regressions of the cyberattack indicator on each topical

[17] Available at https://content.naic.org/consumer_glossary
[18] Available at https://www.uscourts.gov/glossary
[19] Available at https://www.gemini.com/cryptopedia

indicator variable, plus the usual controls and fixed effects. For each topical search we keep only those cyber terms for which the OR is greater than unity. In other words, we construct eight sets of validated cybersecurity libraries $\tilde{\mathbb{C}}^{k}$, one per each topic. Having built the validated topical libraries, we measure topical cybersecurity exposure by counting the number of times terms from each $\tilde{\mathbb{C}}^{k}$ appear in each transcript. For completeness, we show the definition of a relative-frequency topical measure below:

$$\text{CyberRisk Topic}^{\text{R}}_{\text{i,t}} = \frac{\sum_{\text{b}}^{\text{B}_{\text{i,t}}} \left(1[\text{b} \in \tilde{\mathbb{C}}^{\text{Topic}}]\right)}{\text{B}_{\text{i,t}}} \tag{4}$$

As before, superscript R stands for relative frequency. Absolute frequency and binary variants of each measure are built accordingly. The net sentiment measures are defined as: $\text{CyberRisk NetSentiment}_{\text{i,t}} = \text{CyberRisk PosSentiment}_{\text{i,t}} - \text{CyberRisk NegSentiment}_{\text{i,t}}$.

Before we proceed with further validation steps and statistical analyses, it is useful to briefly summarize our constructed cybersecurity measures. Table 1 provides basic summary statistics for all measures in absolute frequencies. The average number of counts, per transcript (p.t.) across all quarters, is 1.33. The range of counts is wide: from 0 to 244. Among topical measures, the highest average count is for Insurance (0.37 p.t.). Average net sentiment is negative: -0.18. The Disease topic recorded an average count of close to 0 with the maximum of just 3. For the majority of our statistical exercises we will therefore ignore the Disease topic. Table D.1 shows pairwise correlation coefficients between all our measures, in relative frequencies, together with p-values in the parentheses. The Insurance topic has the highest unconditional correlation with the baseline measure (0.659 with statistical significance at the 1% level), followed by Negative Sentiment (0.530 with statistical significance at the 1% level). Net Sentiment is strongly negatively correlated with the baseline measure (correlation coefficient of -0.425). Figure C.1 in the Online Appendix shows distributions of term frequencies in the form of histograms. Generally, all distributions are highly right-skewed. Tables C.1 and C.2 provide additional summary statistics by country and industry, respectively.

3 Validation

In this section we validate our baseline exposure measures with a series of tests. First, we test whether our measures pick up high exposure from affected (cyberattacked) firms and cybersecurity providers. Second, we use our measure to predict actual, reported cyberattacks. Third, we provide detailed snippets of select transcripts of heavily exposed firms. Finally, we compare our measures to complementary indices built in Florackis et al. (2022) on the basis of 10-K filings.

3.1 Case Studies

The first major validation test of our baseline measure - $\text{CyberRisk}_{\text{i,t}}$ - is whether it can pick up high exposure for firms that we know should be heavily exposed. This may be because the firm reported a cyberattack or because the said firm is involved in the IT services sector and thus must be exposed by the nature of its business.

We begin with case studies of 9 well-known historical cyberattacks. First, in 2017q3, the American credit bureau Equifax reported that private records of about 150 million American and 15 million British citizens were stolen. To this day, the Equifax breach remains one of the biggest data compromises in history. Second, the 2017-2018 Bank of Montreal breach. BMO acknowledged that vulnerabilities in its online banking applications, existing between June 2017 and January 2018, allowed attackers to breach its security safeguards, take over online banking accounts, and exfiltrate the personal information of 100,000 of its customers in two separate attacks (OPC, 2021). Third, the 2018-2019 Marriott Hotels cyber incident, which led the UK's data privacy watchdog to fine the Marriott Hotels chain £18.4m for a major data breach that could have affected up to 339 million guests.[20] Fourth, the 2013 Adobe data compromise where it was believed that usernames and encrypted passwords had been stolen from about 38 million of the company's active users.[21] Fifth, the 2019 First American Financial (the second largest U.S. title insurer) data leak announcement that left exposed approximately 885 million records related to mortgage deals going back to 2003. The firm was charged by New York's top financial regulator over the cybersecurity gap.[22] Sixth, the 2013 Target data breach that affected 40+ million customers. The company was forced to pay an \$18.5 million multistate settlement, the largest ever for a data breach at the time.[23] Seventh, the 2014 Home Depot data breach which forced the firm to pay a \$17.5 million settlement to resolve a multistate probe into the breach where hackers accessed payment card data belonging to 40 million customers.[24] Eighth, the 2020q4-2021q1 SolarWinds cyberattack where advanced persistent threat (APT) actors infiltrated the supply chain of SolarWinds, inserting a backdoor into the product of the sotware developer. In January 2021, a class action lawsuit was filed against SolarWinds in relation to its security failures and subsequent

[20]https://www.bbc.co.uk/news/technology-54748843

[21]https://www.bbc.co.uk/news/technology-24740873

[22]https://kfgo.com/2020/07/22/new-york-charges-big-title-insurer-first-american-over-security-gap/

[23]https://eu.usatoday.com/story/money/2017/05/23/target-pay-185m-2013-data-breach-affected-consumers/102063932/

[24]https://www.reuters.com/article/us-home-depot-cyber-settlement-idUSKBN2842W5

fall in share price.[25] Ninth and finally, as was reported in 2021q3, a Chinese software developer illegally collected more than 1.1 billion pieces of user information from Alibaba's Taobao shopping platform before Alibaba noticed the scraping.[26]

Figure 1 depicts the dynamic of (standardized) $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ for the aforementioned 9 cyberattacked firms. We notice that the index correctly captures the exact timing of each incident in most cases. For example, it spikes by one or more standard deviations for Equifax in 2017q4, Bank of Montreal in 2018q1, SolarWinds in 2021q1, or Target in 2014q1. The figure also plots $\text{CyberRisk Insurance}^{\text{R}}_{\text{i,t}}$ and $\text{CyberRisk NetSentiment}^{\text{R}}_{\text{i,t}}$ - the two topical indices that are most strongly correlated with the baseline measure. Spikes in $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ around cyber incidents are consistently associated with increases in $\text{CyberRisk Insurance}^{\text{R}}_{\text{i,t}}$ and sharp declines in $\text{CyberRisk NetSentiment}^{\text{R}}_{\text{i,t}}$. Conversations around realized cyber events are pessimistic in nature and involve a large amount of insurance-related nuance.

In the Online Appendix, we also look at 6 of the world's largest listed cybersecurity firms by revenue (as of 2021q4). Cisco Systems, CyberArk, Jupiter Networks, Oracle, Palo Alto Networks, and Synopsys. Figure D.1 plots the time series of $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$, $\text{CyberRisk Insurance}^{\text{A}}_{\text{i,t}}$, and $\text{CyberRisk NetSentiment}^{\text{A}}_{\text{i,t}}$ for these companies. In absolute terms, firms such as these consistently record counts that are in the right tail of the distribution. For example, the average absolute frequency over time for Oracle is 7.81 counts per transcript (with a standard deviation of 8.67), which is several times the sample average. Interestingly, the baseline measure of Cyber Risk is also strongly positively correlated with $\text{CyberRisk Insurance}^{\text{A}}_{\text{i,t}}$ and negatively correlated with $\text{CyberRisk NetSentiment}^{\text{A}}_{\text{i,t}}$. The latter relationship suggests that even for cybersecurity-related service providers net sentiment is generally negative.

3.2 Predicting Cyberattacks

Our second validation step is a test of predictability of realized, reported cyberattacks. Recall that each term that constitutes $\text{CyberRisk}_{\text{i,t}}$ has been validated to be able to predict realized cyberattacks individually. Our measures are thus engineered such that they are forward looking and have predictive power; we believe that this is a fundamental quality of any reliable cybersecurity exposure measure. In order to confirm and quantify our measures' predictive ability, we run a similar specification as in the dictionary validation exercise. Specifically, we run a quarterly firm-level logit regression of the cyberattack indicator variable on our measures, plus sector and quarter fixed effects and the usual firm controls (size, age,

[25] https://www.cisecurity.org/solarwinds

[26] https://www.wsj.com/articles/alibaba-falls-victim-to-chinese-web-crawler-in-large-data-leak-11623774850

Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets). To gauge the extensive and intensive margins of cyber risk exposure, we use as our main regressor either $\mathrm{CyberRisk}^{\mathrm{I}}_{\mathrm{i,t}}$ or $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ (std.). We allow for three different specifications where the cyberattack indicator takes the value of 1 if the attack takes place within 1, 4, or 8 quarters (excluding the current quarter), and 0 otherwise.

Table 3 reports the results. Panel A (B) shows results for $\mathrm{CyberRisk}^{\mathrm{I}}_{\mathrm{i,t}}$ ($\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$). In both panels, odd (even) columns show results without (with) all firm controls. In columns (1)-(2) the cyberattack occurs within 1 quarter, (3)-(4) - within 4 quarter, and (5)-(6) - within 8 quarters. Across twelve specifications that we report, we see that our measure has a significant positive effect on the OR of future cyberattacks. The extensive margin of exposure is particularly strong, as can be seen from Panel A: going from zero to positive cyber risk exposure increases the OR of an attack by 33.7% within 1 quarter (column 2), and by 35.3% within 4 and 8 quarters (columns 4 and 6). In Panel B, the main regressor - $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ - is standardized so that the interpretation of the intensive margin is the following: a one-standard-deviation increase in $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ increases the OR of an attack by 13.2% within 1 quarter (column 2), by 13.5% within 4 quarters, and by 15.9% within 8 quarters (column 6). In absolute frequency terms, one standard deviation of $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ equals approximately 3.2 counts per transcript.

Table D.2 shows the results for topical cyber risk measures. Our main regressor of interest in this instance is CyberRisk x $\mathrm{Topic}^{\mathrm{I}}_{\mathrm{i,t}}$, i.e. topical indicator variables. For simplicity, we focus on the cyberattack indicator which takes the value of 1 if the attack takes place within 8 quarters (excluding the current quarter), and 0 otherwise. Results show that the Insurance, Law, and Negative Sentiment have large and significant effects on the attack odds ratio. The magnitudes are 1.443, 1.619, and 1.536, respectively. Recall that the corresponding value for the baseline measure and horizon is 1.353 (Panel (a), column (6) in Table 3). This suggests that topical analysis improves predictability of actual cyberattacks. In the case of Insurance, Law, and Negative Sentiment topics predictability has improved by 6.67%, 19.67%, and 13.5%, respectively. For the other topical indices we do not find any significant effects.

3.3 Snippets

In order to provide further context and color on cybersecurity related chatter, and to complement our case-study analysis, we now share and discuss snippets from earnings calls transcripts of select firms. We identify exact transcripts (firm x quarter combinations) with significant spikes in $\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}}$ around six known cyber incidents: Equifax Inc in 2017q4, Target Corp in 2014q1, SolarWinds Corp in 2021q1, First American Financial in 2020q3, Home Depot Inc in 2015q1, and Marriott International in 2019q1. We also show snippets of three large cybersecurity firms: Cisco Systems from 2018q4, Oracle Corp from 2020q2, Palo Alto Networks from 2019q3.[27] In every snippet, terms of interest that are identified by our algorithm are highligted by dashes, e.g. -personalinformation-. We concatenate all bigrams into unigrams for consistency and remove all capital letters. Apart from these modifications, we do not make any linguistic cosmetic tweaks to any sentence and present text as it appears in transcript files exactly. Note that some grammatical mistakes are to be expected since since these texts are transcribed from audio files.

Table B.1 presents the snippets along with the $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$ count. The Equifax Inc. snippet is one of the most illustrative ones we have encountered. For example, just the first few lines concern the potential identity of the attacker: "has there been any further progress in identifying whether the hack was done by a foreign state actor"; as well as insurance for the incident: "how youre thinking about total costs of the breach and how much youre accruing for breach costs". A variety of terms is captured, ranging from -breach- to -cyberevent-, -securitysystems-, -personaldata-, and -data-. The Insurance and Legal topics receive a considerable degree of coverage with terms such as -insurance-, -cost-, and -policy-. The role of the Q&A session is also apparent from a line that is clearly a question from an analyst that is addressed at the Equifax manager: "whats your overall level of comfort that the majority of the cyber costs would be covered by -insurance- as opposed to being more equifax ultimately?" The immediate response from the manager was "yes so were not going to specifically disclose the specific amount of the coverage". This reply demonstrates explicitly that the company would most likely not have provided additional detail on an important topic (cyber insurance coverage) if not for the direct question by the call participant. Thus, the Q&A session at the end of each earnings call is essential for uncovering material information about exposure.

The remaining snippets showcase how our algorithm captures a variety of information from "announcement that -criminals- had -gained- access to guest payment card -data-" (Target Corp) to "we could not find -compromise- that was idiosyncratic to the solarwinds environment" (SolarWinds Corp) and "time of the -incident- and the adequacy of our -disclosure- controls there are also class actions pending" (First American Financial). One of the recurring themes is that the term -breach- seems to be effective at picking up realized incidents. Another consistent observation is that the Insurance topic is very prevalent in virtually every snippet. The context of snippets of the three cybersecurity firms is slightly different. Discussions center around more business-related terms such as -data-, -computer-

[27] More snippets can be made available upon request.

, -informationtechnology-, and -digital-. For example, the top line from the Oracle Corp snippet reads: "i want to explain why were -computer- oracle cloud infrastructure is the worlds only secondgeneration autonomous cloud autonomous software". However, there are still conversations about data breaches such is in Palo Alto Networks' top line: "leadership position and customer happiness and customer success out in -breach- market not only that we are not going to rest on our laurels". However, in these contexts, companies are discussing breaches that affected their clients or the market in general, not necessarily their own businesses. All in all, analysis of text snippets reveals that the algorithm does a fairly good job at capturing exposure of both negatively and neutrally/positively affected firms.

3.4 Comparison to Florackis et al. (2022)

As a final validation check, we compare our measures to cybersecurity risk proxies that were developed in Florackis et al. (2022) (FLMW, henceforth). The reason why this is a useful comparison for us is two-fold. First, like us FLMW use natural language processing techniques and textual analysis. Second, they leverage 10-K filings that listed firms supply to the SEC. First-quarter investor earnings calls are typically held soon after the Form 10-K (i.e. annual report) is made public. Thus, our indices should be able to pick up the same slow-moving trends in cyber risk exposure.

Since the baseline FLMW index is only available at the yearly level, in order to compare and contrast our measures at higher frequencies, we proceed with the comparison of *factors*. Specifically, we construct a simple cybersecurity risk factor in two basic steps. First, at the end of each quarter - in line with the release schedule of earnings calls - we sort all stocks in CRSP into two groups based on our $\mathrm{CyberRisk}_{\mathrm{i,t}}$ measure.[28] Second, we build value-weighted portfolios for each group (which we label high- and low-cyber-risk) at the quarterly frequency. The factor is then computed as the difference between returns on the high- and low-cyber-risk portfolios. We have obtained the daily factor from FLMW, whom we thank for sharing this data, and aggregated to the quarterly frequency.

Figure D.2 in the Online Appendix plots the result of this exercise in two panels. Panel (a) shows our baseline $\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}}$ index (bottom x-axis) (quarterly) and the main index from FLMW (top x-axis) (annual). Both series have been standardized. As can be seen, both measures are picking up a similar rise in cyber risk exposure. Panel (b) plots the quarterly cybersecurity risk factor from FLMW together with our own factor. Both series have been CAPM residualized and standardized. The correlation coefficient between the two series is

[28] Our baseline approach is to sort based on $\mathrm{CyberRisk}^{\mathrm{I}}_{\mathrm{i,t}}$ and thus have all stocks with zero exposure in group one and stocks with positive exposure in group two. Sorting based on (the median of) $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ yields the same results.

0.39 with a p-value of 0.0186. These results suggest that our measures are in line with the information that one can extract from 10-K filings. Additionally, the longer time series, the quarterly nature of our indices and the fact that risk factors are not correlated perfectly indicate that our measures bring new information and value-added to the literature.

4 Cyber Risk Facts and Trends

In this section we discuss time-series, regional, and sectoral properties of $\mathrm{CyberRisk}_{\mathrm{i,t}}$. We also study firm-level determinants of high exposure.

4.1 Time Series

Figure 2 plots the time series of $\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}}$ and $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ on the left panel (a). Recall that $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ adjusts for transcript length while $\mathrm{CyberRisk}^{\mathrm{A}}_{\mathrm{i,t}}$ simply measures the absolute frequency (number of counts). We can observe a sharp, three-fold increase in both measures over the past decade, starting from around 2013. This structural break closely corresponds to the 2011-2012 SEC mandate for listed firms to begin to report material cybersecurity incidents and exposure. Another possible explanation is that 2013 is the year of the Snowden leaks and a year where hackers operated on a massive scale: Target was attacked in 2013q4-2014q1 by the POS malware and 40 million clients were affected. Adobe was also hacked in 2013q4 (153 million people were affected). Furthermore, 2014q4 saw the high profile hacking of Sony by North Korea. It is therefore possible that these very salient events were both the symptoms of and increased the awareness of cyber risk going forward. For completeness, panel (b) plots relative frequencies of our three underlying source dictionaries: FSB, NCSC, and CISA. These are the raw measures, i.e. not validated with realized cyberattacks. All three measures have steadily and similarly risen over the past decade.

Figure 3 plots the time series of our 8 topical indices. Panel (a) shows our 4 novel topics: Insurance, Law, Crypto, and Social Media, while Panel (b) shows the 4 existing topics from Hassan et al. (2019) and Hassan et al. (2023): Uncertainty, Net Sentiment, Politics, and Diseases. All measures are in relative frequency and have been standardized. CyberRisk $\mathrm{Insurance}^{\mathrm{R}}_{\mathrm{i,t}}$ stands out as an index that seems to track the baseline $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ closely: it has risen roughly by the same magnitude since 2013. We also verify this statistically by reporting that the pairwise correlation between these two indices is the highest among all pairs (0.66 with a p-value of 0.00). The Law index has interestingly trended down and become less prominent in relative terms. The Social Media topic was at its highest in the 2011-2014 period and has dwindled down since then. That episode coincides with a surge in phishing attacks that targeted social media companies.[29] We see that the Crypto topic has spiked in the latter part of 2020 and first half of 2021, which coincide with the local peaks in the price of Bitcoin. Interestingly, 2017q4 is another local peak of the Crypto topic, which also historically coincided with high Bitcoin prices. We return to this question in Section 5.6. The Politics topic peaked around 2016-2017, coinciding with the U.S. presidential election and the onslaught of international state-sponsored cyberattacks in 2017. Net Sentiment surrounding cybersecurity discussions is currently severely negative, having reached its global negative peak during the COVID-19 pandemic as the number of attacks increased multi-fold. In absolute frequency terms (not shown), Net Sentiment is negative on average and has been trending down heavily over the past 10 years. Finally the Diseases topic is generally close to 0 and peaks around known pandemics (COVID-19 pandemic in 2020, the 2014-2016 Ebola outbreak in Africa, and the 2009 H1N1 pandemic).

4.2 Decomposition by Region

We now provide decompositions of cyber risk exposure by region. Figure 4 presents the 2021 global heatmap of $\text{CyberRisk}^{A}_{i,t}$ by region, defined as the location of firms' headquarters. The most exposed regions are the United States and Canada, Western Europe, the UK, Australia and some parts of Asia such as India, Japan and China. In Latin America, Brazil is most at risk followed by Chile and Mexico. Figure 5, Panel (a), shows the evolution of the regional composition of exposure over time. We observe that the vast majority of cyber chatter still originates in US-based firms. However, this trend has been going through a structural change since the beginning of our sample. Cyber risk is becoming an increasingly global phenomenon affecting all continents, in particular Europe and Asia.

Gravity Model of International Cyber Risk Exposure What can explain the rich geography of cyber risk exposure in Figure 4? A natural candidate would be the "gravity" model, which has proven to be highly successful not only at explaining the patterns of cross-border goods trade but also financial asset flows (Portes et al., 2001; Portes and Rey, 2005) and therefore generally interactions across countries. Ingredients of a textbook gravity model include market size in origin and destination countries, as well as distance - a proxy for bilateral trade and information costs. In our context, as the main dependent variable we use country-level aggregate exposure $\text{CyberRisk}^{A}_{c,t}$ where subscript c stands for a country. We use the United States as the centre country because the U.S. has the greatest number

[29] https://www.kaspersky.com/about/press-releases/2013_kaspersky-lab-report-37-3-million-users-experienced-phishing-attacks-in-the-last-year

of raw cyber risk related keyword counts, and because it is the global leader in technology and innovation. Thus, countries that are larger in market size and/or closer to the U.S. (in terms of having lower bilateral transaction costs) should also plausibly have higher cyber risk exposure. We estimate a panel regression for the sample of 40+ countries, for which we were able to obtain gravity-related data, and over 2005-2019.

Table 12 presents the results with $\text{CyberRisk}^{\text{A}}_{\text{c,t}}$ as the outcome variable and various sets of covariates. Column (1) shows results from the basic specification with real Gross Domestic Product (GDP) and GDP per capita in country c, as well physical distance to the U.S. Both measures of size are positive and statistically significantly different from zero at the 1% level. Moreover, the coefficient on distance is negative and statistically significant. The R^2 in excess of 50% suggests that the basic gravity model explains a considerable fraction of the variation in global cyber risk exposure. In subsequent columns, we consider various extensions to the basic model starting from the "institutions block" that includes dummies for common legal origins, common language, and a common colonizer. This data was obtained from the U.S. International Trade Commission. Then, we introduce goods trade as proxied by manufacturing exports to and from the U.S. This data comes from the World Trade Organization.

Finally, we introduce portfolio investment, to and from the U.S, total or by asset class (equity vs debt). This data is from the IMF's Coordinated Portfolio Investment Survey. In our richest general specification with total portfolio flows (column (4)), we find that market size, the common legal origin dummy, the common colonizer dummy, as well as portfolio investment into and from the U.S. are all important. Investment originating from the U.S. is relatively more economically significant than flows into the U.S.. In columns (5) and (6) we further find that equity portfolio flows appear to be more important than debt flows. The importance of portfolio flows (particularly equity) as a predictor of international cyber security risk is very interesting and consistent with the view that large and interconnected advanced economies, which are responsible for the bulk of international portfolio investments are, like the US, the most susceptible to be cyber exposed. In addition, the modified gravity model tells as that the more they are connnected to the US, the more exposed to cyber they are. To additionally visualize the above results, Figure 9 presents binned scatter plots of the relationship between $\text{CyberRisk}^{\text{A}}_{\text{c,t}}$ and GDP, distance from U.S., portfolio investment from U.S., and portfolio investment into U.S.

4.3 Decomposition by Industry

Figure 5, Panel (b), decomposes $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$ by industry, proxied by two-digit NAICS codes. It shows the industrial composition of all discussions in percent. We document that the IT and services sectors (which include various IT-related consulting companies) have historically dominated our exposure measures, and understandably so. However, since about 2013 the percentage of cyber risk discussions attributed to the finance sector has been steadily growing and currently stands at about 20%. In other words, one fifth of all worldwide cyber risk related discussions now occurs in the finance industry. This compositional change has taken place seemingly at the expense of the decline in manufacturing and IT firms.

Panel (c) of Figure 5 offers a more granular look at the financial industry. The breakdown of cyber exposure based on 4-digit NAICS codes appears to be broadly 45% for financial intermediaries, 35% for insurance companies, 15% for broker-dealers, and 5% for all the rest. Interestingly, the insurance sector has been steadily exposed to cyber risk with a mild decline in the recent years. Within the financial intermediary sector, the most exposed types are depository institutions (banks), followed by other intermediaries (e.g. mortgage companies) and non-banks.

4.4 Determinants of Firm-Level Cyber Risk Exposure

What are the characteristics of firms that have high cyber risk exposure? In order to answer this question, we merge the quarterly earnings call data with Compustat and CRSP and construct an array of firm-level balance sheet and income statement characteristics. Variable construction is detailed in Appendix A. Our main model is a probit regression of firm characteristics on $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$. Recall that this indicator variable takes the value of 1 if a transcript records positive exposure, 0 otherwise. The same exercise is run on all of our topical indices. All specifications include country, industry and time fixed effects, unless specified otherwise.

Table 4 reports the results. Overall, we see that firms which have a higher likelihood of having positive exposure to cyber risk typically fit into the following profile: high ratio of intangible assets to total assets, high liquidity, high growth opportunities (as proxied by Tobin's Q), and large size (as measured by total assets). These characteristics seem to be recurring across studies who look at determinants of cyberattacks or exposure (Kamiya et al., 2021; Florackis et al., 2022). For most of our topical indices, we see that these four firm characteristics are the most robust predictors of exposure. In terms of explanatory power, the pseudo-R^2 of our regressions is at most 0.244; a large fraction of cyber risk exposure is left puzzlingly unexplained. In the Online Appendix we provide three sets of additional results where we explore heterogeneity by region, industry, and financial sub-sector. Tables C.3, C.4, and C.5 report those results. The importance of size and liquidity ratios are relatively homogenous across industries and regions. However, there is wide heterogeneity for other characteristics whereas certain characteristics are more prevalent for certain areas or sectors. For example, intangibles and Tobin's Q are important for the U.S., Americas, and Europe but not for U.K. firms for whom the S&P Rating seems to be more useful. Another example is that intangibles seem to be only important for the Finance and Real sectors and the "other" sector but not for trade or IT. Within the financial sector, intangibles matter most for Broker-Dealers, while size seems to be again a robust predictor of exposure. Liquidity ratios are a good predictor for banks, non-banks and other intermediaries (such as mortgage companies) but not so much for insurance.

5 Cyber Risk and Economic Implications

The key research question of our paper is whether cyber risk exposure and uncertainty have meaningful economic implications. In this section we first document that our measures impact realized firm-level stock market and balance sheet aggregates. We then turn to the option market and study firm- and sector-level option market outcomes. Finally, we trace out spillovers from affected to non-affected peer firms.

5.1 Stock Market Effects

The first test of economic significance is whether our measures of cyber risk exposure have any economically meaningful effects on firms' stock market performance. Recall that cyber risk *exposure* does not necessarily imply an actual incident; it is fundamentally a forward-looking measure which implies a heightened likelihood of a future cybersecurity crisis or event. This uncertainty alone can affect asset prices today. To test this theory, we run quarterly firm-level regressions of (std.) weighted stock returns (WRet), cumulative stock returns (CRet), and realized stock market volatility (RV) on $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ and $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (std.). Specifications include firm and quarter fixed effects as well as the usual firm controls.

Results are reported in Table 5. First, we find that both $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ and $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ have negative and significant effects on stock returns, as can be seen from columns (1)-(2) and (4)-(5). The extensive margin, as in the case with cyberattack forecasting, is especially strong: switching from zero to positive cyber risk exposure lowers cumulative quarterly stock returns by 1.1% of the variable's standard deviation. Similar magnitudes have been obtained elsewhere in the literature (Kamiya et al., 2021; Tosun, 2021). Second, both $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$

and $\text{CyberRisk}_{\text{i,t}}^{\text{R}}$ have a large and significant positive effect on realized volatility in the order of 1.4%-2.1% of the variable's standard deviation. The fact that cyber risk exposure is associated with elevated volatility is an important validation of our measure.

5.2 Balance Sheet Effects

We now ask whether cyber risk exposure drives economic outcomes of firms beyond stock prices and volatility. To this end, we run predictive regressions of firms' return on assets (RoA), cash flow / assets, and valuation on all three of our baseline measures. All dependent variables are standardized and with a one-quarter lead (t+1).

Table 6 reports the results of this exercise. Cyber risk exposure has negative and significant effects on future RoA, cash flow, and valuation. These associations are consistent with previous findings. Coefficients on $\text{CyberRisk}_{\text{i,t}}^{\text{R}}$ suggest that a one-standard deviation increase in exposure lowers future RoA by 2.5%, cash flow by 2.3%, and market value 0.6% of the variables' standard deviation. The top panel of Figure 6 shows (binned) scatter plots that relate $\text{CyberRisk}_{\text{i,t}}^{\text{R}}$ to these three financial variables. The Figure shows strong negative associations, especially in the case of RoA. Notice how results are not driven by singular outliers in any of the three plots. Panels (a)-(c) in Figure 7 further show dynamic effects for the three variables of interest. All coefficients are negative and significant for up to 8 quarters, spiking on impact and reverting to zero slowly. This suggests that the impact of cyber risk exposure can be highly persistent. One economic mechanism that can rationalize this finding is laid out in Akey et al. (2021) and centers around the role of corporate reputation. High exposure to cyber risk constitutes a negative shock to the firm's reputation, causing a long-lasting deterioration in reputational capital, profitability, and franchise value.

We can further quantify the effects on net income in terms of more easily interpretable dollar amounts. A one-standard deviation swing in RoA in our sample is roughly 4.51%. This translates into an RoA decline of the order of 0.11% (percentage points) for the average firm. The average firm in the sample possesses assets of about \$25,572M. This yields a loss of income for the average firm of \$27.79M or about \$28 million. To compute the loss of income for the aggregate economy we have to make some rough assumptions. The number of unique firms in our estimation sample (i.e. after merging StreetEvents with Compustat and performing all the data cleaning steps) for which the value of total assets is not missing is 2,023. Thus, for the aggregate "economy" the total loss in response to a one-standard deviation rise in cyber risk exposure is about \$56,664M per quarter or \$226,576M per year. This roughly estimated amount is in fact very close to more rigorous calculations of the global cost of cyber risk. For example, Bouveret (2018) estimates that the annual average loss to banks from cyber attacks amounts to US$100 billion. In a RAND Corporation Research Report, Dreyer et al. (2018) estimate the direct global cost of cyber crime of at least $275 billion per year. Our simple calculation doesn't capture the firms' precautionary investment motive that arises endogenously in response to the presence of background cyber risk (e.g. cyber insurance, operational analysts, cybersecurity software and services, etc.), so the true material cost of cyber uncertainty is indeed higher in practice. If one factors in both direct and systemic costs to global GDP, the cost of cyber crime can reach into trillions of U.S. dollars in some of the worst-case scenarios (Dreyer et al., 2018).

5.3 Firm-Level Option Market Effects

We now turn to a key empirical exercise of our study. To quantify the impact of firm-level cyber risk exposure and uncertainty we run regressions of our three measures of cyber risk exposure on the three main option market measures: implied volatility (IV), variance risk premium (VRP), and implied volatility slope (SlopeD). Variables are defined in Appendix A. Our main specification focuses on 91-day options with results on additional maturities available in the Appendix. All specifications include a firm and quarter fixed effects and the usual set of controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors are clustered at the firm level. All dependent variables have been standardized.

Table 7 reports the results. Columns (1)-(3), (4)-(6), and (7)-(9) show results for $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$, $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$, and $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ respectively. Across all 9 specifications, we see that cyber risk uncertainty has a positive and significant effect on costs of protection against general price, variance, and downside risks. Going from zero to positive exposure increases IV, VRP, and SlopeD by 3%, 1.5% and 1.6% of the variables' standard deviations, respectively. A one-standard deviation increase in relative exposure $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ raises the three option variables by 2.2%, 1.1% and 0.6% of their standard deviations. As mentioned in the Introduction, these effects are quantitatively in the same range as what Hassan et al. (2019) find in the case of political risk and Sautner et al. (2023) find in the case of climate-change risk. The bottom panel of Figure 6 shows (binned) scatter plots that relate $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ to the three option market variables. Positive associations are seen on all three plots. Outliers do not seem to drive the results. Results on SlopeD, as seen also from the point estimates above, are relatively less stark. Panels (d)-(f) of Figure 7 show dynamic effects for the three option market variables. Cyber risk exposure leads to persistently positive effects on IV, VRP, and SlopeD, lasting for up to 8 quarters. In the case of IV, the effect spikes on impact and slowly reverts to zero over time. For VRP and SlopeD, however, effects seem to remain high and not to mean-revert. This suggests that cyber risk could cause long-lasting if not permanent damages to variance and downside risks of exposed firms.

We now extend our firm-level analysis to the 8 topical indices. We run the same quarterly firm-level regressions with firm controls and time and quarter fixed effects with the only difference being that the main regressor is now CyberRisk $\text{Topic}^{\text{I}}_{\text{i,t}}$, a topical indicator variable. Table 8 reports the findings. Panels A, B, and C show the results for IV, VRP, and SlopeD, respectively. In each column CyberRisk $\text{Topic}^{\text{I}}_{\text{i,t}}$ takes on the value of the topical measure specified in the column. From Panel A, we find that all topical measures have positive and significant effects on IV. From Panel B, we see that all topical measures except Crypto, Sentiment, and Politics have positive and significant effects on the VRP. Finally, from Panel C we see that all topics except Crypto and Social media have positive and significant effects on the SlopeD. The magnitudes of the effects are generally in the same ballpark as for the baseline $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$. A notable exception is the Uncertainty topic which exhibits notably larger effects (8.88%, 10.8%, and 5.2% of variables standard deviations, respectively).

All in all, we find that at the firm level, cyber risk exposure and uncertainty is priced. Obtaining protection against price, variance, and tail risks comes at a premium, which increases for firms that face higher exposure. This finding is consistent with a theory that links cyber risk exposure at present times with probabilities of future realized attacks and related monetary or reputational damage through forward-looking option market variables. Results hold for the baseline exposure measure as well as specific topical measures such as Insurance, Law, and Uncertainty.

5.4 Industry-Level Effects

One central research question for us is whether firm-level cyber risk exposure and uncertainty wash out in the aggregate or instead have industry-level effects. We now aggregate all key variables, including the three option-market measures, the RoA, and the usual controls to various levels of sectoral aggregation: the 3 and 4 digit NAICS classifications.[30] Our main regressor of interest in this exercise is now $\text{CyberRisk}^{\text{R}}_{\text{s,t}}$ (where subscript s stands for sector), which is the average of firm-level $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$. We construct either equally-weighted or size-weighted averages, where in the latter case size is proxied by book total assets. All dependent and independent variables have been standardized as before. All specifications include industry and interacted country x time fixed effects.

Table 9 reports the results. Even at the industry level, $\text{CyberRisk}^{\text{R}}_{\text{s,t}}$ has a robustly positive and statistically significant effect on IV, VRP, and SlopeD. This outcome holds

[30]Our sample includes 84 unique 3-digit and 232 4-digit NAICS industries.

for both 3- (Panel A) and 4-digit (Panel B) NAICS industries, as well as for both equally and asset-weighted aggregation approaches. For example, a one-standard deviation spike in $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{s,t}}$ raises IV, VRP, and SlopeD by 2.4%, 2.6%, and 1.5% by the variables' standard deviations, respectively, as per columns (1)-(3). Furthermore, we also document a negative association with future returns on assets. A one st. dev. increase in $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{s,t}}$ lowers future sectoral RoA by 1.8%-2.6% of the variable's standard deviation across all specifications. Overall, our findings in this section allow us to make claims on the importance of cyber risk beyond firm-level direct effects. Cyber risk has industry-level aggregate implications.

5.5 Spillover Effects

Can idiosyncratic, firm-level cyber risk exposure and uncertainty spill over across markets and generate systemic ripple effects?[31] To further test the notion that cyber risk can be a source of systemic risk for firms we now perform the following exercise. We estimate the impact on firm-level outcomes of cyber risk exposure $\mathrm{CyberRisk}^{\mathrm{R}}_{\mathrm{i,t}}$ that has been aggregated to the country x industry x quarter level. Industries are defined by the 4-digit NAICS codes. We partition the full sample of firms into the affected and the peers. Affected firms are those with positive firm-level exposure. Note how this definition restricts the full impact of our measure to the intensive margin only. Peers, on the other hand, are defined as companies which are headquartered in the same country and operate in the same industry as the affected firm but have zero firm-level exposure of their own. Thus, this empirical strategy attempts to trace out indirect, spillover effects of cyber risk exposure on firms that are not impacted directly but are "connected" to the exposed firms because they belong to the same tightly defined industry, and could thus be affected by association. In other words, we conjecture that financial markets begin to perceive certain firms as being operationally risky if new information about cyber risk exposure of their *peers* gets revealed to the public. All specifications include the usual set of firm controls and interacted industry x time fixed effects.

Table 10 reports the results. Columns (1)-(4) show results for the sample of all firms (affected and peers), columns (5)-(8) zoom in on affected firms (those with positive exposure), and columns (9)-(12) focus on the peers (those with zero exposure). Two main observations come out from this exercise. First, direct effects are positive and statistically significant. This is consistent with our firm-level results and does not come as a surprise. Second, cyber risk exposure has a significant effect on profitability and option market variables of peer firms.

[31] Crosignani et al. (2022) document that cyberattack-driven disruptions propagate across supply chains. Eisenbach et al. (2022) reach a similar conclusion but in the context of the U.S. wholesale payments network. Our focus is on the propagation through financial markets, specifically the option market.

This is evidence of *spillover* effects as peer firms have by construction 0 contemporaneous cyber risk exposure but yet suffer elevated costs of protection against price, variance, and downside market risks as well as lower returns. An important caveat is that this is firm-level, idiosyncratic exposure that causes spillovers. Correlated exposures, i.e. those that affect multiple firms at once (e.g. global cyber attack, state-sponsored hacking operation), could have much stronger systemic implications.

To complement the above exercise that uncovers average spillover effects, we also consider whether spillovers operate heterogeneously based on some firm characteristic such as size. Specifically, in each country, industry, and quarter, we construct percentiles of the distribution of the preferred proxy of firm size - market value. Then, we re-run spillover repressions on the sub-sample of peers that are larger than the respective percentile. Figure D.3 presents the outcome. Each of the four panels show results for a dependent variable of interest. On the x-axes are always percentiles of the market value distribution, ranging from 1 to 75. On the vertical axes we show standardized point estimates with the 90% confidence intervals. We uncover that spillover option market (IV or slopeD) effects are not concentrated in any particular corner of the distribution of firm size and are instead fairly homogeneous across the economy. In the case of the variance risk premium and the ROA, we see that larger firms tend to be marginally more affected. It may be because larger firms are more in the spotlight and hence are more susceptible to contagion effects but this is just a conjecture. Although the differentials in estimates for the 1st and the last percentiles of firm size are large, they are not statistically significantly different from each other.

5.6 Cyber Risk and Cryptocurrencies

By eyeballing Figure 3 one can speculate that the CyberRisk $\mathrm{Crypto_t}$ topical index seems to peak around local maxima in cryptocurrency valuations. The link between crypto coins and ransomware risk has been noted by commentators. In this section we provide tentative statistical correlations between the price of Bitcoin (which dominates the total crypto coin market cap with a 42% share as of July 2022) and some of our topical measures. We obtain the price of Bitcoin from Coinmarketcap.com, which is a leading source of cryptocurrency price and volume data. We aggregate the price to the quarterly frequency by averaging. Panel (a) of Figure 8 plots the resulting standardized series together with our CyberRisk $\mathrm{Crypto_t^A}$ measure. Correlation coefficient between the two series is 95% and statistically significant at the 1% level. Panel (b) plots the two series in first differences, with the correlation coefficient of 57% (significant at the 1% level). There appears to be a very strong contemporaneous link between earnings calls discussions that simultaneously cover cyber risk and cyptocurrencies and the market value of Bitcoin.

Studies such as Wang and Vergne (2017) and Liu and Tsyvinski (2021) find that investor attention, as proxied by Google searches or newspaper headlines, forecasts future cryptocurrency performance. We conduct a simple statistical test in the spirit of these studies under the assumption that our topical measure CyberRisk $\text{Crypto}_{\text{t}}^{\text{A}}$ approximates analyst interest in crypto related affairs. At the quarterly frequency, we regress future price of Bitcoin (at horizons of one to four periods) on current CyberRisk $\text{Crypto}_{\text{t}}^{\text{A}}$. Panel (A) is in levels; Panel (B) is in first differences. All variables are standardized. Table 11 reports the results in columns (1)-(4) across the two panels. We see that our CyberRisk $\text{Crypto}_{\text{t}}^{\text{A}}$ topical measure is strongly associated with the *level* of future Bitcoin price appreciations. This finding is consistent with a theory that links cyber risk exposure to elevated crypto-related analyst attention and market valuations. For example, analysts and firm managers may internalize that crypto coins are typically the currencies of ransomware attacks. Greater risk of potential future attacks raises interest and attention towards the topic of cryptocurrencies, and the market prices in future potential demand for crypto transactions through appreciations. In Panel (B), the association in first differences is not present, however.

It is also possible that cyber criminals intensify their activities *in response* to appreciations of notable coins, to get higher dollar returns from their bitcoin denominated attacks. It could also be that cyber criminals have more resources when bitcoins prices are high and they scale up their activities. This could cause analysts to conduct more crypto-centered conversations in a reactive rather than proactive fashion. We can test the extent to which past cryptocurrency prices influence current analyst attention to crypto and cyber risk. We regress current levels of CyberRisk $\text{Crypto}_{\text{t}}^{\text{A}}$ on current and past levels and differences in the price of Bitcoin. Results are summarized in columns (5)-(9), panels (A) and (B), in Table 11. For both levels and differences, we see that past Bitcoin prices *are* positively and significantly associated with current levels of CyberRisk $\text{Crypto}_{\text{t}}^{\text{A}}$. This positive correlation, however, does not persist past two quarters. These findings imply that the "reactive theory" has some empirical support, as does potentially the "proactive theory". Identifying the direction of causality is beyond the scope of our paper and would require a serious quasi-experimental setting. However, we believe that future research can benefit from these insights and conduct more comprehensive analysis on this topic.

6 Robustness Checks

In this section we provide many robustness checks on our main findings.

First, we perform an alternative dictionary validation procedure by running predictive regressions recursively. Second, we run a test of asymmetric effects by utilizing terms that reduce the likelihood of future cyberattacks. Third, we ask whether option market effects are driven more by firm-level or aggregate cyber risk. Fourth, we replicate our main regressions on options of different maturities to confirm that our results are not driven only by 91-day options. Fifth, we re-run our main specifications on a restricted time period of 2005q1-2021q3 to account for any potential data issues in the first few years of our sample. Finally, we run placebo exercises where we randomly re-assign the main regressor within firms and across time.

Recursive Dictionary Validation Our baseline dictionary validation procedure from Section 2.3 runs predictive logit regressions on the full sample in one step. This approach is potentially restrictive in the sense of requiring $\tilde{\mathbb{C}}$ to be time-invariant. With 20+ years of quarterly data, this is an assumption that demands an independent robustness check. An alternative approach would be to run the same validation analysis recursively, i.e. utilizing only data that was available at the time.[32] Specifically, we now run the same predictive logit model 15 times, once per each year, over the 2005-2019 period for which PRC cyberattack data is available. We then discard, year by year, terms with an odds ratio of less than or equal to one such that the set of validated terms is allowed to be time-varying. Finally, we construct a new measure $\bar{\mathrm{CyberRisk}}_{\mathrm{i,t}}$ and re-run our main firm-level analysis. Table D.3 reports the results of firm-level economic and option market effects conditional on the new, recursively validated cyber risk exposure measure. We find that all estimates remain the same.

Asymmetric Effects Our baseline cyber risk exposure measure is comprised of terms which are useful for predicting future cyberattacks. Our procedure discards some 63 cyber terms that are associated with a *reduction* in the probability of future attacks. One can potentially utilize this set and ask whether cyber risk is *symmetric*, i.e. whether an index that is built on those 63 terms has any reversed relationship with economic and financial aggregates of interest. To this end, based on these 63 terms we have constructed new cyber risk exposure measures $\tilde{\mathrm{CyberRisk}}_{\mathrm{i,t}}$ and have re-done our analysis. Table D.4 reports the results of firm-level economic and option market effects. From Panel (b) we find negative and mostly statistically significant coefficients. This suggests that certain cyber risk terms have a calming effect on financial markets; in this sense cyber risk is priced into the option market symmetrically. On the other hand, from Panel (A) we find zero effects on balance sheet variables such RoA or cash flow. We also found that neither our sector-level or spillover analysis

[32]We thank Christodoulos Louca, our discussant, for suggesting this idea.

produced any economically or statistically significant results (not shown). We conclude that the upside from cyber risk-related discussions in the earnings calls is generally limited to option markets with no observed pass-through to balance sheets, no propagation or spillovers, and no sectoral or aggregate effects. Thus, cyber risk exposure can be thought of as an *asymmetric* source of risk, with limited upside and considerable downside implications.

Firm-Level or Aggregate Cyber Risk Do option market effects that we uncover in this paper run through *firm-level* or aggregate cyber risk channels? In other words, what fraction of firm-level effects is driven by the time-series dimension? To answer this question, we aggregate $\text{CyberRisk}^{\text{R}}_{\text{i,t,}}$ by averaging to the quarterly level and include it in our baseline firm-level regression of Sector 5.3. Table D.5 in the Online Appendix reports the results. In columns (1)-(2), (3)-(4), and (5)-(6) the dependent variable is IV, VRP, and SlopeD, respectively. We include the Mean $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (std.) in columns (2), (4), and (6) and remove the time fixed effect. All specifications include all the controls and firm fixed effects. Inclusion of the mean of $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ lowers the coefficients by 15%, 12.5%, and 60%, respectively for the three option market variables. Coefficients on $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ remain significant at the 1% level for the cases of IV and VRP but significance drops to 10% for SlopeD. Coefficients on average $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ itself imply that a one-standard deviation rise in the time-series (a value which is smaller than in the panel by an order of magnitude) raises IV, VRP, and SlopeD by 7.5%, 3.7%, and 19% of their standard deviations, respectively. The time-series dimension therefore also matters.

Different Option Maturities Are our baseline results robust to different option maturities? Table D.6 reports estimates from firm-level regressions for 30-, 60-, and 182-day options. Results are presented for $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$, $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$, and $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$, in line with baseline estimates in Table 7. We see that our results dot not change and we obtain 23 statistically significant coefficients out of 27.

Restricted Sample Figure 2 (left panel), which plots the absolute and relative frequencies of our aggregated measures, shows that the first few years of our sample exhibit a peculiar decoupling between the two series. This occurs because the denominator in $\text{CyberRisk}^{\text{R}}$, i.e. the total number of words in earnings call transcripts, increases by roughly two standard deviations over 2002q2-2005q1 and then stabilizes (not shown). One concern is that this feature of the data affects our results. We therefore conduct a robustness check where we re-run our main specifications on a restricted sample of 2005q1-2021q3. Table D.7 reports main results from our firm-level analysis of economic and option market effects. None of the estimates change.

Placebo Tests Our final robustness exercise involves running a falsification exercise: placebo regressions for our firm-level specification in Section 5.3. Specifically, we regress our key firm-level variables on $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ where the time series of $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ of every firm has been randomly assigned with replacement. Figure D.4 displays histograms of the t-statistics from 500 regressions. In all six panels, distributions are centered around 0 and are symmetrical. The fraction of false-positive and false-negative cases (defined as the two-sided 95% confidence band) is 2.4%, 2.6%, 2.4%, 2.8%, 1.6%, and 2.6% for the six panels, respectively. We conclude that achieving our baseline results by pure chance would have been highly unlikely.

7 Conclusion

Automation, disruptive technologies like cloud services, the growth of DeFi, the work-from-home revolution are all factors that are rapidly increasing the likelihood of idiosyncratic and global cyberattacks. Uncertainty surrounding exposure to potential future attacks is hard to quantify, primarily due to measurement issues. Reliance on reported cyberattacks is an imperfect solution for all the reasons the literature already documents. Alternative approaches to measuring cyber risk are required.

In this paper, we provide one such alternative by leveraging tools from natural language processing and quarterly earnings calls of listed firms to build a text-based measure of cyber risk exposure. Our measure builds on term libraries of three reputable institutions and validates them with realized cyberattacks. We supplement our core exposure measures with 8 topical indices that capture various contexts in cyber risk discussions. We provide extensive evidence that our measures are valid and truly reflect economically meaningful firm-level variation in cyber risk: we provide case studies of cyberattacked and cybersecurity firms, present snippets from actual call transcripts of select firms, show that our measures can predict reported cyberattacks 1, 4, and 8 quarters in the future, demonstrate that our measures are strongly associated with stock market outcomes and realized volatility, and validate our measures against 10-K files. We are able to provide simple back-of-the-envelope calculations for the aggregate cost of cyber risk exposure which amounts to $226 billion in net income lost per year. This is a lower bound on the cost magnitude as multiple indirect, precautionary, and systemic costs are not accounted for in this calculation.

Unlike most of the existing literature, we are able to provide a global description of cyber risk exposure since our data contains firms from 85 countries and to document shifting geographical patterns. We also present the dynamics of cyber exposure across sectors and characterise the firms which are more likely to be cyberattacked. Using our measures, we show that cyber risk uncertainty is priced in the option market. To the best of our knowledge, we are the first to report this result. Market-based costs of protection against price, variance, and downside risks are greater for firms with higher cyber risk exposure. It is known that option market variables are forward-looking and can be used to predict future stock market and real economic performance. Thus, cyber risk exposure at present times signals future potential stock market or real economic deterioration.

We move beyond firm-level analysis and find that idiosyncratic cyber risk can potentially have systemic implications. Firm-level exposure does not wash out in the aggregate and has significant sector-level effects. Moreover, option market effects spill over across firms; affected firms have a negative effect on their peers, defined as firms in the same country and industry as the affected firm. Financial markets can thus propagate firm-level cyber risk exposure, amplify singular incidents, and have "systemic risk" type implications.

We hope that our results open several avenues for future research. First, all our exposure measures will be made publically available. Our data could be used to establish novel causal effects of cyber risk on employment or other real economic aggregates. Our topical measures - Insurance, Cryptocurrencies, Social Media, and Law - could be useful for various analyses of the links between cyber risk exposure and the cryptocurrency world. Finally, our measures can help calibrate a new generation of equilibrium models that aim to quantify the welfare cost of cyber risk based on empirical firm-level variation.

A Appendix

Table 1: Key Variable Definitions

Variable	Definition	Source
$\text{CyberRisk}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts. $\text{CyberRisk}^{A}_{i,t}$: absolute frequency, i.e. the total number of such bigrams in the transcript of firm i in quarter t. $\text{CyberRisk}^{R}_{i,t}$: relative frequency, normalized by the total number of bigrams in transcripts. $\text{CyberRisk}^{I}_{i,t}$: indicator variable which takes the value of 1 if $\text{CyberRisk}^{A}_{i,t}$ is positive and 0 otherwise.	Thomson Reuters StreetEvents. Self-constructed.
CyberRisk $\text{Insurance}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 50 word distance from terms in the Insurance topic, summarized in Table A.2. Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Thomson Reuters StreetEvents. Self-constructed.
CyberRisk $\text{Legal}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 50 word distance from terms in the Law topic, summarized in Table A.2. Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Thomson Reuters StreetEvents. Self-constructed.
CyberRisk $\text{Crypto}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 50 word distance from terms in the Cryptocurrencies topic, summarized in Table A.2. Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Thomson Reuters StreetEvents. Self-constructed.
CyberRisk $\text{SocialMedia}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 50 word distance from terms in the Social Media topic, summarized in Table A.2. Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Thomson Reuters StreetEvents. Self-constructed.
CyberRisk $\text{Uncertainty}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 10 word distance from terms in the Risk and Uncertainty topic, summarized in Hassan et al. (2019). Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Hassan et al. (2019)
CyberRisk $\text{PosSentiment}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 10 word distance from words with the Positive Sentiment tone, summarized in Loughran and McDonald (2011). Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Loughran and McDonald (2011), Hassan et al. (2019)
CyberRisk $\text{NegSentiment}_{i,t}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 10 word distance from words with the Negative Sentiment tone, summarized in Loughran and McDonald (2011). Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Loughran and McDonald (2011), Hassan et al. (2019)
CyberRisk $\text{NetSentiment}_{i,t}$	Difference between CyberRisk $\text{PosSentiment}_{i,t}$ and CyberRisk $\text{NegSentiment}_{i,t}$. Absolute frequency, relative frequency, and indicator variants are built the same way as in $\text{CyberRisk}_{i,t}$.	Loughran and McDonald (2011), Hassan et al. (2019). Self-constructed.

Variable	Definition	Source
CyberRisk $\mathrm{Politics}_{\mathrm{i,t}}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 10 word distance from terms in the Political topic, summarized in Hassan et al. (2019). Absolute frequency, relative frequency, and indicator variants are built the same way as in $\mathrm{CyberRisk}_{\mathrm{i,t}}$.	Hassan et al. (2019)
CyberRisk $\mathrm{Disease}_{\mathrm{i,t}}$	Frequency with which validated bigrams related to cybersecurity occur in quarterly earnings call transcripts within a 10 word distance from terms in the Disease topic, summarized in Hassan et al. (2023). Absolute frequency, relative frequency, and indicator variants are built the same way as in $\mathrm{CyberRisk}_{\mathrm{i,t}}$.	Hassan et al. (2023)
Cyberattack Indicator	Dummy variable which takes the value of 1 if a firm reported a cyberattack in the present quarter, and 0 otherwise.	PRC
IV	Implied volatility of (log) returns computed from 91-day options. Quarterly measure is constructed by averaging daily values. Similar measures using 30-, 60-, and 182-day maturity options are constructed. Winsorized at the 1% level.	Ivy DB OptionMetrics Volatility Surface File.
VRP	Variance risk premium, defined as the daily difference between the implied variance of (log) returns (IV^2) from t to t+91 calendar days and realized variance of daily (log) returns over the same period (t, t+91). Quarterly measure is constructed by averaging daily values. Similar measures using 30-, 60-, and 182-day maturity options are constructed. Winsorized at the 1% level.	Ivy DB OptionMetrics Volatility Surface File.
SlopeD	Slope of the function that relates implied volatility to the Black-Scholes delta for OTM put options with a 91-day maturity. We run a daily, within-quarter regression of implied volatilities (IV) of OTM put options (deltas between -0.5 and -0.1) on a constant and corresponding deltas. The slope coefficient of the regression constitutes the firm $\times$ quarter SlopeD measure. Similar measures using 30-, 60-, and 182-day maturity options are constructed. Winsorized at the 1% level.	Ivy DB OptionMetrics Volatility Surface File.
WRet	Weighted average quarterly returns, computed as means (weighted by valuation) of daily (log) returns in CRSP. Winsorized at the 1% level.	CRSP
CRet	Cumulative returns, computed as quarterly sums of (log) returns in CRSP. Winsorized at the 1% level.	CRSP
RV	Realized volatility of (log) returns over the period of t and t+91 calendar days in CRSP. Winsorized at the 1% level.	CRSP
Assets	Total assets at the end of the quarter (in logs). ATQ variable in Compustat. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Age	Firm age (in logs) in Compustat. Self-constructed. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Tobin's Q	(Total assets (ATQ) - total common equity (CEQ) + share price (PRCCQ) $\times$ common shares outstanding (CSHOQ)) / total assets (ATQ). We drop observations with PRCCQ<=1 (penny stocks) and >1000. We drop observations with Tobin's Q >1000. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Leverage	(Long term debt (DLTTQ) + debt in current liabilities (DLCQ)) / total assets. We drop observations with Leverage >1. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Liquidity	Cash and short-term investments (CHEQ) / total assets (ATQ). Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Intangibles / Assets	Intangible assets (INTANQ) / total assets (ATQ). We drop observations with Intangibles / Assets of >1. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Operational Costs / Assets	Operating expense (XOPRQ) / total assets (ATQ). Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly

Variable	Definition	Source
Market Beta	Sensitivity of quarterly stock returns to quarterly S&P returns. For each firm and quarter, we run daily regressions of excess (log) returns on a constant and the market factor. For each firm x quarter combination, Market Beta corresponds to the estimated regression coefficient. Winsorized at the 1% level.	CRSP, Kenneth French's website.
RoA	Net income (NIQ) / total assets. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Cash Flow / Assets	(Income before extraordinary items (IBQ) + depreciation and amortization (DPQ)) / total assets (ATQ). We drop observations with Cash Flow / Assets of >1 or <-1. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Market Value	Market value (in logs). MKVALTQ in Compustat. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
S&P Rating	S&P quality ranking (SPCSRC variable in Compustat).	Compustat Global - Fundamentals Quarterly
CAPEX / Assets	Invested capital (ICAPTQ) / total assets. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Book to Market Ratio	Total common equity / (share price × common shares outstanding). Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
PP&E / Assets	Property plant and equipment (PPENTQ) / total assets. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Debt Maturity Ratio	Long-term debt / (long-term debt + debt in current liabilities). Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Equity Issuance Ratio	Common shares issued (CSHIQ) / total assets. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly
Turnover Ratio	Sales (SALEQ) / total assets. We drop observations with SALEQ<0. Winsorized at the 1% level.	Compustat Global - Fundamentals Quarterly

References

Adeney, R., J. Healey, P. Mosser, and D. M. Waiss, "Cyber Risk and Financial Stability - An Atlas for Macroprudential Analysis," *Working Paper*, 2022.

Adrian, T. and C. Ferreira, "Mounting Cyber Threats Mean Financial Firms Urgently Need Better Safeguards," *IMF Blog*, 2023.

Akey, Pat, S. Lewellen, I. Liskovich, and C. Schiller, "Hacking Corporate Reputations," *SSRN Working Paper*, 2021, *3143740.*

Aldasoro, I., L. Gambacorta, P. Giudici, and T. Leach, "The Drivers of Cyber Risk," *Journal of Financial Stability*, 2022, *60.*

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

Anand, K., C. Duley, and P. Gai, "Cybersecurity and Financial Stability," *Deutsche Bundesbank Discussion Paper*, 2022, *08.*

Anhert, T., M. Brolley, D. Cimon, and R. Riordan, "Cyber security and ransomware in financial markets," *CEPR Discussion Paper 17403*, 2022.

Baker, S., N. Bloom, S. Davis, and A. Notes, "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 2016, *131(4).*

Bali, T. and H. Zhou, "Risk, uncertainty, and expected returns," *Journal of Financial and Quantitative Analysis*, 2016, (51 (3)).

Beber, A. and M. Brandt, "The effect of macroeconomic news on beliefs and preferences: Evidence from the options market," *Journal of Monetary Economics*, 2006, *53.*

Biener, C., M. Eing, and J. H. Wirfs, "Insurability of Cyber Risk: An Empirical Analysis," *The Geneva Papers on Risk and Insurance - Issues and Practice*, 2015, *40.*

BoE, "Bank of England Systemic Risk Survey," 2020.

Bollerslev, Tim, G. Tauchen, and H. Zhou, "Expected stock returns and variance risk premia," *Review of Financial Studies*, 2009, *22.*

Bouveret, A., "Cyber risk for the financial sector: A framework for quantitative assessment," *IMF Working Paper 18/143*, 2018.

Carr, P. and L. Wu, "Variance risk premiums," *Review of Financial Studies*, 2009, *22.*

Chang, B.Y., P. Christoffersen, and K. Jacobs, "Market skewness risk and the cross section of stock returns," *Journal of Financial Economics*, 2013, *107.*

Crosignani, Matteo, M. Macchiavelli, and A. F. Silva, "Pirates without Borders: The Propagation of Cyberattacks through Firms' Supply Chains," *Journal of Financial Economics*, 2022, *Forthcoming.*

CSIS, "Net Losses: Estimating the Global Cost of Cybercrime," *Report*, 2014.

Dreyer, Paul, T. M. Jones, K. Klima, J. Oberholtzer, A. Strong and. J. Welburn, and Z. Winkelman, "Estimating the Global Cost of Cyber Risk: Methodology and Examples," *RAND Corporation Research Report*, 2018.

Duffie, D. and J. Younger, "Cyber Runs," *Hutchins Center Working Paper*, 2019, *51.*

Eisenbach, T., A. Kovner, and M. J. Lee, "Cyber Risk and the U.S. Financial System: A Pre-Mortem Analysis," *Journal of Financial Economics*, 2022, *145(3).*

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel, "Hedging climate change news," *Review of Financial Studies*, 2020, *33(3).*

ESRB, "Systemic Cyber Risk," *Report*, 2020, *February.*

Floracki, Chris, C. Louca, R. Michaely, and M. Weber, "Cybersecurity Risk," *Review of Financial Studies*, 2022, *Forthcoming.*

Gentzkow, Matthew, B. T. Kelly, and M. Taddy, "Text as Data," *Journal of Economic Literature*, 2019, *57(3).*

Hassan, T., J. Schreger, M. Schwedeler, and A. Tahoun, "Sources and Transmission of Country Risk," *Review of Economic Studies*, 2022, *Forthcoming.*

_ , S. Hollander, L. v. Lent, and A. Tahoun, "Firm-Level Political Risk: Measurement and Effects," *Quarterly Journal of Economics*, 2019, *134(4).*

_ , _ , _ , and _ , "The Global Impact of Brexit Uncertainty," *Journal of Finance*, 2022, *Forthcoming.*

_ , _ , _ , and _ , "Firm-Level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1," *Review of Financial Studies*, 2023, *Forthcoming.*

Healey, J., P. Mosser, K. Rosen, and A. Wortman, "The Ties That Bind: A Framework To Assess The Linkage Between Cyber Risks And Financial Stability," *Journal of Financial Transformation*, 2021, *53.*

Hentschel, L., "Errors in implied volatility estimation," *Journal of Financial and Quantitative Analysis*, 2003, *38.*

Hilscher, J., A. Raviv, and R. Reis, "Inflating Away the Public Debt? An Empirical Assessment," *Review of Financial Studies*, 2022, *35(3).*

Hollander, S., M. Pronk, and E. Roelofsen, "Does silence speak? An empirical analysis of disclosure choices during conference calls," *Journal of Accounting Research*, 2010, *48(3).*

Huang, A., R. Lehavy, A. Zang, and R. Zheng, "Analyst information discovery and interpretation roles: A topic modeling approach," *Management Science*, 2018, (64 (6)).

Ilhan, Emirhan, Z. Sautner, and G. Vilkov, "Carbon Tail Risk," *Review of Financial Studies*, 2020, *34(3).*

Jiang, H., N. Khanna, and Q. Yang, "The Cyber Risk Premium," *SSRN Working Paper 3637142*, 2020.

Kamiya, S., J. Kang, J. Kim, A. Milidonis, and R. Stulz, "Risk Management, Firm Reputation, and the Impact of Successful Cyberattacks on Target Firms," *Journal of Financial Economics*, 2021, *139(3).*

Kashyap, A. and A. Wetherilt, "Some Principles for Regulating Cyber Risk," *AEA Papers and Proceedings*, 2019, *109*, 482–487.

Kelly, Bryan, L. Pastor, and P. Veronesi, "The Price of Political Uncertainty: Theory and Evidence from the Option Market," *Journal of Finance*, 2016, *71(5).*

Koijen, R., T. Philipson, and H. Uhlig, "Financial health economics," *Econometrica*, 2016, *84(1).*

Kotidis, Antonis and S. L. Schreft, "Cyberattacks and Financial Stability: Evidence from a Natural Experiment," *Finance and Economics Discussion Series*, 2022, *025.*

Lallie, Harjinder Singh, Lynsay A. Shepherd, Jason R.C. Nurse, Arnau Erola, Gregory Epiphaniou, Carsten Maple, and Xavier Bellekens, "Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," *Computers Security*, 2021, *105.*

Lhuissier, S. and F. Tripier, "Measuring Cyber Risk," *Working Paper*, 2021.

Liu, Yukun and A. Tsyvinski, "Risks and Returns of Cryptocurrency," *Review of Financial Studies*, 2021, *34.*

Loughran, T. and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *Journal of Finance*, 2011, *66 (1).*

_ and _ , "Textual analysis in accounting and finance: A survey," *Journal of Accounting Research*, 2016, *54(14).*

Makridis, C. and B. Dean, "Measuring the Economic Effects of Data Breaches on Firm Outcomes: Challenges and Opportunities," *Journal of Economic and Social Measurement*, 2018, *43.*

Neuhierl, A. and M. Weber, "Monetary policy communication, policy slope, and the stock market," *Journal of Monetary Economics*, 2020.

OPC, "Security deficiencies at BMO lead to large-scale breach," *Office of the Privacy Commissioner of Canada*, 2021, *https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-003/.*

ORX, "2020 Annual Banking and Insurance Operational Loss Reports," 2020.

Pastor, Lubos and P. Veronesi, "Political uncertainty and risk premia," *Journal of Financial Economics*, 2013, *110(3).*

Patell, J. and M. Wolfson, "Anticipated information releases reflected in call option prices," *Journal of Accounting and Economics*, 1979, *1.*

Portes, R. and H. Rey, "The determinants of cross-border equity flows," *Journal of International Economics*, 2005, *65(2).*

_ , _ , and Y. Oh, "Information and capital flows: The determinants of transactions in financial assets," *European Economic Review*, 2001, *45.*

Salton, G. and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, 1988, *24(5).*

Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang, "Firm-level climate change exposure," *Journal of Finance*, 2023, *78(3).*

Tosun, O. K., "Cyber Attacks and Stock Market Activity," *International Review of Financial Analysis*, 2021, *76.*

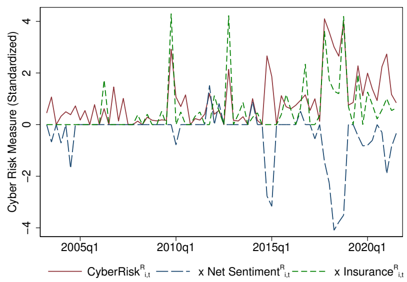
Vanden, J. M., "Information quality and options," *Review of Financial Studies*, 2008, *21.*

Wang, S. and J. Vergne, "Buzz Factor or Innovation Potential: What Explains Cryptocurrencies' Returns?," *PLoS ONE*, 2017, (12 (1)).

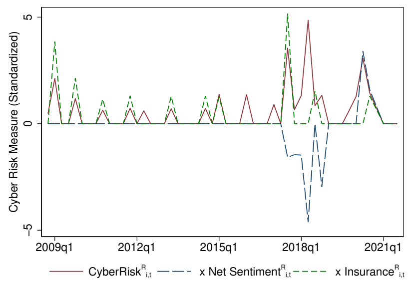
WEF, "Understanding Systemic Cyber Risk," *World Economic Forum: Global Agenda Council on Risk and Resilience*, 2016.

Woods, D., T. Moore, and A. Simpson, "The County Fair Cyber Loss Distribution: Drawing Inferences from Insurance Prices," *Working Paper*, 2019.

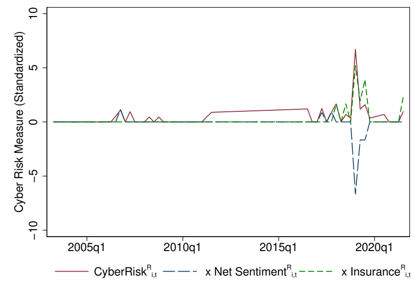
Figure 1: Case Studies - Select Cyberattacked Firms



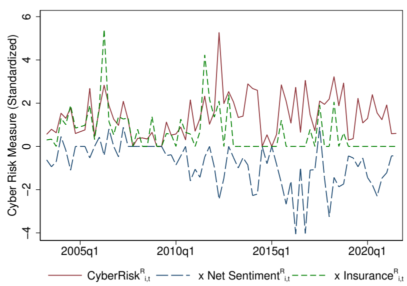
(a) Equifax



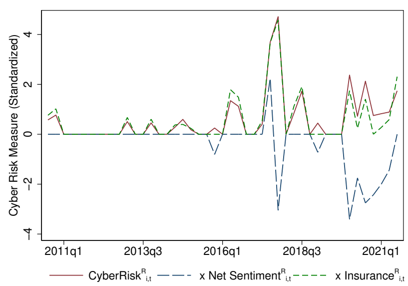
(b) Bank of Montreal



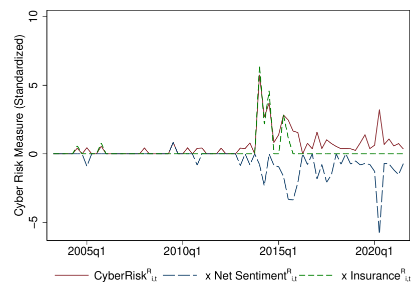
(c) Marriott



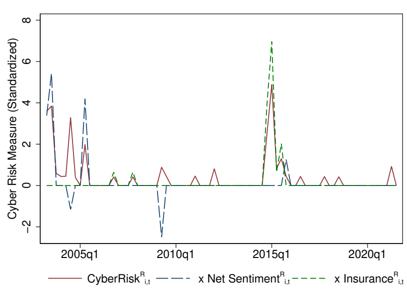
(d) Adobe



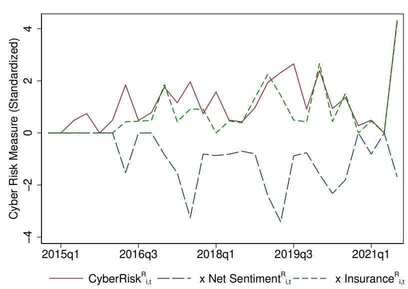
(e) First American Financial



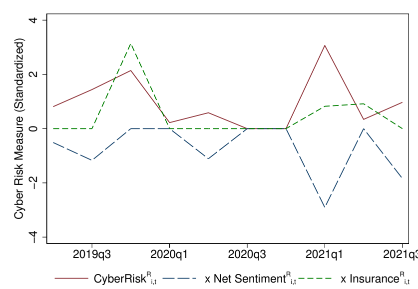
(f) Target



(g) Home Depot



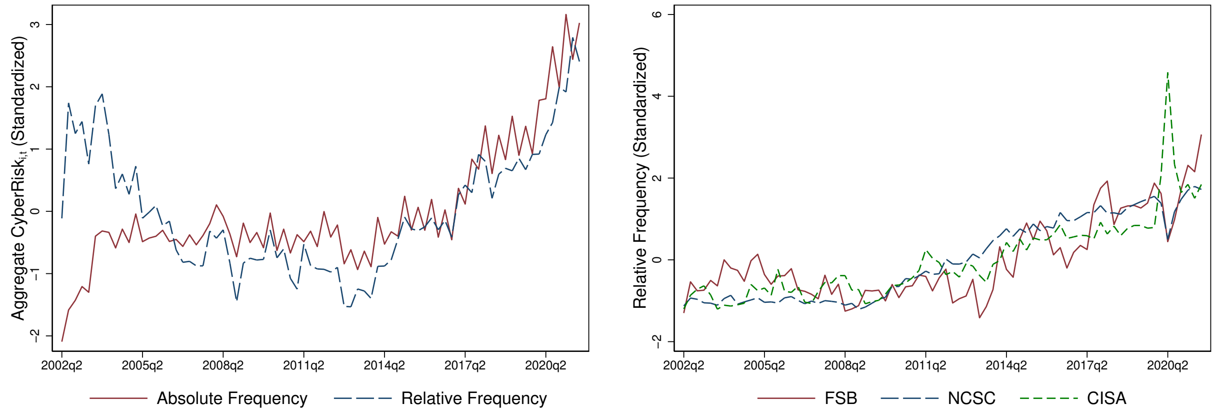
(h) Alibaba



(i) SolarWinds

Notes: This figure plots the time series of our cyber risk measures for select cyberattacked firms.

Figure 2: Cyber Risk over Time

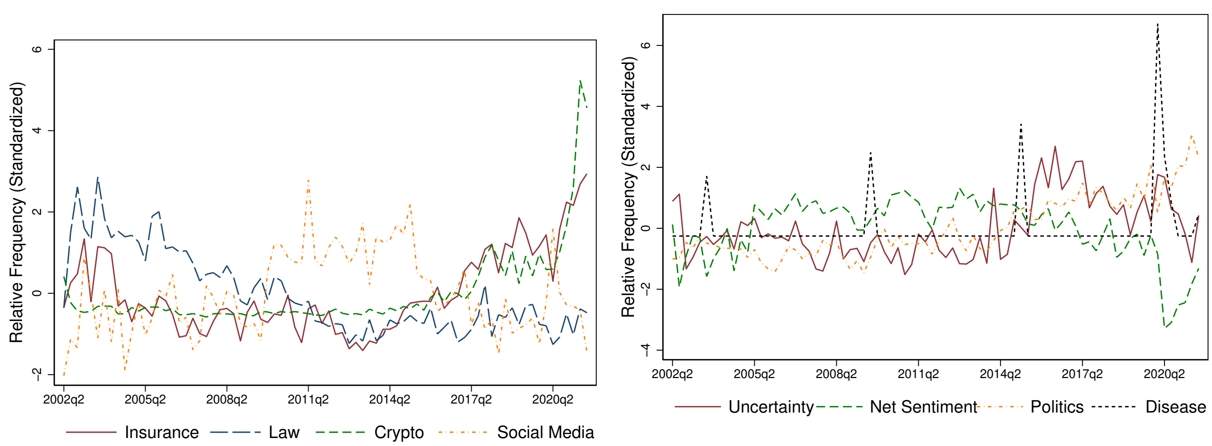


(a) Main Validated Index

(b) Source Libraries

Notes: This figure plots our baseline indices CyberRisk$^{\mathrm{A}}$ and CyberRisk$^{\mathrm{R}}$ on the left panel and unvalidated raw indices from our three source libraries on the right panel.

Figure 3: Cyber Risk by Topic over Time

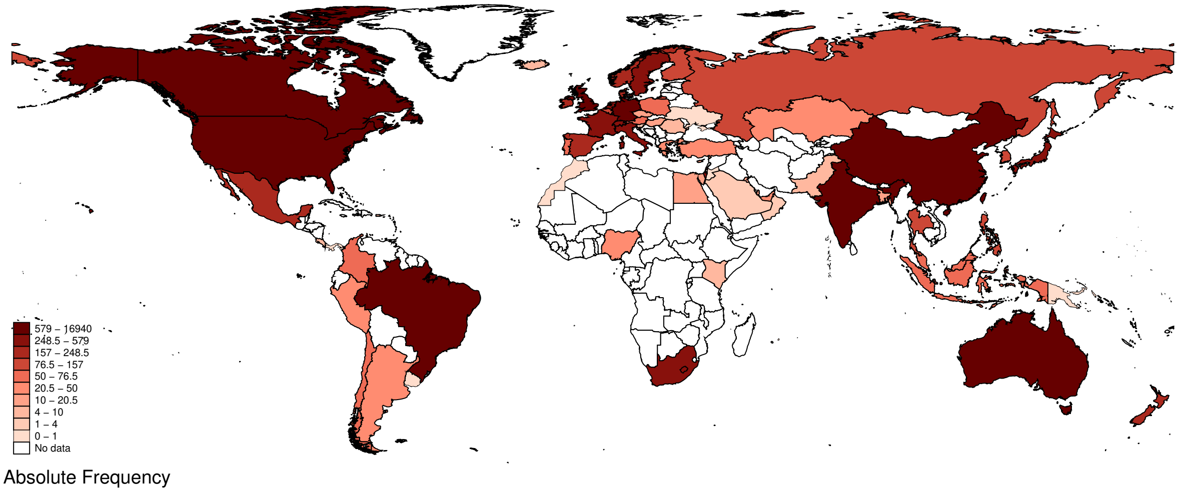


(a) Novel Topics

(b) Topics from Hassan et al. (2019) and Hassan et al. (2023).

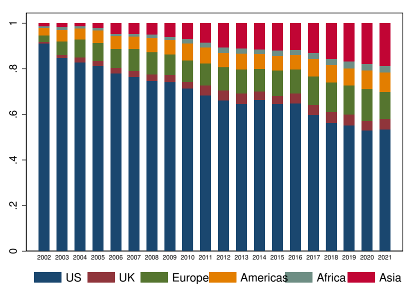
Notes: This figure plots our newly constructed topical indices on the left panel and existing indices on the right panel. All measures are in relative frequencies and standardized.

Figure 4: Global Distribution of Cyber Risk Exposure

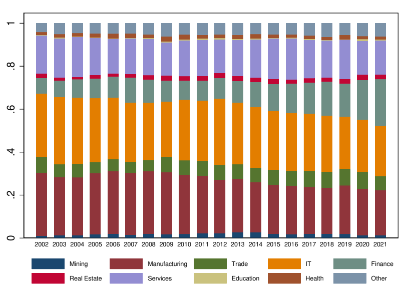


Notes: Regional distribution of $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$. Darker shades of brown indicate higher exposure. The sample is for 2021 only.

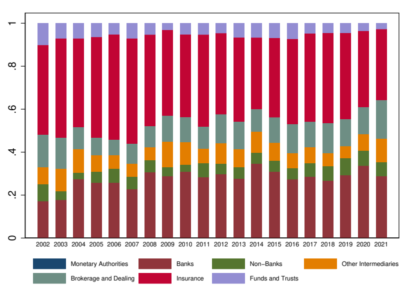
Figure 5: Regional and Industrial Decompositions over Time



(a) By Region



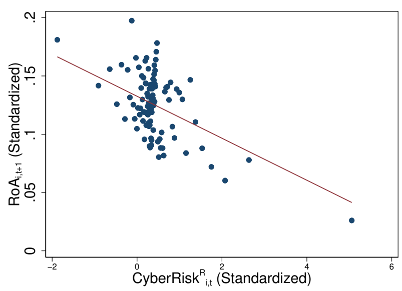
(b) By Industry



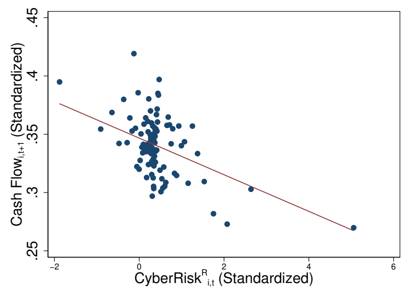
(c) By Finance Sub-Industry

Notes: Panel (a) plots the dynamic of the regional distribution of $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$ over time. Panels (b) and (c) plot the dynamic of the sectoral distribution of $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$ over time. Panel (b) plots major 2-digit NAICS industries, and Panel (c) plots finance sub-industries only.

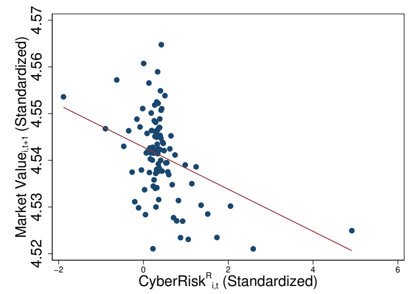
Figure 6: Scatterplots of Firm-Level Effects



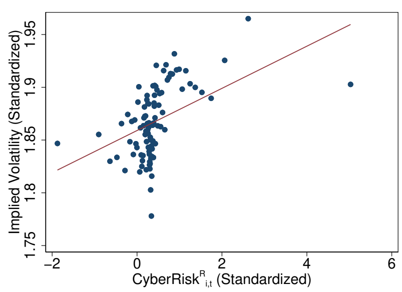
(a) Return on Assets



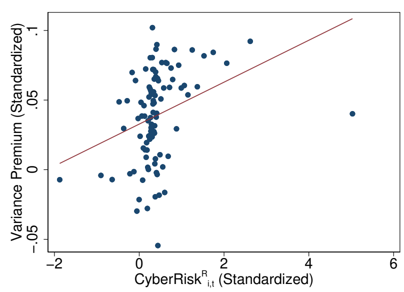
(b) Cash Flow Ratio



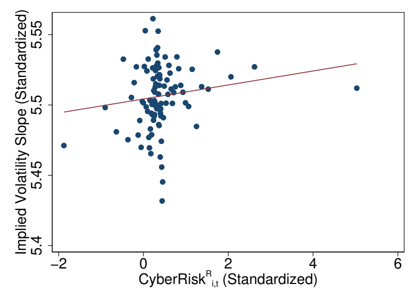
(c) Market Value



(d) Implied Volatility



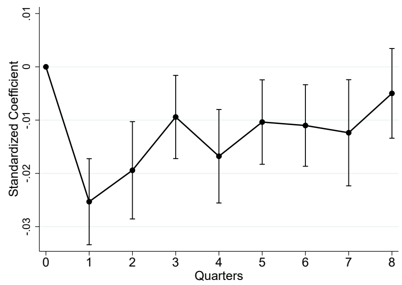
(e) Variance Risk Premium



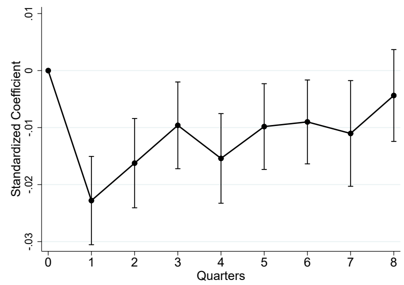
(f) Implied Volatility Slope

Notes: This figure plots (binned) scatterplots of firm-level regressions of balance sheet and option market aggregates on $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$. Each plot includes 100 equally-sized bins. Specifications include firm and quarter fixed effects as well as the following controls: firm size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets.

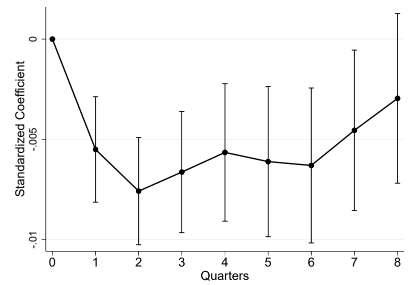
Figure 7: Dynamics of Firm-Level Effects



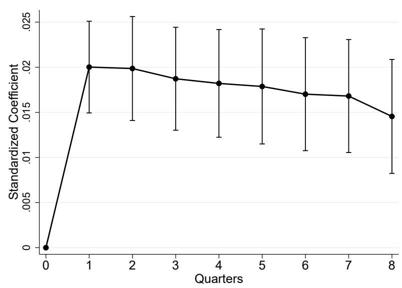
(a) Return on Assets



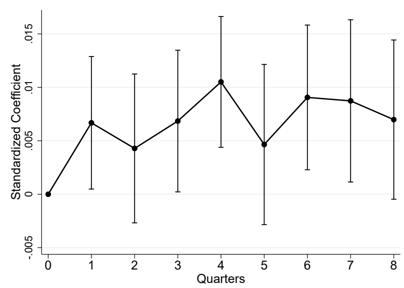
(b) Cash Flow Ratio



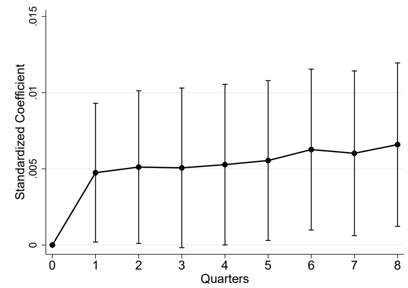
(c) Market Value



(d) Implied Volatility



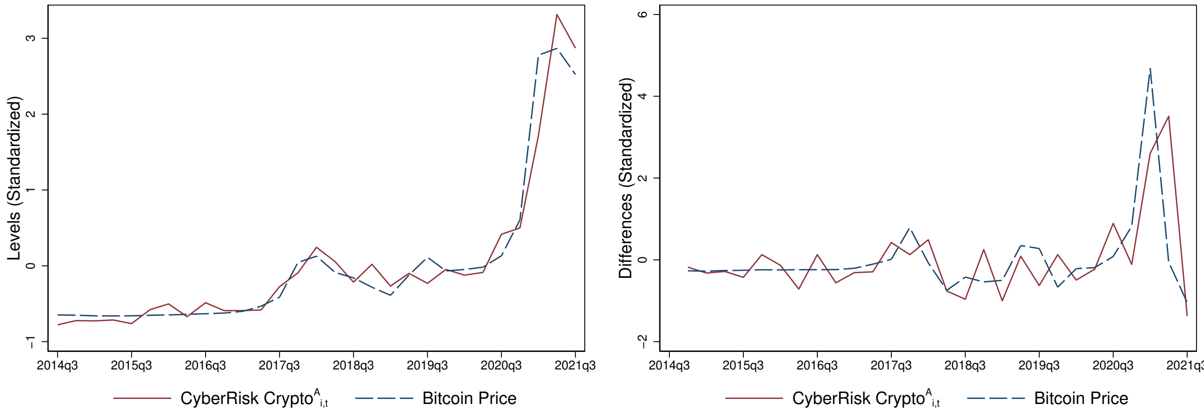
(e) Variance Risk Premium



(f) Implied Volatility Slope

Notes: This figure plots dynamic effects of firm-level regressions of balance sheet and option market aggregates on $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$. Each sub-plot shows relative quarters on the x-axis and standardized estimates with 90% confidence bands on the y-axis. Contemporaneous effects are normalized to 0. Specifications include firm and quarter fixed effects as well as the following controls: firm size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets.

Figure 8: Cyber Risk and Bitcoin

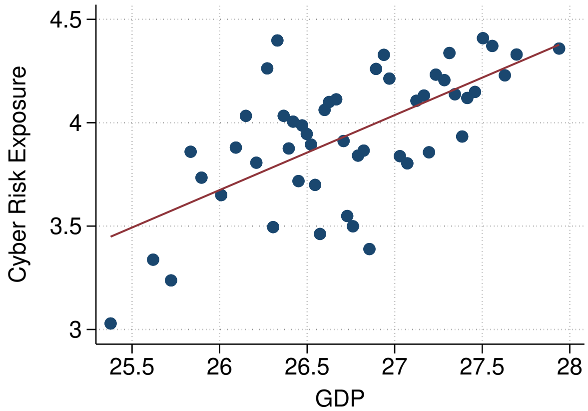


(a) Levels

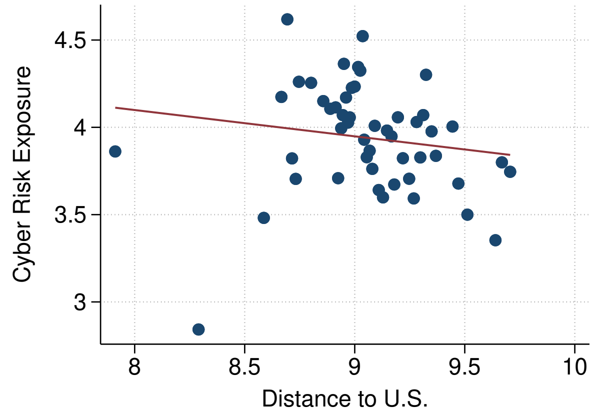
(b) First Differences

Notes: This figure plots the price of Bitcoin and CyberRisk Crypto$^{A}_{t}$; in levels (left panel) and first differences (right panel).

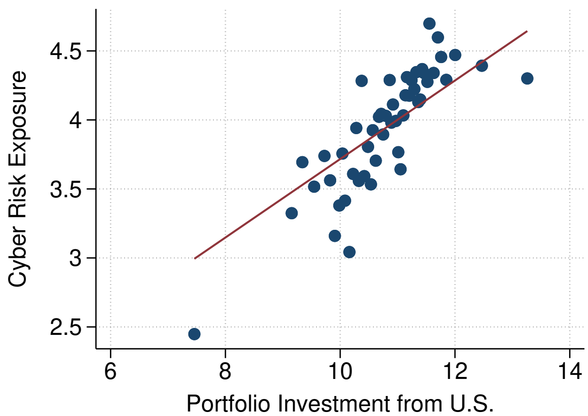
Figure 9: Gravity Model of International Cyber Risk Exposure



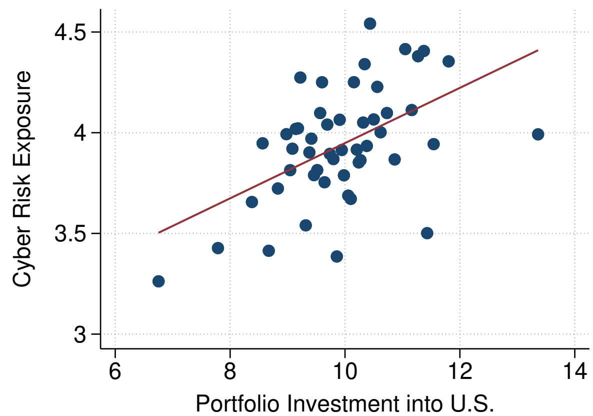
(a) GDP in Country c



(b) Distance to USA



(c) Portfolio Investment from USA



(d) Portfolio Investment into USA

Notes: This figure plots binned scatter plots and linear regression fit lines based on gravity panel regressions of $\text{CyberRisk}^{A}_{c,t}$ on the corresponding aggregates shown on the x-axes, as well as the time fixed effect and the usual country-level controls minus the variable on the x-axes.

Table 1: Summary Statistics

	N	Mean	St. Dev.	Min	Max
Cyber Security (absolute frequency)					
$\text{CyberRisk}^{A}_{i,t}$	134,103	1.33	3.23	0.00	244.00
$\text{CyberRisk Insurance}^{A}_{i,t}$	134,103	0.37	1.53	0.00	75.00
$\text{CyberRisk Legal}^{A}_{i,t}$	134,103	0.10	0.63	0.00	32.00
$\text{CyberRisk Crypto}^{A}_{i,t}$	134,103	0.03	0.87	0.00	214.00
$\text{CyberRisk SocialMedia}^{A}_{i,t}$	134,103	0.11	0.94	0.00	63.00
$\text{CyberRisk Uncertainty}^{A}_{i,t}$	134,103	0.00	0.11	0.00	19.00
$\text{CyberRisk PositiveSentiment}^{A}_{i,t}$	134,103	0.06	0.39	0.00	14.00
$\text{CyberRisk NegativeSentiment}^{A}_{i,t}$	134,103	0.24	1.00	0.00	56.00
$\text{CyberRisk NetSentiment}^{A}_{i,t}$	134,103	-0.18	1.05	-52.00	14.00
$\text{CyberRisk Politics}^{A}_{i,t}$	134,103	0.31	1.14	0.00	41.00
$\text{CyberRisk Disease}^{A}_{i,t}$	134,103	0.00	0.01	0.00	3.00
Stock Market (std.)					
Weighted Average Returns	133,209	0.13	1.00	-3.12	3.22
Cumulative Returns	133,210	0.02	1.00	-3.44	2.70
Realized Volatility	133,178	1.72	1.00	0.52	5.64
Option Market (std.)					
Implied Volatility	131,898	1.90	1.00	0.69	5.77
Variance Risk Premium	131,883	0.07	1.00	-5.10	4.07
Implied Volatility Slope	131,790	5.52	1.00	2.21	6.85
Firms (std.)					
Assets (log)	113,196	4.24	1.00	2.16	6.95
Firm Age (log)	113,196	4.27	1.00	0.87	5.52
Tobin's Q	112,543	1.34	1.00	0.46	6.21
Debt / Assets (Leverage)	107,269	1.29	1.00	0.00	3.96
Cash / Assets (Liquidity)	113,142	0.88	1.00	0.00	4.45
Intangibles / Assets	112,568	0.93	1.00	0.00	3.72
Operational Costs / Assets	113,118	1.14	1.00	0.02	5.13
Market Beta	133,209	3.00	1.00	0.87	5.83
Net Income / Assets (RoA)	113,196	0.12	1.00	-21.57	53.11
Cash Flow / Assets	113,196	0.34	1.00	-21.72	20.62
Market Value	97,223	4.54	1.00	-2.80	8.64

Notes: Select summary statistics of key variables used throughout the paper. Details on variable construction are provided in Appendix A.

Table 2: All Validated Terms Used in the Construction of $\text{CyberRisk}_{\text{i,t}}$

Term	Count	Term	Count	Term	Count	Term	Count
data	61111	securitysystems	255	personaldata	23	hacked	3
software	26418	operationalrisk	239	electronicsignature	22	plaintext	3
digital	25314	networkservices	230	softwareassurance	20	securityarchitecture	3
network	21859	login	190	dataintegrity	19	securityautomation	3
accountability	9179	credentials	189	spyware	19	attackpattern	2
availability	5960	datamining	182	systemarchitecture	19	behaviormonitoring	2
computer	3488	bot	124	antispyware	18	operationalincident	2
compromise	3291	exploit	120	password	15	systemdevelopment	2
disclosure	3030	cipher	117	situationalawareness	14	unauthorizedaccess	2
spam	1646	digitalsignature	106	spearphishing	14	whaling	2
router	1624	informationtechnology	100	blackhat	13	whitelist	2
vulnerabilitymanagement	1220	datacenter	98	unauthorized	13	zeroday	2
domain	1019	incidentresponse	97	dataarchitecture	11	airgap	1
encryption	916	accesscontrol	92	encode	11	attacksignature	1
firewall	758	username	85	threatassessment	11	securityengineering	1
antivirus	714	threatanalysis	84	datarecovery	10		
confidentiality	674	dataaggregation	81	securitybreach	10		
datasecurity	630	systemoutage	78	informationcompliance	9		
bug	580	cyberevent	63	whitehat	9		
app	493	cyberattack	61	cardfraud	8		
accessmanagement	467	privacy	60	hacker	8		
criticalinfrastructure	457	blueteam	57	maliciouscode	8		
vpn	447	spillage	53	operationalevent	8		
identitymanagement	433	cyberspace	48	pharming	8		
ict	428	authenticate	47	collectionoperation	7		
breach	426	securityevent	46	cyberthreat	7		
intrusiondetection	409	worm	42	hack	7		
insiderthreat	374	informationplatform	39	operationstechnology	7		
informationsharing	330	cyberoperations	33	publickey	7		
personalinformation	305	networkresilience	30	honeypot	6		
virus	305	threatintelligence	26	spoofing	6		
incidentmanagement	294	decryption	25	operationaldisruption	5		
networksecurity	270	systemadministration	24	digitalforensics	4		
securitymanagement	259	emailcompromise	23	authenticity	3		

Notes: The list of all terms used in the construction of our baseline cyber risk exposure measures. This list corresponds to the set $\tilde{\mathbb{C}}$ in main text.

Table 3: Predicting Cyberattacks

Panel A: Independent Variable - $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$						
Dependent Variable:	Future Cyberattack					
	Within 1 Quarter		Within 4 Quarters		Within 8 Quarters	
	(1)	(2)	(3)	(4)	(5)	(6)
Odds Ratio	1.461***	1.337**	1.420***	1.353***	1.415***	1.353***
	(0.171)	(0.196)	(0.136)	(0.144)	(0.126)	(0.128)
Controls	-	✓	-	✓	-	✓
Sector FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	90664	70795	98868	79118	101860	81518
Pseudo R^2	0.146	0.208	0.137	0.195	0.130	0.182

Panel B: Independent Variable - $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (std.)						
Dependent Variable:	Future Cyberattack					
	Within 1 Quarter		Within 4 Quarters		Within 8 Quarters	
	(1)	(2)	(3)	(4)	(5)	(6)
Odds Ratio	1.100***	1.132***	1.103***	1.135***	1.124***	1.159***
	(0.034)	(0.044)	(0.029)	(0.035)	(0.035)	(0.041)
Controls	-	✓	-	✓	-	✓
Sector FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	90657	70789	98861	79112	101853	81512
Pseudo R^2	0.144	0.208	0.135	0.195	0.129	0.183

Notes: predictive logit regressions of the future cyberattack indicator on the present measures of cyber risk. Panel (A) reports results on the extensive margin, i.e for $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$. Panel (B) reports results on the intensive margin, i.e for $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table 4: Cyber Risk and Firm Characteristics

Dependent Variable:	$\text{CyberRisk}^{I}_{i,t}$	$\text{CyberRisk x Topic}^{I}_{i,t}$						
Topic:		Uncertainty	Neg Sentiment	Crypto	Legal	Insurance	Social Media	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Size)	0.1239***	0.0683***	0.0744***	0.0839***	0.0396**	0.0932***	0.1292***	0.1015***
	(0.0105)	(0.0241)	(0.0157)	(0.0274)	(0.0161)	(0.0113)	(0.0220)	(0.0118)
Market Beta	0.0228	-0.0428	0.0364	0.1598	0.0100	0.0111	0.2304**	-0.0618
	(0.0431)	(0.1344)	(0.0655)	(0.1169)	(0.0627)	(0.0447)	(0.1075)	(0.0515)
Intangibles / Assets	0.3727***	-0.0229	0.5514***	0.5231	0.1055	-0.0063	0.5732***	0.4113***
	(0.1064)	(0.2603)	(0.1409)	(0.3405)	(0.2100)	(0.1201)	(0.2220)	(0.1141)
Liquidity Ratio	0.9314***	0.0156	0.6941***	0.4572	0.1909	0.3154**	0.8975***	1.0677***
	(0.1258)	(0.3544)	(0.1539)	(0.3141)	(0.2205)	(0.1405)	(0.2498)	(0.1396)
S&P Rating	0.0336***	-0.0505**	0.0099	0.0213	0.0369**	0.0422***	-0.0004	0.0302***
	(0.0096)	(0.0257)	(0.0138)	(0.0332)	(0.0175)	(0.0108)	(0.0201)	(0.0111)
Tobin's Q	0.0558***	-0.0053	0.0165	0.0602***	-0.0086	0.0357***	-0.0017	0.0699***
	(0.0106)	(0.0262)	(0.0127)	(0.0195)	(0.0150)	(0.0110)	(0.0152)	(0.0101)
CAPEX / Assets	0.1566	0.2938	-0.2630*	0.3688	0.2234	0.0057	0.6502***	0.2575**
	(0.1139)	(0.3453)	(0.1549)	(0.3161)	(0.2026)	(0.1207)	(0.2449)	(0.1226)
Cash Flow / Assets	1.4022*	-3.2428*	3.1314*	4.0712*	1.0924	1.3453	4.2128**	1.0868
	(0.8205)	(1.9474)	(1.7582)	(2.1732)	(1.1282)	(0.8548)	(1.6604)	(0.8469)
Log (Age)	0.0123	-0.1462**	0.0639	0.0611	0.0525	-0.0576*	-0.0561	0.0657*
	(0.0279)	(0.0691)	(0.0420)	(0.0870)	(0.0479)	(0.0311)	(0.0595)	(0.0344)
Book to Market Ratio	-0.0100	-0.0254	-0.0431	-0.1140	-0.0149	0.0247	-0.0510	-0.0598**
	(0.0214)	(0.1210)	(0.0384)	(0.1025)	(0.0352)	(0.0251)	(0.0537)	(0.0282)
Leverage	-0.0108	-0.7640***	-0.0360	-0.0371	-0.0751	0.0333	-0.1785	0.0619
	(0.0828)	(0.2311)	(0.1225)	(0.2242)	(0.1611)	(0.0888)	(0.1586)	(0.0863)
ROA	-2.6089***	1.6450	-3.4023**	-2.5618	-0.9916	-2.6085***	-4.6030***	-1.9248**
	(0.7801)	(1.8636)	(1.6428)	(1.8885)	(1.0937)	(0.8230)	(1.5852)	(0.8056)
PP&E / Assets	-0.2952**	-0.4977	-0.4281**	0.1555	-0.2221	-0.2218	0.3451	-0.1753
	(0.1231)	(0.3268)	(0.1879)	(0.3771)	(0.2139)	(0.1398)	(0.2528)	(0.1424)
Debt Maturity Ratio	0.1084**	0.3267**	-0.0019	0.1370	0.0984	0.0402	-0.0643	0.0950*
	(0.0451)	(0.1532)	(0.0645)	(0.1493)	(0.0624)	(0.0476)	(0.0897)	(0.0496)
Equity Issuance Ratio	0.3517**	-1.2219	0.2342	0.4433	-0.1625	0.1394	0.9194***	0.3795**
	(0.1761)	(0.7588)	(0.2191)	(0.5186)	(0.2459)	(0.1768)	(0.2768)	(0.1801)
Turnover Ratio	-1.0399***	2.6806*	0.7454	-0.1935	-0.3653	-0.0458	0.5091	-1.5918***
	(0.3136)	(1.5234)	(0.4821)	(1.0376)	(0.4544)	(0.3683)	(0.6107)	(0.3623)
Operat. Costs / Assets	1.0756***	-2.7637*	-0.6287	0.5601	0.5186	-0.0047	-0.1705	1.4760***
	(0.3283)	(1.4946)	(0.5139)	(1.0743)	(0.4619)	(0.3813)	(0.6142)	(0.3774)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	67738	44650	67727	49681	67336	67738	65761	67672
Pseudo R^2	0.103	0.155	0.180	0.190	0.064	0.084	0.248	0.101

Notes: firm-level probit regressions of the indicator variable of cyber risk $\text{CyberRisk}^{I}_{i,t}$ on various firm-level aggregates. All firm-level variables are lagged by 1 quarter. Details on variable construction are provided in Appendix A. Specifications include country, sector, and quarter fixed effects. Standard errors clustered at the firm level are in parentheses.

Table 5: Firm-Level Stock Market Effects

Independent Variable:	$\text{CyberRisk}^{\text{I}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (std.)		
Dependent Variable (std.):	$\text{WRet}_{\text{i,t}}$	$\text{CRet}_{\text{i,t}}$	$\text{RV}_{\text{i,t,m}}$	$\text{WRet}_{\text{i,t}}$	$\text{CRet}_{\text{i,t}}$	$\text{RV}_{\text{i,t,m}}$
	(1)	(2)	(3)	(4)	(5)	(6)
Cyber Risk Measure	-0.012**	-0.011**	0.021***	-0.002**	-0.002**	0.014***
	(0.006)	(0.005)	(0.005)	(0.001)	(0.001)	(0.003)
Controls	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105903	105903	105893	103376	103376	103376
R^2	0.315	0.317	0.657	0.317	0.319	0.653

Notes: firm-level regressions of stock market aggregates on measures of cyber risk. WRet, CRet, and RV stand for weighted-average stock returns, cumulative stock returns, and realized stock volatility, respectively. Details on variable construction are provided in Appendix A. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table 6: Firm-Level Economic Effects

Independent Variable:	$\text{CyberRisk}^{I}_{i,t}$			$\text{CyberRisk}^{A}_{i,t}$			$\text{CyberRisk}^{R}_{i,t}$ (std.)		
Dependent Variable (std.):	$\text{RoA}_{i,t+1}$	$\text{CashFlow}_{i,t+1}$	$\text{Valuation}_{i,t}$	$\text{RoA}_{i,t+1}$	$\text{CashFlow}_{i,t+1}$	$\text{Valuation}_{i,t}$	$\text{RoA}_{i,t+1}$	$\text{CashFlow}_{i,t+1}$	$\text{Valuation}_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
Cyber Risk Measure	-0.027***	-0.024***	-0.006***	-0.007***	-0.006***	-0.001**	-0.025***	-0.023***	-0.006***
	(0.006)	(0.006)	(0.002)	(0.001)	(0.001)	(0.000)	(0.005)	(0.005)	(0.002)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	99060	99060	86188	99060	99060	86188	99056	99056	86184
R^2	0.410	0.455	0.965	0.410	0.455	0.965	0.410	0.455	0.965

Notes: firm-level regressions of balance sheet aggregates on measures of cyber risk. RoA, CashFlow, and Valuation stand for return on Assets, cash flow / assets, and market valuation, respectively. Details on variable construction are provided in Appendix A. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table 7: Firm-Level Option Market Effects

Independent Variable:	$\text{CyberRisk}^{\text{I}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{A}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (std.)		
Dependent Variable (std.):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk	0.030***	0.015**	0.016***	0.006***	0.002**	0.003***	0.022***	0.011***	0.006**
	(0.005)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105,272	105,263	105,192	105,272	105,263	105,192	102,749	102,740	102,662
R^2	0.793	0.380	0.855	0.793	0.380	0.855	0.791	0.379	0.855

Notes: firm-level regressions of option market aggregates on measures of cyber risk. IV, VRP, and SlopeD stand for implied volatility, variance risk premium, and implied volatility slope, respectively. Details on variable construction are provided in Appendix A. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table 8: Cyber Risk Topics and Firm-Level Option Market Effects

Panel A

Dependent Variable	Implied Volatility (std.)							
Topic:	Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CyberRisk x $\text{Topic}^{I}_{i,t}$	0.029***	0.012*	0.034**	0.023***	0.088***	0.049***	0.019***	0.024***
	(0.004)	(0.007)	(0.017)	(0.008)	(0.024)	(0.008)	(0.005)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105272	105272	105272	105272	105272	105272	105272	105272
R^2	0.792	0.792	0.792	0.792	0.792	0.792	0.792	0.792

Panel B

Dependent Variable	Variance Risk Premium (std.)							
Topic:	Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CyberRisk x $\text{Topic}^{I}_{i,t}$	0.022***	0.034***	0.016	0.027**	0.108***	0.041	0.013	0.009
	(0.008)	(0.011)	(0.023)	(0.013)	(0.034)	(0.412)	(0.008)	(0.008)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105263	105263	105263	105263	105263	105263	105263	105263
R^2	0.380	0.380	0.380	0.380	0.380	0.380	0.380	0.380

Panel C

Dependent Variable	Implied Volatility Slope (std.)							
Topic:	Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CyberRisk x $\text{Topic}^{I}_{i,t}$	0.013***	0.031***	0.010	0.006	0.052**	0.022***	0.011***	0.008**
	(0.004)	(0.006)	(0.012)	(0.008)	(0.026)	(0.007)	(0.004)	(0.004)
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105192	105192	105192	105192	105192	105192	105192	105192
R^2	0.855	0.855	0.855	0.855	0.855	0.855	0.855	0.855

Notes: firm-level regressions of option market aggregates on topical measures of cyber risk. IV, VRP, and SlopeD stand for implied volatility, variance risk premium, and implied volatility slope, respectively. Details on variable construction are provided in Appendix A. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table 9: Industry-Level Option Market and Economic Effects

Panel A: NAICS3

Aggregation:	Equally-Weighted				Assets-Weighted			
Dependent Variable (std.):	IV	VRP	SlopeD	RoA_{t+1}	IV	VRP	SlopeD	RoA_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$CyberRisk^{R}_{s,t}$ (std.)	0.024***	0.026***	0.015***	-0.019*	0.028***	0.028***	0.013**	-0.026**
	(0.009)	(0.010)	(0.005)	(0.012)	(0.009)	(0.009)	(0.006)	(0.012)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓	✓	✓	✓
Country x Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	14851	14850	14844	14322	14851	14850	14844	14322
R^2	0.794	0.553	0.872	0.437	0.796	0.518	0.864	0.45

Panel B: NAICS4

Aggregation:	Equally-Weighted				Assets-Weighted			
Dependent Variable (std.):	IV	VRP	SlopeD	RoA_{t+1}	IV	VRP	SlopeD	RoA_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$CyberRisk^{R}_{s,t}$ (std.)	0.021***	0.018**	0.009*	-0.018**	0.019***	0.017**	0.011*	-0.024***
	(0.008)	(0.008)	(0.005)	(0.009)	(0.007)	(0.007)	(0.006)	(0.008)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓	✓	✓	✓
Country x Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	24829	24828	24818	24138	24829	24828	24818	24138
R^2	0.794	0.526	0.846	0.391	0.796	0.494	0.846	0.392

Notes: Results from sector-level regressions. Specifications include industry and country x time fixed effects as well as usual controls that are aggregated to the sector-time level by averaging. Panels (A) and (B) report results for different levels of industry aggregation: 3-digit and 4-digit NAICS codes, respectively. Standard errors clustered by industry are in parentheses.

Table 10: Cyber Risk Spillovers Effects

	All Firms				Affected Firms				Peer Firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	IV	VRP	SlopeD	RoA_{t+1}	IV	VRP	SlopeD	RoA_{t+1}	IV	VRP	SlopeD	RoA_{t+1}
CyberRisk (std.)	0.006**	0.016***	0.004**	-0.010**	0.009**	0.017***	0.008**	-0.018***	0.013**	0.019**	0.009*	-0.023**
	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)	(0.006)	(0.010)	(0.005)	(0.011)
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry x Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	98965	98956	98875	99588	37035	37033	36992	35589	56754	56747	56716	54509
R^2	0.826	0.412	0.887	0.563	0.838	0.398	0.900	0.606	0.823	0.430	0.886	0.510

Notes: Results from regressions of firm-level outcomes on country x industry x time cyber risk exposure, constructed by averaging the firm-level $\mathrm{CyberRisk}^{R}_{i,t}$ measure. Affected firms are firms with positive firm-level exposure. Peer firms are defined as firms with zero firm-level exposure but which belong to a country, industry, and quarter with positive exposure. Industries are defined by the 4-digit NAICS code. All specifications include the usual firm controls as well as firm and industry x time fixed effects. Every dependent and independent variable has been standardized. Standard errors are double-clustered by industry and time.

Table 11: Cyber Risk and Bitcoin

Panel A: Levels

Dependent Variable:	P_{t+1}	P_{t+2}	P_{t+3}	P_{t+4}	CyberRisk Crypto_t^A (std.)				
CyberRisk Crypto_t^A (std.)	1.016***	1.561***	1.933***	1.740***					
	(0.214)	(0.292)	(0.500)	(0.579)					
P_t					0.965***	0.493***	0.469***	0.457***	0.448***
					(0.122)	(0.041)	(0.029)	(0.031)	(0.033)
P_{t-1}						0.590***	0.705***	0.716***	0.717***
						(0.059)	(0.045)	(0.039)	(0.045)
P_{t-2}							-0.138***	-0.176***	-0.127**
							-0.041	(0.060)	(0.060)
P_{t-3}								0.081	-0.186
								(0.153)	(0.253)
P_{t-4}									0.255
									(0.260)
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	28	27	26	25	29	28	27	26	25
R^2	0.729	0.662	0.482	0.321	0.932	0.983	0.984	0.985	0.986

Panel B: First Differences

Dependent Variable:	P_{t+1}	P_{t+2}	P_{t+3}	P_{t+4}	CyberRisk Crypto_t^A (std.)				
CyberRisk Crypto_t^A (std.)	-0.088	0.158	-0.155	-0.750**					
	(0.126)	(0.495)	(0.213)	(0.339)					
P_t					0.572***	0.438***	0.397***	0.396***	0.418***
					(0.050)	(0.048)	(0.044)	(0.046)	(0.032)
P_{t-1}						0.722***	0.766***	0.765***	0.767***
						(0.043)	(0.036)	(0.038)	(0.043)
P_{t-2}							-0.188***	-0.177**	-0.155**
							(0.042)	(0.064)	(0.063)
P_{t-3}								-0.050	-0.198
								(0.323)	(0.347)
P_{t-4}									0.528
									(0.315)
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	27	26	25	24	28	27	26	25	24
R^2	0.007	0.012	0.005	0.107	0.328	0.810	0.840	0.840	0.862

Notes: time-series regressions for the price of Bitcoin and cyber risk measures. Panel (A) and (B) report results in levels and first differences, respectively. Columns (1)-(4) are for specifications where the dependent variable is future price of Bitcoin and independent variable is CyberRisk Crypto_t. Columns (5)-(9) are for specifications where the dependent variable is CyberRisk Crypto_t and independent variables are contemporaneous and lagged prices of Bitcoin. Robust standard errors are in parentheses.

Table 12: Gravity Model of International Cyber Risk Exposure

Dependent Variable:	Cyber Risk Exposure in Country c					
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.67***	0.73***	0.40***	0.36***	0.31***	0.39***
	(0.03)	(0.03)	(0.06)	(0.06)	(0.06)	(0.07)
GDP Per Capita	0.46***	0.39***	0.34***	0.01	0.06	0.15**
	(0.04)	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)
Distance to U.S.	-0.38***	-0.46***	-0.34***	-0.15	-0.04	-0.17
	(0.12)	(0.10)	(0.12)	(0.18)	(0.11)	(0.15)
Common Legal Origin		1.18***	1.39***	1.04***	0.63***	1.35***
		(0.13)	(0.14)	(0.20)	(0.12)	(0.15)
Common Language		0.16*	0.03	0.10	0.08	0.13
		(0.08)	(0.09)	(0.08)	(0.07)	(0.09)
Common Colonizer		-0.04	-0.53***	-0.50***	-0.29**	-0.61***
		(0.11)	(0.14)	(0.14)	(0.12)	(0.14)
Exports to U.S.			0.07	0.07	0.04	0.07
			(0.05)	(0.05)	(0.04)	(0.05)
Imports from U.S.			0.19***	-0.08	-0.07	0.02
			(0.06)	(0.07)	(0.05)	(0.06)
Portfolio Investment into U.S. (Total)				0.14***		
				(0.04)		
Portfolio Investment from U.S. (Total)				0.28***		
				(0.11)		
Portfolio Investment into U.S. (Equity)					0.07**	
					(0.03)	
Portfolio Investment from U.S. (Equity)					0.39***	
					(0.04)	
Portfolio Investment into U.S. (Debt)						0.10***
						(0.03)
Portfolio Investment from U.S. (Debt)						0.10***
						(0.04)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	669	669	533	509	506	502
R^2	0.51	0.64	0.64	0.68	0.73	0.63

Notes: Results from panel regressions of the aggregated CyberRisk^A measure on various country-level indicators over the 2005-2019 period. Manufacturing trade data is from the WTO. Real GDP, common language dummy, common legal origin dummy, common colonizer dummy, and physical distance are all relative to the U.S. and the data comes from the U.S. International Trade Commission. Bilateral portfolio investment data is from the IMF Coordinated Portfolio Investment Survey. Dependent and independent (except for dummies) variables have been logged and standardized. Robust standard errors in parentheses.

Online Appendix for "The Anatomy of Cyber Risk"

Rustam Jamilov   Hélène Rey   Ahmed Tahoun

May 10, 2023

Contents

A Text Libraries **2**

B Earnings Call Snippets **5**

C Additional Summary Statistics **9**

D Additional Results and Robustness Checks **16**

A Text Libraries

Table A.1: Cyber Risk Terms Libraries

Source	Term
Financial Stability Board	accesscontrol accountability advancedpersistentthreat asset authenticity availability campaign compromise confidentiality courseofaction cyber cyberadvisory cyberalert cyberevent cyberincident cyberincidentresponseplan cyberresilience cyberrisk cybersecurity cyberthreat databreach defence-in-depth denialofservice detection distributeddenialofservice exploit identityandaccessmanagement incidentresponseteam indicatorsofcompromise informationsharing informationsystem integrity malware multi-factorauthentication non-repudiation patchmanagement penetrationtesting reliability situationalawareness socialengineering tacticstechniquesandprocedures threatactor threatassessment threatintelligence threatvector threat-ledpenetrationtesting trafficlightprotocol verification vulnerability vulnerabilityassessment
National Cyber Security Centre	allowedlist antivirus app attacker authentification botnet breach bringyourowndefice browser bruteforceattack byod certificate cloud credentials cyberattack cyberincident cybersecurity dataatrest denialofservice denylist dictionaryattack digitalfootprint downloadattack encryption enduserdevice exploit firewall hacker honeynet honeypot incident insiderrisks internetofthings iot macro malvertising malware mitigation network patching pentest pharming phishing platform ransomware router saas sanitisation smishing socialengineering softwareasaservice spearphishing trojan twofactorauthentification virtualprivatenetwork virus vpn vulnerability waterholing whaling zeroday
Cybersecurity and Infrastructure Security Agency	access accessandidentitymanagement accesscontrol accesscontrolmechanism activeattack activecontent advancedpersistentthreat adversary airgap alert allsourceintelligence antispyware antispywaresoftware antivirussoftware attackmethod attackmode attackpath attackpattern attacksignature attacksurface authenticate authenticity authorization availability behaviormonitoring blacklist blueteam bot botherder botmaster botnet bug buildsecurityin capability cipher ciphertext cloudcomputing collectandoperate collectionoperations computerforensics computernetworkdefense computernetworkdefenseanalysis computersecurity computersecurityincident confidentiality continuityofoperationsplan criticalinfrastructure cyberarchitecture cyberecosystem cyberexercise cyberindicdentresponseplan cyberinfrastructure cyberoperations cyberspace cyberspace dataadministration dataaggregation dataarchitecture dataintegrity datalossprevention datamining datarecovery dataspill datatheft decode decrypt decryption denialofservice digitalforensics digitalrightsmanagement digitalsignature disruption distributeddenialofservice dynamicattacksurface electronicsignature encipher encode encrypt encryption exfiltration exploitationanalysis exposure firewall forensics hashvalue hashing hazard ict ictsupplychain ictthreat identityandaccessmanagement incidentmanagement incidentresponse industrialcontrolsystem informationandcommunicationtechnology informationassurance informationcompliance informationrecovery informationsecurity informationsecuritypolicy informationsharing informationsystem informationtechnology insiderthreat integratedriskmanagement intrusion intrusiondetection itasset keylogger macrovirus maliciousapplet maliciouscode maliciousemail maliciouslogic maliciousmessage networkresilience networkservices operationalexercise operationalrisk operationstechnology oversightanddevelopment passiveattack password penetrationtesting personaidentifyinginformation phishing plaintext precursor privacy privatekey publickey secretkey securityarchitecture securityautomation securitybreach securityengineering securityevent securityincident securitymanagement securitypolicy securityprogram securitysystems softwareassurance spam spillage spoofing spyware systemadministration systemintegrity systemintrusion systemsarchitecture systemsdevelopment tabletopexercise tailoredtrustworthspace threatactor threatanalysis threatassessment trafficlightprotocol trojanhorse unauthorizedaccess vulnerability vulnerabilityassessment vulnerabilitymanagement whitelist

Notes: This table reports keywords - and their sources - that constitute the dictionary set $\mathbb{C}$ from main text. All keywords have been concatenated into single words for readability.

Table A.2: **Topical Terms Libraries**

Topic	Source	Term
Insurance	National Association of Insurance Commissioners	accidentinsurance accumulationperiod actualcashvalue actuarialreport actuary adjuster admittedassets admittedcompany advancepremium adverseselection annuitant annuity appraisal arbitration assessedvalue assetrisk assignedrisk assumedreinsurance authorizedcompany authorizedreinsurance beneficiary blanketcoverage businessinterruption businessownerspolicy captiveagent captiveinsurer carryingvalue casualtyinsurance catastrophebonds catastropheloss cededpremium cedingcompany charteredlifeunderwriter claim claimsadjustmentexpenses classrating coinsurance commercialgeneralliability commercialmultipleperil commercialpackagepolicy commercialproperty commission completedoperationsliability comprehensivegeneralliability comprehensivepersonalliability concurrentcausation constructionandalterationliability contingentliability contractualliability convertibleterminsurancepolicy copay correctiveorder coveredlives creditaccidentandhealth creditdisability credithealthinsurance creditlifeinsurance creditpersonalpropertyinsurance creditplacedinsurance cyberinsurance cyberriskinsurance deductible deferredannuity demutualization differenceinconditionsinsurance directincurredloss directloss domesticinsurer earlywarningsystem earnedpremium EDPpolicy employeebenefitliability employerliability encumbrance environmentalimpairmentliability environmentalpollutionliability excessandumbrellaliability excessoflossreinsurance extraexpenseinsurance facultativereinsurance fairplan fairvalue federalemergencymanagementagency federalfloodinsurance financialguaranty financialresponsibilitylaw firelegalliability foreigninsurer fraternalinsurance generalliability generallyacceptedaccountingprinciples goodwill grosspremium groupaccidentandhealth groupannuities groupannuity guarantyfund healthinsurance hullinsurance humanerror incurredclaims incurredlosses indemnity independentadjuster independentagent indexannuity insurableinterest insurance insuranceholdingcompanysystem insuranceregulatoryinformationsystem insurancetovalue insured insurer irrevocablebeneficiary jointlifeannuity jointunderwritingassociation levelpremiuminsurance liability lifeinsurance lifesettlement lifetimebenefit limitedpaymentlifeinsurance limitedpolicy lineofbusiness loanbackedsecurities loss lossadjustmentexpense lossesincurred lossesincurredbutnotreported lossesnotreported lossfrequency lossofuseinsurance losspayableclause lossratio lossreserve malpractice mandatedbenefits manufacturersoutputpolicy marginpremium mechanicalbreakdowninsurance moralhazard morbidityrisk multiperilinsurance municipalliability mutualinsurancecompany namedinsured namedperilcoverage negligence netadmittedassets netpremiumsearned nonadmittedassets nonadmittedinsurer nonproportionalreinsurance notionalvalue occurrence operationalinsurance operationalriskinsurance otheraccidentandhealth otherliability otherunderwritingexpenses packagepolicy planenrollment policy policydividend policyholderssurplus policyperiod policyreserve preferredrisk premium premiumsearned premiumsnet premiumswritten primaryinsurance priorapprovallaw productliability professionalerrorsandomissionsliability proratareinsurance protectedcell protectionandindemnityinsurance provisions proximatecause publicadjuster purepremium purerisk qualifiedactuary reinsurance reinsurer renewableterminsurance residualmarketplan retentionlimit retrocession retrospectiverating riskretentiongroup securitizationofinsurancerisk selfinsurance situsofcontract socialinsurance specialrevenuebond standardrisk stateofdomicile statutoryaccounting statutoryaccountingprinciples stockinsurancecompany stoploss subrogationclause substandardrisk superfund suretybond surplusline terminsurance titleinsurance totalliability umbrellaandexcess unallocatedlossadjustmentexpense unauthorizedreinsurance underlyinginterest underwriter underwritingrisk unearnedpremium universallifeinsurance unpaidlosses valuedpolicy valuedpolicylaw variablelifeinsurance variableuniversallife wholelifeinsurance writtenpremium

Table A.2: **Topical Terms Libraries (Continued)**

Topic	Source	Term
Law	Administrative Office of the United States Courts	acquittal activejudge admissible adversaryproceeding affidavit alternatejuror alternativedisputeresolution amicuscuriae appeal appelant appelate appellee arbitragecourt arbitragesettlement arraignment attorney automaticstay bail bankruptcy bankruptcyadministrator bankruptcycode bankruptcycourt bankruptcyestate bankruptcyjudge bankruptcypetition benchtrial burdenofproof businessbankruptcy capitaloffence casefile caselaw caseload causeofaction chambers Chaptereleven chapterfifteen chapternine chapterseven Chapterthirteen Chaptertwelve chiefjudge classaction commonlaw communityservice complaint concurrentsentence consumerbankruptcy contingentclaim contract conviction councel counseling court damages declaratoryjudgement defaultjudgement defendant deposition discharge dischargeabledebt disclosurestatement discover dismissal dueprocess exclusionaryrule exculpatoryevidence exemptassets exemptproperty facesheetfiling federalpublicdefender felony fraud fraudulenttransfer grandjury impeachment inculpatoryevidence indictment injunction interrogation interrogatories jointpetition judge judgeship judiciary jurisdiction jury juryinstructions law lawsuit lawyer legal legalclaim legalmotion legalpanel legalsentence legalsettlement lien liquidatedclaim liquidation litigation magistratejudge meanstest mistrial noassetcase nodischargeabledebt nonexemptassets oralargument parole petition petitionpreparer plaintiff plea pleading prebankruptcy prebankruptcyplanning precedent preferentialdebtpayment presentencereport pretrial pretrialconference pretrialservice priorityclaim probation proofofcaim propertyofestate prosecute prosecution reaffirmationagreement remand sanction sentencingguidelines sequester serviceofprocess standardofproof statementofintention statute statuteoflimitations subordination subpoena testimony tort undersecuredclaim unduehardship unliquidatedclaim unscheduleddebt unsecuredclaim verdict voluntarytransfer warrant witness
Crypto	Cryptopedia	aaveprotocol accountcheckertool adminkey airno alamedaresearch algorand alphahomora alphax altcoin anchorprotocol aragonclient aragonnetwork arpanet asymmetricencription atoken atomicswap automatedclearinghouse automatedmarketmaker avalabs backtesting baltoken bandchain bandprotocol binance binancecoin binancesmartchain binanceusd bitcoin bitcoincash bitcoindominance bitcoiner bitcoingenesisblock bitcoinnetwork bitcoinprotocol bitcointalk blockchain blockchainledger blockchainprotocol cardano casascius cashfusion cashshuffle chainlink coinbase coingecko coinjoin coinmarketcap coinmining coinmixer coinswap collateraltoken coloredcoin communitybackedstablecoin consortiumblockchain cryptanalysis crypto cryptoart cryptobackedloan cryptocollateralizedloan cryptocurrency cryptocurrencyexchange cryptocurrencypair cryptocurrencyprotocol cryptocurrencywallet cryptodefense cryptodotcom cryptographicalgorithm cryptographicallyverfiable cryptographicproof cryptography cryptojacking cryptolocker cryptology cryptomine cryptomining cryptomixer cryptoprotocol cryptoransomware cryptotoken cryptotumbler cryptowall cToken daicoin dataledger decentralizedapplication decentralizedexchange decentralizedexchangeaggregator deposittoken devnetcoin diem digitalasset digitalcurrency digitaldollar distributedledger distributedledgertechnology dogecoin dollarcoin dotcoin elrondegold eoscoin eosnetwork ether ethereum ethereumclassic ethereumtransaction ethereumvirtualmachine exchangecoin factom fiatbackedcoin fiatbackedstablecoin filecoin filecoinnetwork flexacoin flexanetwork fractionalownerhip fungibletoken gascoin graphnode graphprotocol greenlist happs holochain huobi huobiglobal hyperledger initialcoinoffering initialdataoffering ledger ledgerprotocol libra lighteningnetwork liquidnetwork litecoin mastercoin memecoin minebitcoin miningfarm miningpool miningreward minting mobilewallet monacocoin monero neocoin neofilestorage nestedblockchain nonfungibletoken omni oxprotocol paxgold paxosgold paxosstablecoin paxstandard permissionedledger permissionlessledger physicalbitcoin polkadot polkadotnetwork polygoncoin postmine privacycoin proofofstake proofofstorage proofofvalidation publicledger qtum raidennetwork ravencoin rippleledgernetwork ripplenetwork robinhood rupple satoshi satoshinakamoto securitytoken sidechain slimcoin smartcontract splana stablecoin stellar symmetriccryptography symmetricencryptionalgorithm symmetrickey terracoin testnetcoin tether tezosnetwork thetacoin token tokengenerationevent tokenization tokenswap troncoin uniswap usdcoin vechaincoin vechainnetwork vethor vitalikbuterin vtoken wrappedbitcoin xrp yearnfinance yearnprotocol
Social Media	Various	adblocker adsmanager airbnb API apple appletv applicationprogramminginterface avatar averageresponsetime baidu baidutieba bing blogger blogosphere bolt businesstobusiness businesstoconsumer buzzfeed chatbot clickbait clickthroughrate conversionrate costperclick crowdsourcing darkpost darksocial darkweb directmessage douyin facebook facebookmessenger feed gofundme google goviral hangouts hashtag influencer instagram keyperformanceindicator kuaishou linkedin mailchip mashup newsjacking patreon payperclick pinterest QQ quora qzone rambler reddit sinaweibo snapchat sociallistening socialmedia socialmediaROI socialmonitoring socialselling telegram tencent tiktok traffic tumblr twitch twitter uber viber vlogger webex wechat weibo weixin whatsapp yandex youtube zoom

Notes: This table reports topical keywords - and their sources - that constitute the corresponding topical libraries from main text. All keywords have been concatenated into single words for readability.

B Earnings Call Snippets

Table B.1: Earnings Call Snippets

Quarter	Company	$\text{CyberRisk}_{i,t}$	Text Snippet
2017q4	Equifax Inc.	38	not only tomorrow but going forward into the future so -hack- progressing and progressing very rapidly and as paulino has talked about so our conversations with customers ensuring they understand where we stand and then what were doing going forward; has there been any further progress in identifying whether the hack was done by a foreign -state- actor now bloomberg had run a story saying that there was evidence of that but it didnt sound like anything definitive has come out when is there a pronouncement about that yes what we have as i have my -testimony- declared theres no we; help frame how youre thinking about total costs of the -breach- and how much youre accruing for -breach- costs beyond the; have insurance to cover costs in connection with the data -breach- -incidents- with limits in excess of the current amount of; much of the usis -decline- was due to the data -breach- compared to mortgage market -decline- and if you anticipate customer; time its certainly -lost- its only been months since the -cyberevent- event so the discussions are ongoing so we were characterizing; the type of cost that weve incurred related to the -cyberevent- event are indeed under the general structure of the policy; entire industry to develop solutions to the growing cybersecurity and -data- protection -challenges- we believe the time is right for an; this -incident- requires a revisit to our entire it and -datasecurity- security practice including engaging industry experts to support the effort; are you going to think about that and handle that -disclosure- as we move forward throughout this process yes so we; you all spent maybe million or so on cybersecurity and -network- is that million a year the right base to think; the trust of customers and -improve- the -strength- of our -securitysystems- systems and our it systems then those fundamental capabilities still; this is a turning point for everyone interested in protecting -personaldata- data due to the impact of the cybersecurity -incident- we; what we have seen understood thats helpful and just can -breach- quantify the amount youre -insured- up until like is there a certain dollar amount that youre -insured- up to yes again were not going to disclose the cap on our -insurance- john you mentioned in usis outside of the breach that you thought and i think outside of mortgage too that you saw a little bit of weakness your competitors perhaps havent been seeing that can you just elaborate on trends and whether some of that is vertical marketspecific or perhaps clarify what youre seeing outside to the extent thats; free service which includes unlimited equifax credit reports bureau credit -breach- monitoring the ability to lock your equifax credit report social security number monitoring and identity theft -insurance- consumers can sign up for this free service until january equifax does have -insurance- to cover costs in connection with the data breach incidents with limits in excess of the current amount of the onetime -cost- incurred in the third quarter subject to the terms conditions and exclusions of the policies we are currently in discussions with our insurers regarding the cybersecurity incident as a reminder as our q filings will be made; whats your overall level of comfort that the majority of the cyber costs would be -cyberevent- by -insurance- as opposed to being more equifax ultimately yes so were not going to specifically disclose the specific amount of the coverage and in general we believe that the type of -cost- that weve incurred related to the cyber event are indeed under the general structure of the -policy- and were currently in discussions with the insurers around completing around moving forward with -insurance- claims and we would expect to make very good progress in this quarter on that process understood a quick final question from me you mentioned; our customers are also providing assistance by sharing their views -data- best practices for our integrated cybersecurity program were also working to monitor for the use of stolen personal identifiable information being used for fraudulent transactions and to date we do not have any evidence linking fraudulent problem activity to data stolen from equifax our customers have been generous with their time and willing to work with us the business units with the most direct impact from the incident were us information solutions the global consumer solutions as well as workforce solutions in usis as expected we saw deferrals of customers; support the effort after a comprehensive topdown review with inaudible -accountability- pwc we have taken immediate steps to improve our data security infrastructure we are hardening our networks changing our procedures to require closelooped confirmation when software patches are applied rolling out new vulnerability scanning tools and processes and increasing accountability mechanisms for our security and it team members we have also engaged pwc to assist us with our security program including strategic remediation and transformation initiative that will help us identify and implement solutions in the future so to strengthening our longterm data protection and cybersecurity posture were also working; in no way reflects the normal ongoing spend so obviously -breach- spend this year is up dramatically from what it has been in the past to turn to the dimension side of our spend i think probably the best thing i can reference is we spend in prior to the breach occurring so if you took a look at what our forecast was what we were budgeting we would have spent about of it and security combined on our security specifically so thats probably the best metric to use okay and then you also mentioned that there may be some free

Table B.1: Earnings Call Snippets (Continued)

Quarter	Company	$\text{CyberRisk}^{A}_{i,t}$	Text Snippet
2014q1	Target Corp	15	for those account numbers becomes less -desirable- but didnt the -breach- actually come from systems internally not necessarily coming from the; if traffic was down in the quarter presumably post the -breach- it was down pick a number like or is it; along with costs related to our recent -restructuring- and data -breach- along with small accounting and tax matters as weve worked; any -unauthorized- charges on their card accounts resulting from the -breach- we increased -fraud- detection for redcard holders and extended free; holiday merchandising and marketing plan immediately following news of the -breach- sales turned meaningfully -negative- but began to recover in january; it have -stopped- the actual theft of the credit card -data- or would it have -stopped- the personal information disclosure the; announcement that -criminals- had -gained- access to guest payment card -data- in our us stores in total fourth quarter comparable sales; invest to ensure this recovery continues beyond our efforts in -datasecurity- security and chip -enabled- technology we are applying insights from; our guests that they would have zero liability for any -unauthorized- charges on their card accounts resulting from the -breach- we; active leader in a retail industry cyber security and data -privacy- initiative in addition we are investing million in a new; the breach is and given where we are in the -breach- itd be inappropriate for me to speculate fair enough thank you so much hi thanks i have a couple questions just a quick followup on the breach costs you showed a net you got some -insurance- payments from the breach -cost- that you had is that a should we expect that or do you have any -insurance- for these potential costs whatever they may be or is that sort of a one off in the quarter and then i have a follow up just to be clear that was -insurance-; sentiment and -traffic- we believe that well continue to see -digital- trends in the next few months but the breach impact will diminish throughout the year as we engage in a vigorous effort to address our guests concerns and provide irresistible content and offers driving visits to our stores and digital channels in addition while economic trends are improving we continue to expect our lower and middle -income- guests to shop very cautiously in with that backdrop our current view is that us comparable sales will grow in the range of to in on those sales we expect a us segment
2021q1	Solarwinds Corp	10	potential -litigation- related to sunburst how are you thinking about -software- of these liabilities and customer claims and the degree to which solarwinds might be covered by its licensing agreements thank you for the question the point you made last is the most relevant one which is much like most software companies we have covered through our enduser licensing agreements and as you mentioned sunburst is not just a solarwinds specific issue but its a broader industry issue and as you also know most software vendors unfortunately have vulnerabilities that they disclose and correct on a goforward basis and so we; expecting that headwind to continue in and like we said -breach- going to make subscription sales a priority so if anything that headwind is only going to be even a little bit stronger as we move through right but i guess what im asking is the demand impact from the breach are you expecting the demand for your subscriptions not the mix but just demand for subscriptions in general to kind of hit a bottom here nearterm and then show improvement through the year yes absolutely as weve been building out our forecast for sterling we expect the biggest impact to; anything specific to the solarwinds environment we could not find -compromise- that was idiosyncratic to the solarwinds environment and if anything both our security hygiene security posture security tools consistent with what is practiced in the industry got it and then as you sort of manage through the solar storm compromise and work with your customers i guess the larger question is what is your vision for solarwinds as the company sort of comes out of this and as you look at what the company has been focused on the strategy how they sort of balance growth versus margins should we; combination of both security initiatives that sudhakar talked about as -cyberattack- as just some general increases in some of our expenses such as we expect our -insurance- -cost- to go up in and then theres other charges some of our professional fees will go up as a result of the cyberattack as well so really just the million to million number was to give you some context of what we expected the increase in our expenses to be not just for but as we move forward as well and bart id also like to clarify that these are not necessarily related; of our msp business these statements are based on currently -cyberattack- information and assumptions and we undertake no duty to update this information except as required by -law- these statements are also subject to a number of risks and uncertainties including the numerous risks related to the impact of the cyberattack on our business and a potential spinoff of our msp business additional information considering concerning these statements and the risks and uncertainties associated with them is highlighted in todays earnings release and in our filings with the sec copies are available from the sec or on our investor relations website; nonorion products one update that i believe is critical to -maliciouscode- is that we previously disclosed that the number of customers that may have installed an affected version of the orion software platform was fewer than based on our discussions with customers and our investigations into the nature of sunburst malicious code and the advanced trade craft of the threat actor we believe the number of organizations actually exploited through sunburst is substantially fewer than the number of customers that may have installed an affected version of the orion platform this is consistent with statements by national security advisor for cyber; processes that we believe goes well beyond industry norms to -maliciouscode- the integrity and security of all of our products we firmly believe that the orion software platform and related products as well as all of our other products can be used by our customers without risk of the sunburst malicious code we also formed a new technology and cybersecurity committee of our board current sitting members of our board who are cios with significant cybersecurity experience and i form the member committee this committee has the responsibility to assist our board in overseeing our response to the cyber incident and;

Table B.1: Earnings Call Snippets (Continued)

Quarter	Company	$\text{CyberRisk}^{A}_{i,t}$	Text Snippet
2020q3	First American Financial CP	10	the time of the -incident- and the adequacy of our -disclosure- controls there are also class actions pending which consistent with; have been active recently being conducted by the enforcement division -cyberattack- the new york department of financial services and the other by the sec enforcement staff the new york department of financial services notwithstanding the compliance finding in the examination report i mentioned earlier has alleged violations of new yorks cybersecurity requirements our efforts to resolve the matter have not been successful and as a result yesterday they filed a statement of charges we intend to conduct a vigorous defense which well focus on among other matters the examination report and its conclusion regarding our compliance with new yorks cybersecurity requirements; in our response to the information security incident the resulting -cyberevent- concluded that our it general controls environment is suitably designed and is operating effectively and that we adequately and appropriately detected analyzed contained eradicated and recovered from a security incident and that we are in compliance with new yorks cybersecurity requirements for financial services companies a number of other regulators have closed investigation without any findings only investigations have been active recently being conducted by the enforcement division of the new york department of financial services and the other by the sec enforcement staff the new york department of financial
2015q1	Home Depot Inc	12	-breach- related expenses in the fourth quarter our gross data -breach- expenses were approximately million after estimating our insurance recovery we; know if there was any recovery you can detect post -breach- im not sure this segment is very gasoline price sensitive; our insurance recovery we recorded approximately million of net data -breach- related expenses in the quarter for the year our gross; were approximately million and after expected insurance recovery our net -breach- expenses were approximately million our operating margin for the; think well repeat next year further we had million of -breach- related expenses that we arent projecting for next year; the year our expenses grew at of our sales growth -breach- in line with our most recent guidance and just a comment on data breach related expenses in the fourth quarter our gross data breach expenses were approximately million after estimating our -insurance- recovery we recorded approximately million of net data breach related expenses in the quarter for the year our gross data breach expenses were approximately million and after expected -insurance- recovery our net data breach expenses were approximately million our operating margin for the quarter was and for the year reached interest in investment -income- increased by million in; with our most recent guidance and just a comment on -breach- breach related expenses in the fourth quarter our gross data breach expenses were approximately million after estimating our -insurance- recovery we recorded approximately million of net data breach related expenses in the quarter for the year our gross data breach expenses were approximately million and after expected -insurance- recovery our net data breach expenses were approximately million our operating margin for the quarter was and for the year reached interest in investment -income- increased by million in the quarter reflecting a gain on sale of million shares of hd supply
2019q1	Marriott Intl Inc	19	expenses in the fourth quarter our gross data breach expenses -breach- approximately million after estimating our -insurance- recovery we recorded approximately million of net data breach related expenses in the quarter for the year our gross data breach expenses were approximately million and after expected -insurance- recovery our net data breach expenses were approximately million our operating margin for the quarter was and for the year reached interest in investment -income- increased by million in the quarter reflecting a gain on sale of million shares of hd supply common stock since the beginning of the year and including the fourthquarter transaction; with our most recent guidance and just a comment on -breach- breach related expenses in the fourth quarter our gross data breach expenses were approximately million after estimating our -insurance- recovery we recorded approximately million of net data breach related expenses in the quarter for the year our gross data breach expenses were approximately million and after expected -insurance- recovery our net data breach expenses were approximately million our operating margin for the quarter was and for the year reached interest in investment -income- increased by million in the quarter reflecting a gain on sale of million shares of hd supply; but we dont think well repeat next year further we -breach- million of data breach related expenses that we arent projecting for next year but we are projecting higher investments in our it infrastructure including people so if you look at the good guys that we had including the data breach related expenses it nets out to million if you added the million to the expenses that we reported for you would see that our expense growth factor for the year was more like and then if you run off of that adjusted base for the adjusted expense growth factor is; candidly weve -lost- a little bit of visibility with the -breach- and changeover of cards for the pro and our ability; through a -difficult- environment and even drove productivity through our -network- as we go forward there is a tentative labor agreement

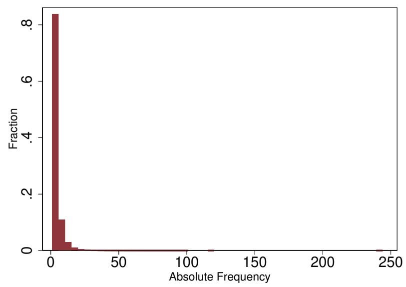
Table B.1: Earnings Call Snippets (Continued)

Quarter	Company	$\text{CyberRisk}^{\text{A}}_{\text{i,t}}$	Text Snippet
2018q4	Cisco Systems Inc	17	experience aienabled devices and enhanced interoperability across our onpremise and -availability- solutions yesterday we expanded our collaboration offerings with a full suite of cloud calling and team collaboration tools to extend our customers onpremise investments with new hybrid solutions from the cloud to the end user these innovations include the availability of broadsoft cloud calling with -webex- teams through service providers inch -webex- board and our new portfolio of huddle room solutions with room kit mini and -webex- share in summary we had a great quarter and our opportunity has never been greater our growth continued to accelerate as we executed; single architecture to provide that capability so one of the -data- things that we talked about this week was the need to drive multidomain architectures for our customers which actually give them the ability and youre seeing us extend and connect like -policy- in the campus with -policy- in the data centers so youre seeing aci being connected into dna and our softwaredefined access technology in the campus so that we can extend -policy- you saw this week with the branch where we integrated our sdwan with our security cloud security portfolio so and i think were seeing that come through; does some pruning if you could address that and thematically -router- like to get an understanding of how you think about the sdwan products youve been in this marketplace for a while now and it looks like its getting traction but my thought is this is a headwind for your router business but a tailwind for the sdwan platform so how do you see this playing out over lets say the next several quarters okay so let me address both the first is that the restructuring thats going on right now is first of all its not an opex reduction and; by (strong) execution differentiated (innovation) and our transition to more -software- and subscription offerings we are well positioned to capture significant;
2020q2	Oracle Corp	32	listing the additional wins i want to explain why were -computer- oracle cloud infrastructure is the worlds only secondgeneration autonomous cloud autonomous software technology the oracle autonomous database oracle autonomous linux autonomy is the defining technology that separates our gen cloud from amazons microsofts and googles generation cloud autonomous selfdriving computer systems eliminate human labor and thus eliminate human error there is nothing for humans to learn and nothing for humans to do eliminating human labor dramatically lowers the -cost- of running an autonomous system eliminating human error dramatically increases data security and system reliability all of the big data losses; and system reliability all of the big data losses at -data- were caused by human error there is no opportunity for any human error if your data is stored in an oracle autonomous system this is a very big deal the oracle autonomous database -provisions- itself configures itself encrypts the data itself patches itself and updates itself automatically scales itself up and down and continuously tunes itself as the database grows and user access patterns change and it does all of those things while the system is running theres no downtime required to patch theres no downtime required to installing new; at a count of we will this fiscal year add -firewall- gen oci regions allowing more customers to run in a public cloud without compromising data locality or data sovereignty requirements for customers who are wanting or needing to run their applications in their own data center behind their own firewall we uniquely offer oracle cloud at customer either for just the oracle database or for all of our oci cloud services including our saas applications none of the other cloud vendors have this kind of cloud customer offering to summarize oracles gen autonomous serverless elastic cloud infrastructure delivers better performance; the -cost- savings they achieved that they decided to move -informationtechnology- of their services out of aws and into oracle once oci demonstrated much better performance at a much better price that sealed the deal for x and its growing base of million monthly active users another win the omani information technology and communications group a year -contract- they built a dedicated oci gen cloud at customer data center in oman offering oci services oci public cloud services to all the different government agencies within oman this means theyll be able to have the full benefits of the public cloud oci
2019q3	Palo Alto Networks Inc	33	leadership position and customer happiness and customer success out in -breach- market not only that we are not going to rest on our laurels we have just announced to our field team were introducing an industryfirst security incident assurance service whereby if any of our customers unfortunately is in a breach situation or any customer in the industry were going to be there available until their breach is resolved irrespective of other what proportion of their products are palo alto products so we continue to want to be at the forefront of customer success and customer happiness in our ability to; because there arent enough alerts already and maybe sometimes were -data- to respond back to the endpoint now thats not enough we have to do something with the networks so theres a whole industry called nta network -traffic- analysis thats doing that on the network same thing they collect deep data from the network using separate sensors into another data lake process that with rules and whatever and machine learning and then maybe respond back usually they generate just more alerts and the same thing happens for iot and the same thing happens for public cloud and the same thing happens; addition to product releases we had several notable wins during -digital- quarter we displaced symantec and zscaler at a fortune us retailer to secure their data center and network of more than retail outlets we displaced zscaler and beat fortinet at a major -european- national health care provider in their digital transformation project theyre securing their hundreds of hospitals along with all of their patients and employees it was a great win for us in the quarter we beat crowdstrike and displaced symantec with our prisma and cortex platforms at a global -insurance- company with more than million policyholders and we

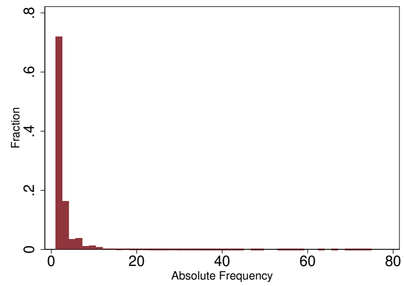
Notes: This table reports extracted snippets of text surrounding relevant discussions of cyber risk for select cyberattacked firms and cybersecurity firms, along with the associated earnings call date, company name, and the value of $\text{CyberRisk}^{\text{A}}_{\text{i,t}}$.

C Additional Summary Statistics

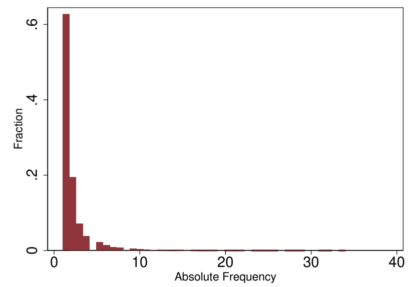
Figure C.1: Term Frequencies of All CyberRisk Measures



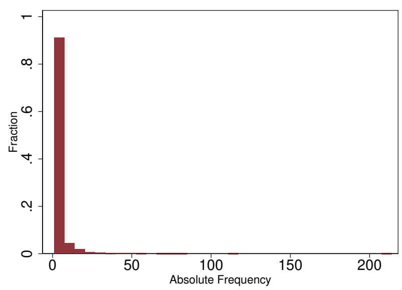
(a) $\text{CyberRisk}_{i,t}^{A}$



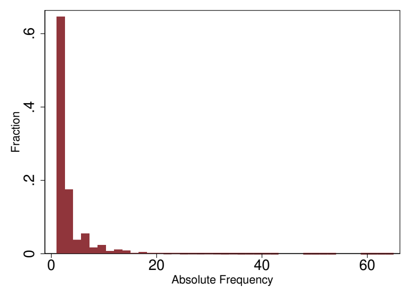
(b) $\text{CyberRisk Insurance}_{i,t}^{A}$



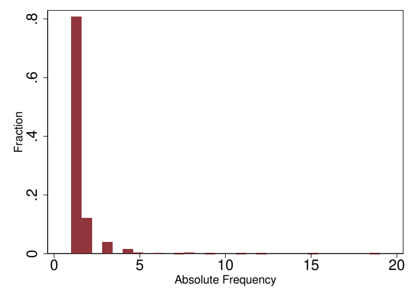
(c) $\text{CyberRisk Legal}_{i,t}^{A}$



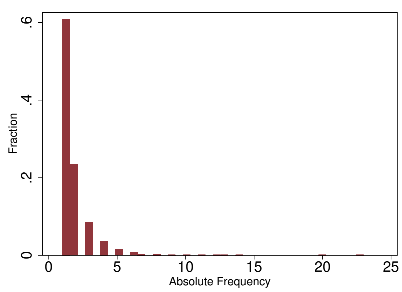
(d) $\text{CyberRisk Crypto}_{i,t}^{A}$



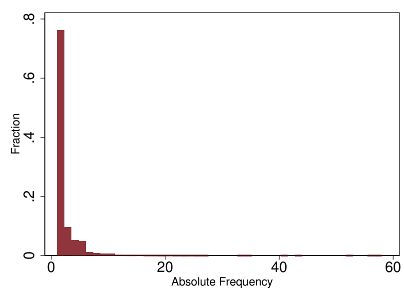
(e) $\text{CyberRisk SocialMedia}_{i,t}^{A}$



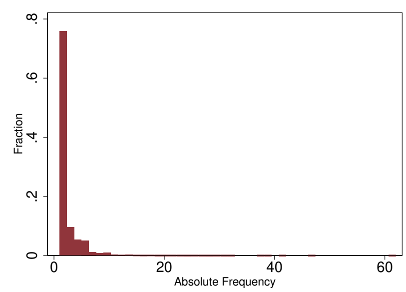
(f) $\text{CyberRisk Uncertainty}_{i,t}^{A}$



(g) $\text{CyberRisk PosSentiment}_{i,t}^{A}$



(h) $\text{CyberRisk NegSentiment}_{i,t}^{A}$



(i) $\text{CyberRisk Politics}_{i,t}^{A}$

Notes: This Figure plots histograms of every cyber risk exposure and topical measure used throughout this paper. In every panel, values have been pooled across all quarters and firms.

Table C.1: Summary Statistics by Country

	CyberRisk	CyberRisk x Topic$^{A}_{i,t}$								
		Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Net Sentiment	Politics
Country Code	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AE	191	84	12	0	13	1	6	27	-21	27
AR	438	147	37	21	109	1	4	41	-37	65
AT	640	407	15	7	40	0	17	65	-48	45
AU	7284	2770	294	123	704	21	443	1303	-860	1052
BD	14	5	0	0	4	0	0	2	-2	3
BE	1962	673	64	2	223	9	26	496	-470	297
BH	58	52	0	0	4	0	0	2	-2	0
BM	3212	2048	198	7	147	12	112	294	-182	289
BR	4375	1774	229	84	590	2	63	641	-578	734
BS	17	3	1	0	0	0	6	1	5	0
CA	19352	6199	1954	478	1980	54	768	3085	-2317	2823
CH	3961	1287	257	81	212	7	237	800	-563	699
CL	659	252	48	8	177	2	28	68	-40	46
CN	9017	3457	491	265	2173	31	279	881	-602	1139
CO	311	133	11	6	2	0	11	55	-44	71
CR	10	4	1	0	0	0	0	0	0	5
CY	156	69	5	5	29	0	1	13	-12	25
CZ	472	114	8	3	123	5	5	101	-96	95
DE	7263	2485	481	62	488	19	177	1546	-1369	1118
DK	1700	591	51	4	79	1	46	184	-138	637
EG	326	101	4	4	107	0	3	11	-8	80
ES	2771	1335	79	24	420	9	51	307	-256	351
FI	1330	326	44	4	164	5	62	376	-314	161
FO	5	3	2	0	0	0	0	0	0	0
FR	6772	2509	315	34	640	16	131	1589	-1458	978
GB	15375	5328	922	131	1001	39	709	2845	-2136	2933
GG	168	65	17	4	0	0	7	33	-26	28
GI	73	20	4	0	9	0	1	20	-19	14
GR	683	355	36	0	54	4	22	53	-31	132
HK	1929	759	94	46	364	2	54	239	-185	275
HU	216	70	3	0	58	0	2	37	-35	39
ID	1213	354	2	8	508	3	11	82	-71	219
IE	2916	631	184	26	84	5	123	338	-215	1325
IL	5684	1892	285	20	498	20	226	986	-760	999
IM	67	20	13	7	3	0	0	13	-13	9
IN	9507	4316	280	248	658	35	156	1906	-1750	1366
IS	34	2	2	1	0	0	1	9	-8	19
IT	4290	2448	87	22	387	13	99	435	-336	495
JE	258	48	4	97	3	0	7	12	-5	74
JO	0	0	0	0	0	0	0	0	0	0
JP	5630	1788	91	281	392	21	138	1237	-1099	1290
KE	59	28	0	0	11	0	0	4	-4	13
KR	2268	1177	13	25	382	7	19	257	-238	304
KW	89	13	12	2	4	0	0	33	-33	25
KY	541	196	38	3	50	0	60	89	-29	85
KZ	243	129	0	0	63	0	0	12	-12	38
LU	1142	301	68	12	180	3	19	261	-242	232
MA	50	7	0	0	19	0	0	0	0	24

Table C.1: Summary Statistics by Country (Continued)

	CyberRisk	CyberRisk x Topic$^{A}_{i,t}$								
		Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Net Sentiment	Politics
Country Code	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MC	107	56	25	0	0	0	1	2	-1	20
MH	7	2	0	0	0	0	0	0	0	3
MO	1	1	0	0	0	0	0	0	0	0
MT	38	7	1	0	2	0	3	6	-3	18
MU	15	9	2	0	0	0	0	1	-1	3
MX	1555	382	56	7	455	5	67	170	-103	298
MY	734	270	15	4	191	1	5	75	-70	157
NG	135	47	1	14	2	4	1	43	-42	17
NL	4649	1092	672	32	927	12	201	635	-434	640
NO	1758	512	133	50	172	2	37	329	-292	373
NZ	1163	598	14	5	118	0	24	202	-178	154
OM	165	48	6	0	54	0	2	3	-1	45
PA	91	18	0	0	63	0	0	2	-2	7
PE	107	56	0	0	1	0	0	27	-27	15
PG	38	21	0	0	1	0	0	0	0	10
PH	918	276	20	29	325	1	1	74	-73	180
PK	26	12	0	2	0	0	0	3	-3	7
PL	766	384	31	3	107	1	12	76	-64	105
PR	238	118	20	0	3	0	7	38	-31	42
PT	533	246	12	0	110	1	5	43	-38	72
QA	160	71	0	3	13	0	0	15	-15	55
RO	30	6	1	0	10	0	0	4	-4	5
RU	1376	497	52	25	352	1	26	132	-106	214
SA	10	3	1	0	0	0	3	0	3	1
SE	2966	815	135	23	509	7	86	605	-519	501
SG	1757	641	37	72	212	4	88	327	-239	318
SI	7	0	0	0	0	0	0	2	-2	4
TH	923	322	10	22	312	2	8	88	-80	136
TR	598	274	7	7	105	1	10	68	-58	108
TW	1421	460	30	8	185	2	153	276	-123	231
UA	3	0	0	0	0	0	0	3	-3	0
US	304123	81765	26862	4852	23746	1117	14096	53431	-39335	68881
UY	6	3	0	0	1	0	0	0	0	1
VE	22	1	1	0	15	0	0	0	0	4
VG	10	4	0	0	0	0	0	1	-1	2
VI	23	0	22	0	0	0	0	0	0	1
ZA	2549	922	94	19	372	6	111	299	-188	467
Total	453759	136714	35016	7332	41519	1514	19077	77769	-58692	93773

Notes: This table reports country-level absolute frequencies (total counts) of every cyber risk exposure and topical measure used throughout the paper. Country-level values are aggregates across all quarters.

Table C.2: Summary Statistics by Industry

Industry	Cyber Risk	Cyber Risk x Topic$^{A}_{i,t}$								
		Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Net Sentiment	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mining	8337	2374	1493	26	56	24	422	557	-135	2329
Manufacturing	112985	26459	8744	720	6117	299	6935	18297	-11362	31973
Trade	30797	8405	1857	851	4971	53	1967	5713	-3746	3601
IT	123380	36103	7418	1786	22902	392	3146	25612	-22466	17277
Finance	57462	32321	3326	2721	826	157	1771	5578	-3807	6950
Real Estate	9003	3054	911	37	710	22	336	1315	-979	1865
Services	77769	16924	7752	969	4580	380	2426	16440	-14014	22224
Education	1944	467	110	3	33	3	244	373	-129	465
Health	7696	3031	520	18	92	38	340	500	-160	2541
Other	24578	7614	2921	201	1237	146	1495	3456	-1961	4573
Total	453951	136752	35052	7332	41524	1514	19082	77841	-58759	93798

Notes: This table reports industry-level absolute frequencies (total counts) of every cyber risk exposure and topical measure used throughout the paper. Industry-level values are aggregates across all quarters.

Table C.3: Firm Characteristics by Geographical Region

Dependent Variable:	$\text{CyberRisk}^{\text{I}}_{\text{i,t}}$					
Region:	USA	Americas	Europe	UK	Asia	Africa
	(1)	(2)	(3)	(4)	(5)	(6)
Log (Size)	0.119***	0.022	0.208***	0.722***	0.031	0.410
	(0.012)	(0.047)	(0.069)	(0.238)	(0.263)	(0.534)
Intangibles / Assets	0.945***	1.700***	0.511	-2.733*	5.615**	6.396***
	(0.102)	(0.367)	(0.584)	(1.498)	(2.739)	(2.065)
Liquidity Ratio	1.433***	1.553***	1.684**	-3.088	-0.955	4.894***
	(0.124)	(0.584)	(0.713)	(1.908)	(0.794)	(1.632)
S&P Rating	0.034***	-0.071*	-0.009	0.383***	-0.068	0.090
	(0.010)	(0.042)	(0.067)	(0.115)	(0.173)	(0.149)
Tobin's Q	0.049***	0.108*	0.182***	0.266	-0.442	-0.377
	(0.009)	(0.055)	(0.039)	(0.218)	(0.431)	(0.410)
CAPEX / Assets	-0.064	-0.991**	0.572	2.124	-3.632***	-2.161
	(0.106)	(0.449)	(0.739)	(1.887)	(1.179)	(1.346)
Cash Flow / Assets	2.516**	5.489	11.047**	25.455	-25.686**	-4.711
	(1.027)	(4.415)	(4.899)	(19.376)	(12.642)	(6.139)
Log (Age)	0.011	0.033	-0.069	0.796*	-0.047	0.290
	(0.031)	(0.088)	(0.105)	(0.425)	(0.281)	(0.441)
Book to Market Ratio	0.021	-0.045	-0.069	-1.273*	-0.044	0.282
	(0.026)	(0.085)	(0.069)	(0.670)	(0.110)	(0.537)
Leverage	-0.055	0.408	-0.812	0.531	-3.452***	-2.237*
	(0.087)	(0.327)	(0.501)	(1.072)	(1.202)	(1.153)
ROA	-3.403***	-5.624	-11.620**	-26.777	23.122**	6.168
	(0.980)	(4.243)	(4.858)	(18.998)	(11.018)	(7.315)
PP&E / Assets	-0.107	-0.096	-0.141	-4.920**	1.903*	2.954
	(0.098)	(0.304)	(0.591)	(2.114)	(1.090)	(2.209)
Debt Maturity Ratio	0.075	0.422**	0.611**	-0.361	1.405***	-0.492
	(0.050)	(0.176)	(0.283)	(0.555)	(0.461)	(0.707)
Equity Issuance Ratio	0.510***	-0.497	2.251	12.169	-2.321	-3.813
	(0.179)	(0.500)	(1.624)	(13.377)	(3.592)	(2.943)
Turnover Ratio	-0.827**	-0.860	-0.689	-6.802	7.480*	2.082
	(0.327)	(1.700)	(2.602)	(11.810)	(4.489)	(13.733)
Operat. Costs / Assets	0.863**	0.017	-0.189	9.327	-11.542**	1.418
	(0.349)	(1.724)	(3.135)	(13.466)	(4.872)	(15.626)
Time FE	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	63697	4189	1298	372	301	158
Pseudo R2	0.053	0.117	0.129	0.209	0.251	0.393

Notes: each column reports results from country- or region-specific firm-level probit regressions of the indicator variable of cyber risk $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ on various firm-level aggregates. All firm-level variables are lagged by 1 quarter. Details on variable construction are provided in Appendix A. Specifications include time fixed effects. Standard errors clustered at the firm level are in parentheses.

Table C.4: Firm Characteristics by Industry

Dependent Variable:	$\text{CyberRisk}^{I}_{i,t}$									
Industry:	Mining	Manufacturing	Trade	IT	Finance	Real Estate	Services	Education	Health	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log (Size)	0.134***	0.115***	0.098***	0.125***	0.140***	0.144*	0.216***	0.130	0.142**	0.111***
	(0.036)	(0.017)	(0.029)	(0.037)	(0.022)	(0.075)	(0.057)	(0.238)	(0.066)	(0.020)
Intangibles / Assets	0.202	0.203	0.364	-0.275	0.863***	1.220***	0.764	-0.943	-0.519	0.893***
	(0.693)	(0.178)	(0.366)	(0.306)	(0.287)	(0.364)	(0.575)	(1.443)	(0.740)	(0.208)
Liquidity Ratio	-0.372	1.098***	0.723**	-0.232	0.933***	0.614	0.804	0.893	0.484	0.969***
	(0.747)	(0.188)	(0.341)	(0.390)	(0.347)	(0.697)	(0.610)	(1.258)	(0.874)	(0.290)
S&P Rating	0.013	0.061***	0.017	0.020	0.041**	-0.073*	0.069*	-0.072	0.043	-0.018
	(0.042)	(0.017)	(0.021)	(0.041)	(0.021)	(0.044)	(0.037)	(0.087)	(0.041)	(0.019)
Tobin's Q	0.127***	0.065***	-0.017	0.010	0.106***	0.131***	0.079***	-0.102*	0.030	0.034
	(0.034)	(0.012)	(0.042)	(0.031)	(0.037)	(0.045)	(0.028)	(0.056)	(0.086)	(0.027)
CAPEX / Assets	-0.096	0.002	-1.605***	0.574	0.139	1.150***	1.064	-0.177	-0.048	-0.578***
	(0.440)	(0.185)	(0.348)	(0.421)	(0.312)	(0.439)	(0.799)	(0.775)	(1.139)	(0.220)
Cash Flow / Assets	5.388**	0.929	-0.457	2.699*	17.678**	5.013	11.959	2.222	-8.324***	5.766**
	(2.406)	(1.137)	(3.622)	(1.560)	(8.365)	(4.368)	(7.657)	(7.687)	(2.277)	(2.727)
Log (Age)	-0.017	0.081*	0.195*	0.089	-0.232***	0.088	-0.131	-0.093	0.140	0.080
	(0.091)	(0.044)	(0.106)	(0.095)	(0.069)	(0.123)	(0.126)	(0.388)	(0.170)	(0.054)
Book to Market Ratio	0.019	-0.009	-0.047	0.126	-0.022	-0.069	-0.025	0.335**	-0.069	-0.040
	(0.038)	(0.048)	(0.052)	(0.095)	(0.058)	(0.091)	(0.219)	(0.153)	(0.200)	(0.048)
Leverage	0.398	-0.170	-0.323	-0.490**	-0.304	-0.337	-0.168	0.830	0.009	0.291*
	(0.293)	(0.135)	(0.222)	(0.229)	(0.229)	(0.360)	(0.399)	(0.811)	(0.575)	(0.167)
ROA	-4.381*	-2.510**	-0.648	-3.346**	-16.829**	-4.152	-12.270	-2.711	0.840	-5.569**
	(2.309)	(1.070)	(3.432)	(1.484)	(8.377)	(4.044)	(7.570)	(6.842)	(1.409)	(2.595)
PP&E / Assets	-0.492	-0.776***	-0.145	-0.367	-1.398***	-0.216	0.675	-1.394*	-0.914	-0.213
	(0.552)	(0.207)	(0.244)	(0.369)	(0.374)	(0.318)	(1.326)	(0.751)	(0.700)	(0.158)
Debt Maturity Ratio	-0.201	0.104	0.111	0.085	0.475***	0.376	-0.545***	-0.313	0.217	0.252*
	(0.163)	(0.067)	(0.129)	(0.154)	(0.108)	(0.381)	(0.183)	(0.628)	(0.293)	(0.133)
Equity Issuance Ratio	-0.918*	0.215	-0.552	0.468	1.118	0.835	0.774	-3.747	-0.010	1.055***
	(0.516)	(0.223)	(0.677)	(0.542)	(0.902)	(1.003)	(0.736)	(3.518)	(1.036)	(0.291)
Turnover Ratio	-1.644*	-1.238**	-1.415**	0.749	-3.767**	-2.665	1.126	-0.489	1.213	-0.394
	(0.930)	(0.504)	(0.640)	(1.205)	(1.908)	(2.674)	(1.318)	(2.544)	(4.057)	(0.938)
Operat. Costs / Assets	1.342	1.044*	0.850	-1.134	4.099**	2.219	-1.532	2.788	-0.826	0.544
	(0.915)	(0.548)	(0.637)	(1.330)	(2.019)	(2.824)	(1.405)	(3.102)	(4.294)	(0.980)
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	3936	29471	8522	5182	7826	1647	2471	359	1429	9192
Pseudo R2	0.040	0.061	0.050	0.036	0.077	0.104	0.063	0.098	0.057	0.046

Notes: each column reports results from industry-specific firm-level probit regressions of the indicator variable of cyber risk $\text{CyberRisk}^{I}_{i,t}$ on various firm-level aggregates. All firm-level variables are lagged by 1 quarter. Details on variable construction are provided in Appendix A. Specifications include time fixed effects. Standard errors clustered at the firm level are in parentheses.

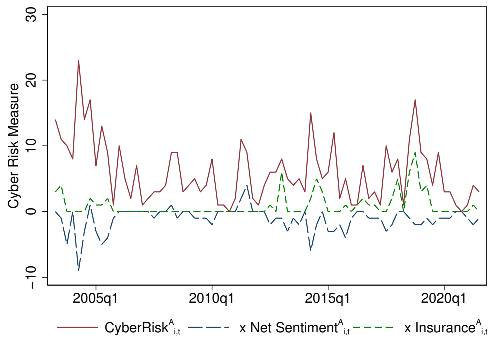
Table C.5: Firm Characteristics by Finance Sub-Sector

Dependent Variable:	$\text{CyberRisk}^{I}_{i,t}$				
Finance Sub-Sector:	Banks	Non-Banks	Other Interms.	Broker Dealers	Insurance
	(1)	(2)	(3)	(4)	(5)
Log (Size)	0.141***	0.215**	-0.213	0.114***	0.150***
	(0.035)	(0.108)	(0.324)	(0.044)	(0.048)
Intangibles / Assets	-3.852*	1.342	1.168	1.468***	0.620
	(2.053)	(1.551)	(0.787)	(0.469)	(0.431)
Liquidity Ratio	1.841***	2.899**	4.051***	0.742	0.287
	(0.547)	(1.338)	(0.773)	(0.543)	(0.477)
S&P Rating	-0.019	-0.443***	0.300**	0.011	0.142***
	(0.026)	(0.154)	(0.150)	(0.049)	(0.032)
Tobin's Q	-1.107	0.084	-0.094	0.081	0.308*
	(1.260)	(0.467)	(0.060)	(0.068)	(0.162)
CAPEX / Assets	0.680	-1.546	-0.311	-0.624	0.450
	(0.991)	(1.230)	(0.956)	(0.401)	(0.505)
Cash Flow / Assets	45.307	-39.933***	9.191	47.893***	-25.219*
	(52.397)	(12.948)	(18.586)	(14.422)	(14.876)
Log (Age)	-0.208	0.214	0.847***	-0.332***	-0.682***
	(0.128)	(0.216)	(0.282)	(0.116)	(0.148)
Book to Market Ratio	-0.020	0.150	-0.339*	0.072	-0.186
	(0.114)	(0.319)	(0.188)	(0.115)	(0.155)
Leverage	-0.930	1.335*	-1.304	-0.485	-1.144*
	(0.784)	(0.811)	(0.939)	(0.342)	(0.617)
ROA	-54.096	28.300	-3.182	-47.425***	26.491**
	(52.000)	(18.766)	(14.611)	(14.239)	(13.491)
PP&E / Assets	1.405	1.589	0.745	-1.129**	1.435
	(8.400)	(3.088)	(3.402)	(0.463)	(1.978)
Debt Maturity Ratio	-0.006	1.674	0.619	0.422*	-0.172
	(0.168)	(1.018)	(0.716)	(0.243)	(0.374)
Equity Issuance Ratio	11.685	-38.372***	3.695	2.294***	-3.888
	(11.721)	(8.327)	(2.941)	(0.874)	(3.399)
Turnover Ratio	6.450	11.095	25.343**	-6.515***	3.565
	(22.053)	(18.752)	(12.400)	(2.344)	(4.412)
Operat. Costs / Assets	21.552	11.604	-37.946**	5.984**	-2.912
	(21.821)	(19.495)	(18.110)	(2.387)	(4.391)
Year FE	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	3974	340	272	1632	1399
Pseudo R2	0.059	0.230	0.213	0.138	0.091

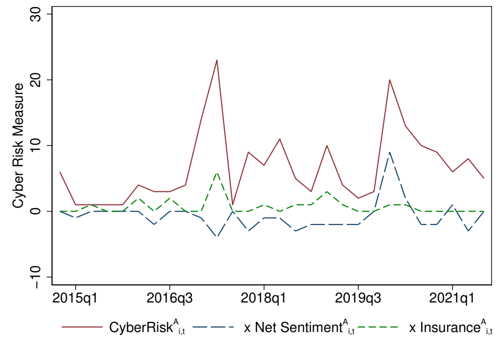
Notes: each column reports results from financial sub-industry-specific firm-level probit regressions of the indicator variable of cyber risk $\text{CyberRisk}^{I}_{i,t}$ on various firm-level aggregates. All firm-level variables are lagged by 1 quarter. Details on variable construction are provided in Appendix A. Specifications include time fixed effects. Standard errors clustered at the firm level are in parentheses.

D Additional Results and Robustness Checks

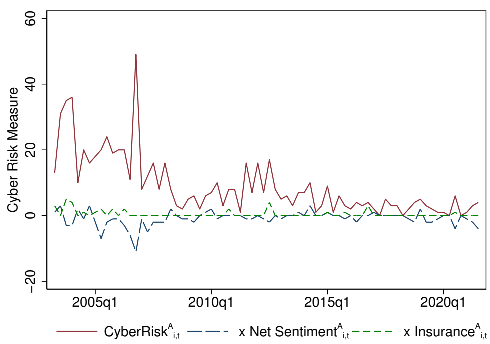
Figure D.1: Case Studies - Select Cybersecurity Firms



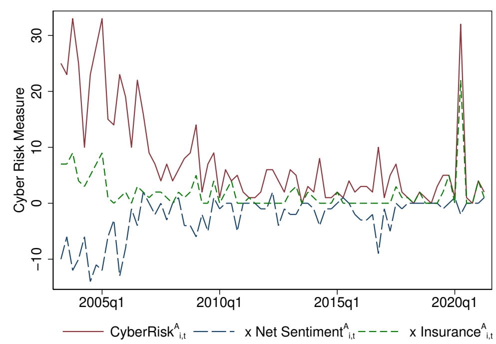
(a) Cisco



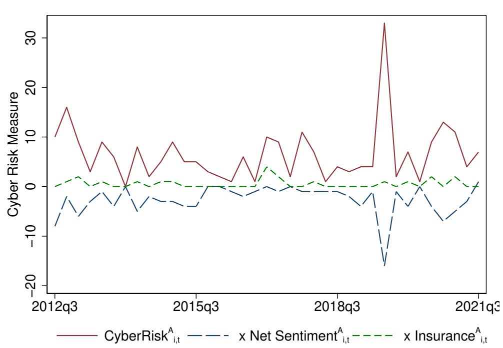
(b) CyberArk



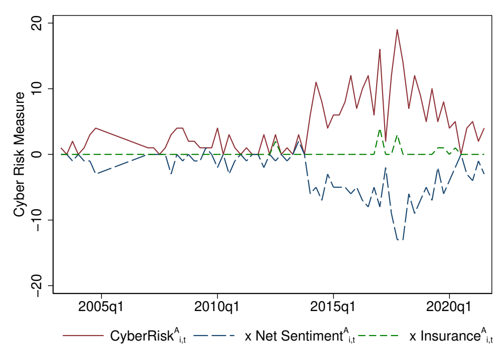
(c) Juniper Networks



(d) Oracle



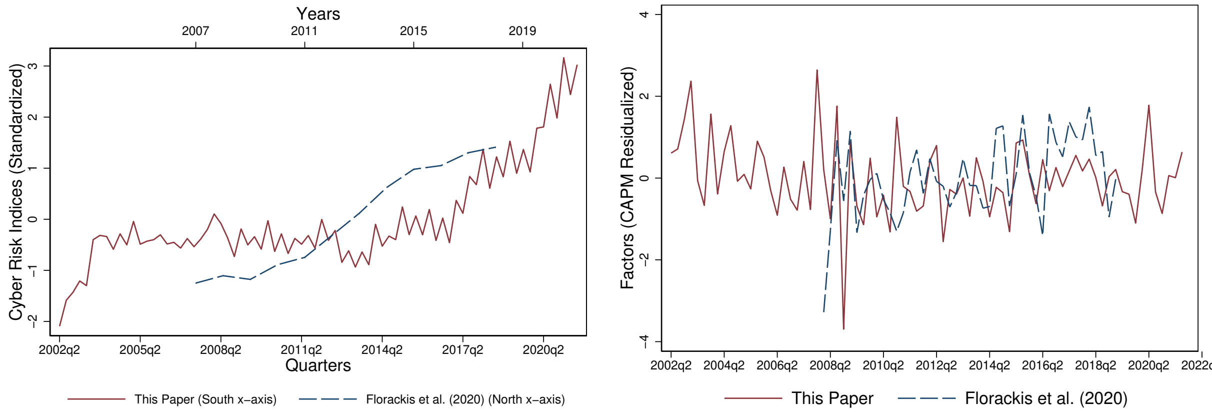
(e) Palo Alto Networks



(f) Synopsys

Notes: This figure plots select time series of our cyber risk measures for select cybersecurity firms.

Figure D.2: Comparison with Florackis et al. (2022)

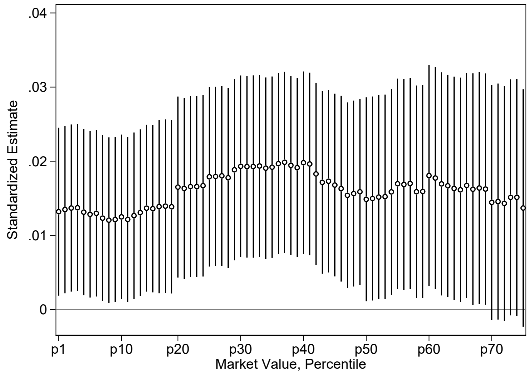


(a) Indices

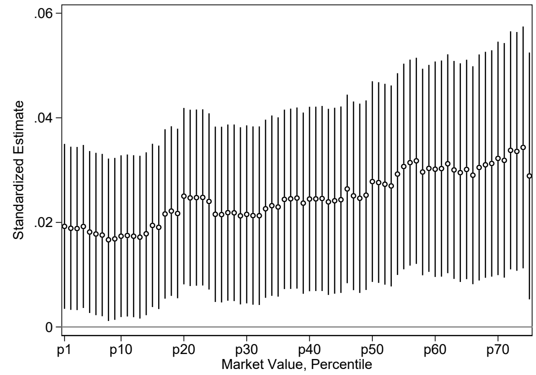
(b) Factors

Notes: This Figure compares this paper's main cyber risk exposure measure with the index that is developed in Florackis et al. (2022). The left panel plots the quarterly time series of CyberRisk$^{\mathrm{A}}$, developed in this paper from earnings calls (bottom x-axis) and the yearly index in Florackis et al. (2022) developed from 10-K files (top x-axis). The right panel plots quarterly factors in the two papers. Correlation between the factors is 0.39 with the p-value of 0.0186.

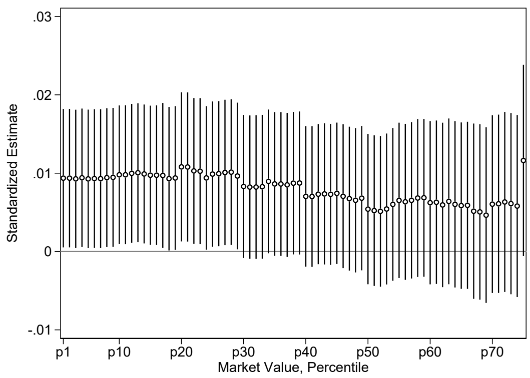
Figure D.3: Heterogeneous Spillovers Effects



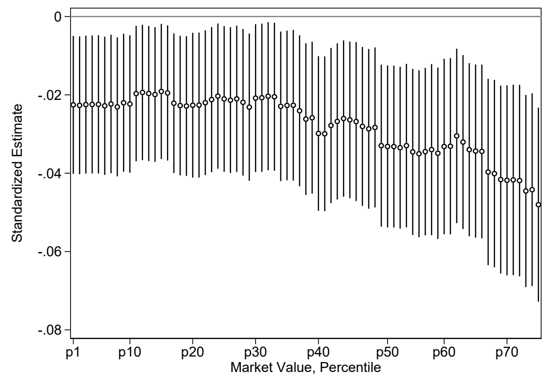
(a) IV



(b) VRP



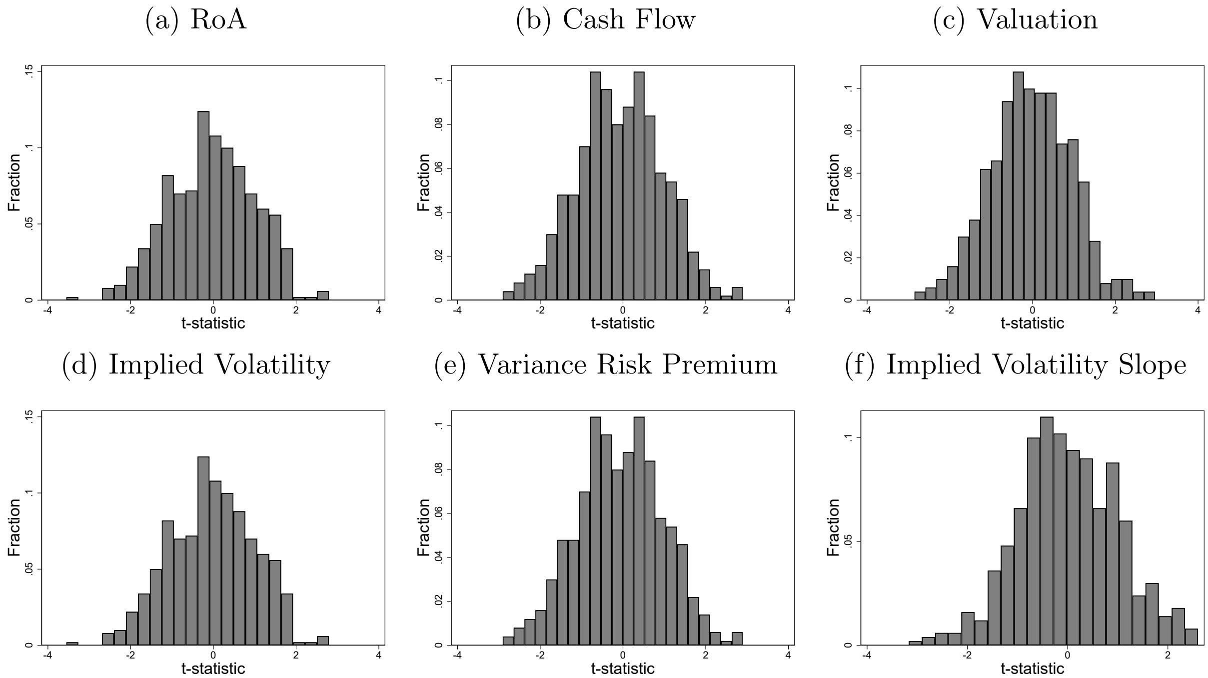
(c) SlopeD



(d) RoA

Notes: This figure plots heterogeneous spillover effects that complement average effects reported in Table 10. On each panel, each value on the horizontal axis restricts the sample to peer firms that are larger than the respective percentile of the following firm characterisc: market valuation. Peer firms are defined as firms with a zero firm-level cyber risk exposure but which belong to a country, industry, and quarter with positive exposure. Percentiles are computed specifically for each quarter, country, and industry. The vertical axis plots point estimates and 90% confidence intervals for the effects of cyber risk exposure at the country x industry x quarter level on the respective firm-level outcome. All specifications include the usual firm controls as well as firm and industry x time fixed effects. Standard errors are double-clustered by industry and time.

Figure D.4: Placebo Regressions: t-statistic Distributions



(a) RoA (b) Cash Flow (c) Valuation

(d) Implied Volatility (e) Variance Risk Premium (f) Implied Volatility Slope

Notes: Each panel on this figure presents a histogram of 500 t-statistics from regressions of corresponding firm-level variables on the $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ measure where the time series of $\text{CyberRisk}^{\text{I}}_{\text{i,t}}$ has been re-assigned randomly, with replacement. Each specification includes the usual firm controls, firm and time fixed effects, and standard errors that are clustered at the firm level.

Table D.1: Index Correlations

$\text{CyberRisk}^{R}_{i,t}$	1									
$\text{CyberRisk Insurance}^{R}_{i,t}$	0.6596	1								
	(0.00)									
$\text{CyberRisk Legal}^{R}_{i,t}$	0.3277	0.0624	1							
	(0.00)	(0.00)								
$\text{CyberRisk Crypto}^{R}_{i,t}$	0.2178	0.0161	0.004	1						
	(0.00)	(0.00)	(0.86)							
$\text{CyberRisk SocialMedia}^{R}_{i,t}$	0.3934	0.0772	0.0003	0.0106	1					
	(0.00)	(0.00)	(1.00)	(0.00)						
$\text{CyberRisk Uncertainty}^{R}_{i,t}$	0.1141	0.0128	0.0186	0.0019	0.0028	1				
	(0.00)	(0.00)	(0.00)	(1.00)	(1.00)					
$\text{CyberRisk PositiveSentiment}^{R}_{i,t}$	0.2136	0.0148	0.0214	0.0018	0.0161	0.0459	1			
	(0.00)	(0.00)	(0.00)	(1.00)	(0.00)	(0.00)				
$\text{CyberRisk NegativeSentiment}^{R}_{i,t}$	0.5304	0.2068	0.156	0.0422	0.0382	0.0974	0.053	1		
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
$\text{CyberRisk NetSentiment}^{R}_{i,t}$	-0.425	-0.191	-0.14	-0.039	-0.03	-0.076	0.319	-0.93	1	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		
$\text{CyberRisk Politics}^{R}_{i,t}$	0.4374	0.0972	0.0212	0.004	0.0307	0.041	0.016	0.0521	-0.044	1
	(0.00)	(0.00)	(0.00)	(0.89)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

Notes: pairwise correlation coefficients between all main cyber risk exposure measures used throughout the paper. P-values are in parentheses.

Table D.2: Predicting Cyberattacks - Topics

Dependent Variable:	Future cyberattack (Within 8 Quarters)							
Topic:	Insurance	Law	Crypto	Social Media	Uncertainty	Pos Sentiment	Neg Sentiment	Politics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CyberRisk x $\text{Topic}^{I}_{i,t}$ (Odds Ratio)	1.443***	1.619**	0.597	1.282	0.905	0.973	1.536***	1.244*
	(0.165)	(0.354)	(0.501)	(0.340)	(0.477)	(0.264)	(0.255)	(0.150)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Observations	81518	81518	81518	81518	81518	81518	81518	81518
Pseudo R2	0.182	0.181	0.180	0.180	0.180	0.181	0.182	0.181

Notes: predictive logit regressions of the future cyberattack indicator on the present measures of topical cyber risk. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table D.3: Recursive Dictionary Validation

Firm-Level Economic Effects

Independent Variable:	$\overline{\text{CyberRisk}}^{I}_{i,t}$			$\overline{\text{CyberRisk}}^{A}_{i,t}$			$\overline{\text{CyberRisk}}^{R}_{i,t}$		
Dependent Variable (standardized):	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk Exposure	-0.028***	-0.027***	-0.010***	-0.006***	-0.005***	-0.002***	-0.018***	-0.016***	-0.010***
	(0.006)	(0.006)	(0.002)	(0.001)	(0.001)	(0.000)	(0.006)	(0.005)	(0.001)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	76297	76297	70375	76297	76297	70375	76293	76293	70371
R2	0.408	0.450	0.969	0.409	0.450	0.969	0.408	0.450	0.969

Firm-Level Option Market Effects

Independent Variable:	$\overline{\text{CyberRisk}}^{I}_{i,t}$			$\overline{\text{CyberRisk}}^{A}_{i,t}$			$\overline{\text{CyberRisk}}^{R}_{i,t}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
Cyber Risk Exposure	0.035***	0.022***	0.016***	0.009***	0.001	0.005***	0.030***	0.005**	0.011***
	(0.004)	(0.007)	(0.003)	(0.001)	(0.001)	(0.001)	(0.004)	(0.003)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	81213	81206	81157	81213	81206	81157	81207	81200	81151
R2	0.791	0.276	0.879	0.791	0.276	0.879	0.791	0.276	0.879

Notes: This table reports the results from baseline firm-level linear regressions of economic and option-market variables on measures of cyber risk exposure that were obtained with the recursive dictionary validation procedure. Keyword-level predictive logit regressions are run recursively, for each year, over the 2005-2019 period for which the PRC cyberattack indicator data is available. Keywords with an odds ratio of less than or equal to unity are discarded for each recursion, and a new measure $\overline{\text{CyberRisk}}_{i,t}$ is constructed from the resulting time-varying dictionary.

Table D.4: Asymmetric Effects

Panel A: Asymmetric Firm-Level Economic Effects

Independent Variable:	$\widetilde{\text{CyberRisk}}^{I}_{i,t}$			$\widetilde{\text{CyberRisk}}^{A}_{i,t}$			$\widetilde{\text{CyberRisk}}^{R}_{i,t}$		
Dependent Variable (standardized):	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk Exposure	-0.008	0.011*	0.002	-0.001	-0.001	0.001**	-0.005	0.006*	0.003*
	(0.006)	(0.006)	(0.002)	(0.001)	(0.001)	(0.000)	(0.003)	(0.003)	(0.001)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	99060	99060	86188	99060	99060	86188	99056	99056	86184
R2	0.409	0.455	0.965	0.409	0.455	0.965	0.409	0.455	0.965

Panel B: Asymmetric Firm-Level Option Market Effects

Independent Variable:	$\widetilde{\text{CyberRisk}}^{I}_{i,t}$			$\widetilde{\text{CyberRisk}}^{A}_{i,t}$			$\widetilde{\text{CyberRisk}}^{R}_{i,t}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
Cyber Risk Exposure	-0.020***	-0.022***	-0.017***	-0.004***	-0.001	-0.003***	-0.015***	-0.003	-0.012***
	(0.004)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105272	105263	105192	105272	105263	105192	102749	102740	102662
R2	0.792	0.380	0.855	0.793	0.380	0.855	0.791	0.379	0.855

Notes: This table reports the results from baseline firm-level linear regressions of economic and option-market variables on measures of cyber risk exposure $\widetilde{\text{CyberRisk}}$ that are built on the 63 terms that were excluded from the baseline measure as a result of the dictionary validation procedure. All specifications include the usual firm controls, firm and time fixed effects, and standard errors clustered at the firm level.

Table D.5: Firm-Level vs Time-Series Dimensions

Dependent Variable (standardized):	IV		VRP		SlopeD	
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (standardized)	0.021***	0.018***	0.016***	0.014***	0.010***	0.004*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Mean $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (standardized)		0.075***		0.037***		0.190***
		(0.006)		(0.005)		(0.007)
Controls	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	102749	102749	102740	102740	102662	102662
R2	0.622	0.626	0.160	0.162	0.724	0.739

Notes: This table reports the results from baseline firm-level linear regressions of option-market variables on measures of firm-level cyber risk exposure $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ and the quarterly average of that measure Mean $\text{CyberRisk}^{\text{R}}_{\text{i,t}}$ (in columns (2), (4), and (6)). All specifications include the usual firm controls, firm and time fixed effects, and standard errors clustered at the firm level.

Table D.6: Robustness to Different Option Maturities

Panel A: 30-day options

Independent Variable:	$\text{CyberRisk}^{I}_{i,t}$			$\text{CyberRisk}^{A}_{i,t}$			$\text{CyberRisk}^{R}_{i,t}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk	0.029***	0.010*	0.013***	0.006***	0.002	0.003***	0.022***	0.010**	0.003
	(0.005)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105384	105366	105336	105384	105366	105336	102861	102843	102806
R^2	0.779	0.347	0.839	0.779	0.347	0.839	0.777	0.342	0.839

Panel B: 60-day options

Independent Variable:	$\text{CyberRisk}^{I}_{i,t}$			$\text{CyberRisk}^{A}_{i,t}$			$\text{CyberRisk}^{R}_{i,t}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk	0.029***	0.015**	0.016***	0.006***	0.003***	0.003***	0.022***	0.012***	0.003
	(0.005)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	105354	105345	105298	105354	105345	105298	102831	102822	102768
R^2	0.788	0.369	0.848	0.788	0.369	0.848	0.786	0.366	0.848

Panel C: 182-day options

Independent Variable:	$\text{CyberRisk}^{I}_{i,t}$			$\text{CyberRisk}^{A}_{i,t}$			$\text{CyberRisk}^{R}_{i,t}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk	0.030***	0.017***	0.017***	0.006***	0.002	0.004***	0.021***	0.009**	0.005*
	(0.004)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	104952	104951	104772	104952	104951	104772	102429	102428	102242
R^2	0.796	0.390	0.858	0.796	0.390	0.858	0.794	0.389	0.858

Notes: This table reports the results from baseline firm-level linear regressions of option-market variables on measures of firm-level cyber risk exposure for 30-day (Panel A), 60-day (Panel B), and 182-day (Panel C) options. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.

Table D.7: Restricted Sample (2005q1-2021q3)

Firm-Level Economic Effects									
Independent Variable:	$\text{CyberRisk}^{\text{I}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{A}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{R}}_{\text{i,t}}$		
Dependent Variable (standardized):	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation	RoA	Cash Flow	Valuation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cyber Risk Exposure	-0.025***	-0.023***	-0.006**	-0.007***	-0.006***	-0.001**	-0.024***	-0.022***	-0.006***
	(0.006)	(0.006)	(0.002)	(0.001)	(0.001)	(0.000)	(0.005)	(0.005)	(0.002)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	92900	92900	86188	92900	92900	86188	92896	92896	86184
R2	0.412	0.457	0.965	0.413	0.457	0.965	0.413	0.457	0.965

Firm-Level Option Market Effects									
Independent Variable:	$\text{CyberRisk}^{\text{I}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{A}}_{\text{i,t}}$			$\text{CyberRisk}^{\text{R}}_{\text{i,t}}$		
Dependent Variable (standardized):	IV	VRP	SlopeD	IV	VRP	SlopeD	IV	VRP	SlopeD
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
Cyber Risk Exposure	0.024***	0.018***	0.016***	0.005***	0.003***	0.003***	0.020***	0.011***	0.005**
	(0.004)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Level	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Frequency	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Observations	98766	98758	98694	98766	98758	98694	96243	96235	96164
R2	0.802	0.398	0.877	0.802	0.398	0.877	0.800	0.397	0.877

Notes: This table reports the results from baseline firm-level linear regressions of economic and option-market variables on measures of firm-level cyber risk exposure for a restricted sample that runs over the 2005:q1-2021:q3 period. Specifications include firm and time fixed effects as well as firm controls: size, age, Tobin's Q, leverage, liquidity, intangibles / assets, market beta, and operational costs / assets. Standard errors clustered at the firm level are in parentheses.