NBER WORKING PAPER SERIES

THE ANATOMY OF CYBER RISK

Rustam Jamilov
Hélène Rey
Ahmed Tahoun

Working Paper 28906
http://www.nber.org/papers/w28906

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

We thank our discussants, Jinfei Sheng and Ishita Sen, as well as Matteo Crosignani, Ralph Koijen, Anna Kovner, Richard Portes, Elias Papaioannou and participants of the AFA 2021 Annual Meeting and the ECB-RFS Macro-Finance Conference for useful comments and suggestions. Aakash Kalyani and especially Markus Schwedeler provided outstanding research assistance. Jamilov 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 view of the Haut Conseil de Stabilité Financière. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Anatomy of Cyber Risk
Rustam Jamilov, Hélène Rey, and Ahmed Tahoun
NBER Working Paper No. 28906
June 2021
JEL No. F3, G0, G12, G14, G15, G30, G32

ABSTRACT

We construct novel text-based measures of firm-level cyber risk exposure based on quarterly earnings calls of 12,000+ firms from 85 countries over 20+ years. We categorize each cyber-related discussion into topics that capture sentiment, monetary loss, country names, etc. We document new facts on the worldwide rise of cyber risk and its industrial and geographical composition. We characterize most affected firms and show that our indices can predict future cyberattacks. Cyber risk exposure has significant direct and contagion effects on stock returns. Finally, there is a factor structure in our firm-level measures and shocks to the common factor are priced.

Rustam Jamilov
London Business School
Regents Park
London NW1 4SA
United Kingdom
jamilovrustam@gmail.com

Hélène Rey
London Business School
Regents Park
London NW1 4SA
United Kingdom
and CEPR
and also NBER
hrey@london.edu

Ahmed Tahoun
London Business School
26 Sussex plc, Regent's Park
NW1 4SA London
atahoun@london.edu

1 Introduction

The World Economic Forum identifies systemic cyber risk as one of the most likely and impactful risks for firms (WEF, 2016). Major institutions have lost nearly $500 billion from operational risk events from 2011 to 2020, predominantly due to cyberattacks (ORX, 2020). The European Systemic Risk Board has recently characterized cyber security as a systemic risk to the European financial system (ESRB, 2020). According to the Center for Strategic and International Studies, cyber-crime caused economic losses up to 1% of global gross output as of 2014 (CSIS, 2014). Recent systemic risk surveys of financial market participants cite cyber security as a Top 2 most challenging risk for managing a firm, falling behind only political risk (BoE, 2020). There is rapidly escalating interest in cyber monitoring and macroprudential regulation from financial market regulators around the world (Kashyap and Wetherilt, 2019). Cyberattacks pose particularly large threats to trading and banking systems, with new and unforeseen avenues for the propagation of idiosyncratic attacks into systemic crises such as "cyber bank runs" (Duffie and Younger, 2019).

Despite such continuous interest from both industry participants and policy makers, empirical research on the economics of cyber security is very limited. The goal of this paper is to fill this gap by constructing comprehensive text-based measures of firm-level exposure to cyber risk by leveraging quarterly earnings calls and computational linguistics, in the spirit of Hassan et al. (2019)[1]. Our method relies on the application of recent advances in numerical natural language processing (NLP) techniques to textual information from quarterly earnings announcements data provided by Thomson Reuters Street Events.[2] 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. Many interesting dialogues take place during post-announcement Q&A sessions where investors, industry analysts, and other

[1] This general approach has been recently applied to the case of epidemic diseases like COVID-19 (Hassan et al., 2020a), the Brexit vote in the United Kingdom (Hassan et al., 2020b), and climate change risk (Sautner et al., 2020)

[2] See Gentzkow et al. (2021) for a general introduction to text-based modelling in economic research.

vested parties can ask questions about various pressing issues. The information contained in these dialogues is much richer and more multidimensional than in standard regulatory filings.

We process all the text from earnings announcements and Q&A sessions and construct firm-level measures of exposure to cyber risk and uncertainty by identifying and capturing relevant "textual bigrams". These bigrams are ordered combinations of words that relate to some topic of interest. For example, in the Hassan et al. (2019) study authors search for bigrams that are related to political uncertainty. In our case, this flexible approach can identify dialogues related to topics such as "cyber attack", "ransomware", "data breach", etc. Our complete quarterly dataset is available for 12,000+ firms in 85 countries and from 2002 till the present at quarterly frequency.

Our first and main measure is firm-level *exposure* to cyber risk. For each transcript, we calculate the number of times each term of interest gets mentioned. The full list of cyber terms is provided in Table 1 of the Online Appendix. We have over 30 unique terms that include monograms and bigrams like "cyber attack", "data loss," "hack", "DDOS", "data leak", etc. In order to construct this list of terms, we follow closely the documents and studies by the Cyber Policy Initiative at the Carnegie Endowment for International Peace. Firm-level total exposure is then defined as the sum of all of these mentions, normalized by the number of words in each transcript. We then aggregate the firm-level measure to the level of an individual industry, country, or the world.

An advantage of observing the full text of each earnings call transcript is that we can run *conditional* searches that identify cyber chatter that is concurrent to other topics of interest. For example, we independently identify any mentions of such words as "loss", "cost", "income", "monetary", etc. Terms like these belong to our "Monetary Loss" category. Then, we flag instances when a monetary loss term appears within 50 words of any of our cyber-related terms. We assign a value of 1 to each such instance and 0 otherwise. We repeat this exercise for 11 different topics that cover country names, sentiment, insurance and legal claims, crypto platforms or currencies, political events, etc. With the help of the country names topic, we are able to construct a global flowmap that identifies common geographical origins and destinations for cyber risk exposure.

Having constructed our indices and validated them with the data on reported cyber attacks, we document several stylized facts on global cyber risk as follows:

Fact 1: *The fraction of conference calls worldwide that discuss cyber risk is growing.* It has more than tripled since 2002, with a sharp increase after 2013. The global intensity of discussions, i.e. number of cyber mentions per call, has also gone up very significantly.

Fact 2: *The sentiment surrounding cyber risk is becoming increasingly negative.* The global cyber risk sentiment index has dropped roughly four-fold since 2002.

Fact 3: *Association of cyber-related discussions with uncertainty and risk is growing.* The global cyber risk uncertainty index has more than quadrupled since 2002, with a particularly marked increase after 2013.

Fact 4: *Most cyber risk related discussions emanate from firms head-quartered in the US. But, over time, and especially after 2013, the regional composition of global cyber risk exposure is shifting to Europe, the UK, Australia and to Asia. The share of cyber risk discussions taking place in the Americas excluding the US is also growing in the recent years while the share of Africa is relatively constant.* Within the US, we document strong clustering of exposure along the West and East coast lines, and in the South. California and Virgina (which includes DC) are the most exposed states followed by New York and Massachusetts. The Midwest is less exposed except may be for Illinois.

Fact 5: *Sentiment is heterogeneous. It is volatile but generally positive in some countries (Canada, India) and negative in others (US, UK).* We also find rich heterogeneity in cyber risk sentiment across US states: California is the most pessimistic state followed by Georgia while Virginia and the District of Columbia are the most optimistic.

Fact 6: *Industrial composition of global cyber risk exposure is shifting towards the financial sector.* The finance industry exhibited very little exposure before 2014. It is now the third most-affected sector after "IT" and "Professional Services" (the sector that includes the cyber-sensitive IT consulting firms) and before "Manufacturing". Within the financial sector, the breakdown of exposure across sub-industries is 40% financial intermediaries such as banks, 40% insurance, 10% broker-dealers, and 10% all others.

Fact 7: *Cyber risk discussions frequently also include terms from the "Insurance and Legal", "Monetary Loss", "Global Events", and "Country Names" topics.* The "Crypto" topic appears more frequently in recent years and spikes around local maxima in the price of Bitcoin. The "Global events" topic is clustered around famous global cyberattacks such as the Wannacry ransomware attack of 2017.

We proceed by studying the determinants of cyber risk exposure at the firm level. We estimate quarterly probit regressions of our measures of cyber risk on an array of balance sheet and income statement characteristics. To that end, we manually merge the Street Events firm announcements data with Compustat. We report three results. First, firm-level cyber risk exposure is robustly positively correlated with firm size, the share of intangible assets, and liquidity. This is also true for most of our topic-related indices. Second, the explanatory power of fundamentals is strong in the finance, real estate, services, and health sectors. Less so in manufacturing and trade. Within the finance sector, the explanatory power is stronger for insurance companies and less so for banks. Third, the explanatory power is stronger in the Americas (ex US), Europe, the UK, and Africa. Less so in the United States and Asia.

Can our text-based firm-level measures predict future cyberattacks? In order to answer this question, we match our data with the Privacy Rights Clearinghouse database of reported cyberattacks. We find that if a firm has positive exposure to cyber risk based on earnings call data, then it is significantly more likely to report a cyberattack within the next 8 quarters.

This result is robust to the inclusion of firm-level controls, time and industry fixed effects. We find that if a cyber risk discussion also mentions terms from the "Insurance and Legal", "Net Sentiment", or "Monetary Loss" topics, the probability becomes slightly greater. Finally, we find that the intensive margin plays a role - the probability increases with the *number* of times cyber risk gets mentioned per call, not just *whether* it does or not: it ranges from 1% for <5 mentions to over 10% for >30 mentions.

Our next major empirical exercise is on the asset pricing implications of cyber risk. We merge the Street Events dataset with CRSP at the level of a CUSIP. Our empirical strategy is to zoom in on weeks surrounding earnings calls with at least one positive firm-level cyber risk exposure. We find that cyber risk exposure has a negative and significant effect on stock returns of affected firms. Results are robust to the presence of country, week, and industry fixed effects and whether returns are value-weighted or equally-weighted. Results are significant not only for total exposure but also for the "Risk and Uncertainty", "Negative Sentiment", "Insurance and Legal", and "Monetary Loss" topics.

We then go beyond estimating direct effects and ask whether *unaffected* peer firms also experience losses. In other words, are there systemic contagion effects from those who are exposed to cyber risk onto those that are not? Formally, our left-hand-side variable is now the average weekly stock return of firms that have no cyber risk exposure of any kind but which are in the same *country and industry* as a firm that has had a positive exposure. The right-hand-side variable is cyber risk exposure of affected firms. We find evidence of spillover effects: high cyber risk exposure is associated with negative stock returns of firms that have *zero* exposure but are in the same country and industry. This result is significant if cyber discussions also include words from the "Risk and Uncertainty", "Negative Sentiment", "Insurance and Legal", and "Monetary Loss" categories. For robustness, we also check whether there is any effect on non-peers, i.e. firms that are from the same country but not in the same industry as the affected firm. Reassuringly, we find zero association. We interpret this broad result as new prima facie evidence that firm-level cyber risk can be a source of *systematic* risk in financial markets due to contagion effects or firm-to-firm networks.

Some of the variation in our aggregate indices of cyber risk is driven by multinational if not worldwide cyber attacks and incidents. Examples include the 2017 "WannaCry" and the 2016-2018 "NotPetya" ransomware attacks. This leads us to conjecture that there is a *factor structure* in firm-level measures of cyber risk. We construct a novel pricing factor based on aggregate cyber risk exposure - CyberE. We build our factor by extracting residuals from an AR(1) model that is fitted into our baseline index. Our monthly factor runs from 2002:m1 until 2020:m12 and will be updated regularly.

We construct 5 CyberE one-way sorted stock portfolios, as well as the high minus low portfolio, by regressing firm-level stock returns on our factor and extracting CyberE betas. We find that CyberE beta-sorted portfolios generate annualized spreads in average excess returns of -3.30%. This shows that stocks which are in the lowest CyberE beta sorted portfolio, i.e. suffer stock market losses when cyber risk is high, demand equilibrium compensation for this additional source of risk. These return spreads cannot be readily explained away with the CAPM or the Fama and French (1993) 3-factor models, nor the momentum factor. The spread is also not correlated with quantile-specific average market values.

In order to estimate the aggregate price of cyber risk, we perform the canonical Fama and MacBeth (1973) exercise. We first construct 10 test assets - CyberE beta-sorted stock portfolios from time-series regressions of returns on our factor. Then, we run the two-stage Fama-MacBeth procedure and produce cross-sectional regressions of average portfolio excess returns on the CyberE betas as well as the three Fama-French and momentum factor betas. We find that our factor is priced with a positive price of risk that is always significant, even in the presence of the market, size, book-to-market, and momentum pricing factors. Moreover, on top of the market factor, CyberE can individually explain around 70% of the cross-section of returns. The mean average pricing error with CyberE and the four canonical factors is only 22 basis points per year.

Literature Review There are several empirical studies that analyze the impact of cyber risk on economic and financial performance. Kamiya et al. (2020) employ the PRC database and estimate the effects of reported cyber attacks on firm-level stock returns and subsequent economic outcomes. Tosun (2019) performs a similar exercise and also reports that short-term market reactions to cyberattacks correlate with increased investors' attention as measured by Google trends. Eisenbach et al. (2020) study how cyber attacks get amplified through the U.S. financial system, with a focus on the wholesale payments network. Crosignani et al. (2020) show that cyber attacks can propagate through firms' supply chain networks. Woods et al. (2019) estimate the theoretical distribution of losses due to cyber attacks by leveraging regulatory filings and data on cyber insurance pricing and premia. Biener et al. (2015) study the distinct characteristics of cyber risks compared to other operational risks and emphasize significant problems resulting from interrelated losses, lack of data, and information asymmetries.

Relative to the aforementioned literature that employs reported cyber attack data, our text-based approach offers several substantial refinements. There are considerable issues when one analyses reported cyber attacks. First, there is a well known problem of significant under-reporting of cyber attacks (Amir et al., 2018). Second, there is potentially a substantial lag (days, if not weeks) between the day an attack gets reported to authorities and when it actually took place. This implies that any sort of event study approach with daily asset prices is problematic, particularly if information leaks prior to the disclosure. Under-reporting and time lags are not a problem in our setting because (a) during the Q&A sessions investors and analysts consistently pressure firm executives on issues that the latter could potentially ignore or postpone otherwise, (b) earnings announcements get recorded and reported to the general public immediately and (c) the Q&A sessions are non-rehearsed and questions are unexpected, deeming information leakage much less likely.

A study that is close to ours is Florakis et al. (2020) who use information in annual 10-K filings of U.S. firms to construct cyber security risk proxies. But we differ for at least three important reasons. First, the quarterly frequency of earnings calls (as opposed to annual data) allows for more robust cyber attack forecasting and asset pricing analyses. Second, earnings calls feature Q&A sessions which make cyber-related chatter much richer, more unrehearsed, multi-dimensional, and timely. Third, the international dimension of our data set gives us a global view of cyber risk and of the key players and affected industries.

The rest of the paper is structured as follows. Section 2 describes the data and our measurement approach. Section 3 presents new stylized facts and trends. Section 4 validates our measure by providing excerpts of transcripts for several notable cases. Section 5 predicts future cyberattacks with our text-based measures. Section 6 studies the determinants of firm-level cyber risk exposure. Section 7 analyzes direct and contagion effects of cyber risk exposure on stock market performance of publically listed firms. Section 8 conducts various time-series and cross-sectional asset pricing tests using our text-based measures. Section 9 studies the determinants of several traded cybersecurity ETFs. Section 10 provides a qualitative discussion of the importance of research on the economics of cyber risk with potential directions for future work. Finally, Section 11 summarizes and concludes.

2 Data and Measurement

Data Our primary data source is quarterly earnings conference calls of publically listed firms. We have collected the complete set of 334,438 English-language transcripts that cover 12,381 unique firms from 85 countries over the 2002q1-2020q4 period. We adopt computational linguistics algorithms to analyse the texts of quarterly earnings conference calls and tag conversations that relate to cyber risk, data breaches, or hacks. Our general approach follows closely the work of Hassan et al. (2019) on political risk and uncertainty.

Cyber Risk Exposure Our main firm-level measure is constructed by finding and counting cyber-related textual bigrams. The list of all cyber terms is based on the documents and studies by the Cyber Policy Initiative at the Carnegie Endowment for International Peace and is provided in Table 1 of the Online Appendix. If our algorithm detects that any bigram from that list gets mentioned X times in a given conference call transcript, it assigns the value of X to that particular transcript. We run a query on all transcripts in the database for each cyber term. Formally, for a given transcript i and quarter t, with total words $\mathrm{N_{it}}$, exposure to cyber risk term c is defined as:

$$\mathrm{CyberTerm}^{\mathrm{c}}_{\mathrm{it}} = \sum_{\mathrm{j=1}}^{\mathrm{N_{it}}} \mathbb{1}[\mathrm{j} = \mathrm{Cyber}^{\mathrm{c}}]\ \forall\ \mathrm{c} \tag{1}$$

The baseline transcript-level total exposure index is then defined as the normalized sum across all terms:

$$\mathrm{CyberExposure}_{\mathrm{it}} = \frac{1}{\mathrm{N_{it}}} \sum_{\mathrm{c}} \mathrm{CyberTerm}^{\mathrm{c}}_{\mathrm{it}} \tag{2}$$

We normalize by the number of words in each transcript to account for the length of discussions. We will be referring to the non-normalized version of the measure as $\mathrm{CyberExposure}^{\mathrm{T}}_{\mathrm{it}}$, which is simply the sum of all cyber terms mentioned per transcript.

Topics Our approach towards capturing the content of each cyber-related discussion involves running joint-search queries between cyber bigrams and terms that we associate with topics of special interest. We construct eight novel topics: "Country Names", "Crypto", "Insurance and Legal", "Monetary Loss", "Pandemics", "Social Media", "Politics", and "Global Events". As in Hassan et al. (2019) we also use the three topics "Risk and Uncertainty", "Positive Sentiment", and "Negative Sentiment". All topics and their associated terms are listed in Table 2 of the Online Appendix.

We now briefly describe the reason for including each of the eight new categories. The "Country Names" topic includes terms such as "Russia" and "North Korea" and helps us locate cyber chatter that also includes spoken country names. This, in turn, will allow us to identify potential geographical origins of cyber risk in Section 3. The "Crypto" topic includes terms such as "crypto", "blockchain", and "bitcoin". Reported and anecdotal evidence suggests that cyber ransomware incidents typically involve ransom demands in cryptocurrency.

The "Monetary Loss" and "Insurance and Legal" topics include terms such as "loss" or "reputation" and "insurance" or "liability", respectively, and flag cyber-related conversations that mention potential monetary and reputation losses or insurance and legal related claims and settlements. The "Global Events" and "Politics" topics include terms such as "Wannacry" or "Wiki leaks" and "election" or "Trump", respectively, and flag cyber-related conversations with references to significant worldwide incidents such as global cyberattacks or major political events and figures. The "Social Media" topic picks up terms such as "Zoom", "Facebook", or "Twitter". Social media platforms are potential targets for cyberattackers given the breadth of sensitive private information that they collect and store. Finally, the "Pandemics" topic includes terms like "corona" and "covid" in order to gauge the rise of digital risks and incidents during the COVID-19 pandemic.

Formally, for each term s of each topic k we count instances of cyber risk bigrams c being within a 50 word distance from s(k). In other words, we assume that if s(k) and c are within 50 words of each other, then the particular exposure to cyber risk was in the context of theme k. We construct topical indices in two steps. First, for a transcript i at time t with $\mathrm{N_{it}}$ words we compute the count of each topical term s(k) in the neighbourhood of c:

$$\mathrm{CyberTopicTerm}_{\mathrm{it}}^{\mathrm{s(k)}} = \sum_{\mathrm{j=1}}^{\mathrm{N_{it}}} \{\mathbb{1}[\mathrm{j} = \mathrm{Cyber}^{\mathrm{c}}] \times \mathbb{1}[|\mathrm{j} - \mathrm{s(k)}| < 50]\} \ \forall \ \mathrm{s(k)} \tag{3}$$

Then, we sum across all terms s(k) at the level of each topic k, thus constructing eleven topical indices:

$$\mathrm{CyberTopic}_{\mathrm{it}} = \frac{1}{\mathrm{N_{it}}} \sum_{\mathrm{s(k)}} \mathrm{CyberTopicTerm}_{\mathrm{it}}^{\mathrm{s(k)}} \tag{4}$$

Henceforth, for simplicity we will be referring to our eleven topical indices as $\mathrm{CyberCountry_{it}}$ (for the "Country Names" topic), $\mathrm{CyberCrypto_{it}}$ ("Crypto"), $\mathrm{CyberLoss_{it}}$ ("Monetary Loss"), $\mathrm{CyberInsurance_{it}}$ ("Insurance and Legal"), $\mathrm{CyberGlobal_{it}}$ ("Global Events"), $\mathrm{CyberPolitics_{it}}$ ("Politics"), $\mathrm{CyberSocial_{it}}$ ("Social Media"), $\mathrm{CyberPandemics_{it}}$

("Pandemics"), $\text{CyberUncertainty}_{\text{it}}$ ("Risk and Uncertainty"), $\text{CyberPosSentiment}_{\text{it}}$ ("Positive Sentiment"), and $\text{CyberNegSentiment}_{\text{it}}$ ("Negative"). The index of "Net Sentiment" $\text{CyberNetSentiment}_{\text{it}}$ is constructed as $\text{CyberPosSentiment}_{\text{it}} - \text{CyberNetSentiment}_{\text{it}}$. We will be referring to the non-normalized version of this measure as $\text{CyberNetSentiment}_{\text{it}}^{\text{T}}$, which is simply the sum of all cyber terms associated with the "Net Sentiment" topic (with similar notation for the other topics).

Summary Statistics Aggregation, unless specified otherwise, is performed by taking unweighted averages of $\text{CyberExposure}_{\text{it}}$ to the level of an industry or country. Similarly for every topical index. In the Online Appendix, Figure 20 plots the number of observations in our dataset per quarter. For all but three quarters, we have at least 2,000 observations. Starting from 2005q4 we have at least 4,000 observations per quarter. Table 4 provides summary statistics for every index by major industry, defined by the 2-digit NAICS classification that we take from Compustat. Table 5 provides summary statistics by individual country, defined by the firm headquarters location from Compustat. All numbers are reported as total counts of each measure per category, i.e. not normalized by transcript length.

3 Cyber Risk Facts and Trends

3.1 Global Patterns

We now present time-series and geographical facts on worldwide cyber risk exposure. Figure 1 presents global heatmaps for 2010, 2015, and 2020. Strikingly, the intensity of exposure has increased considerably for most countries. The most exposed regions are the United States and Canada, Europe, UK, Australia and some parts of Asia such as India, Japan and China. In Africa, we do not cover many countries but among those for which we have data the most affected one is South Africa. In Latin America, Brazil is most at risk followed by Chile and Mexico.

Figure 2 shows the regional heatmap for the United States. We observe much dispersion in both exposure and sentiment to cyber risk across individual US states. In particular, there is noticeable clustering of exposure along the east and west coast lines, with some exposure in the South as well especially in Texas, followed by Georgia and Arizona. In the Midwest only Illinois stands out as relatively exposed. The most exposed states overall are California, Virginia (which includes the District of Columbia), New York, Massachusetts and Maryland. They concentrate a large share of high-tech companies and of the IT industry. Interestingly, the two most exposed states, California and Virginia do not share the same net sentiment: California is the most pessimistic and Virginia is the most positive state regarding cyber risk. We conjecture this could be linked to the relative densities of IT security firms in the two states.

Figure 3 presents a flowmap of $\text{CyberExposure}_{\text{it}}$ by geographical origin and destination.[3] We identify destinations with the location of firm headquarters in Compustat. We distinguish between six major destinations: Africa, Americas (ex US), Asia, Europe, United States, and United Kingdom. We identify origins with the terms behind the "Country Names" topic. Specifically, we distinguish between eight geographical origins: China, Europe, Iran, Israel, North Korea, Russia, United States, and United Kingdom. For example, if country X is mentioned in proximity to some cyber term for firm i with headquarters in region Y, then we assume that an association was made between that particular form of cyber risk and country X (Y) as the origin (destination) of that risk.[4]

We find that, in absolute terms, the United States is the biggest origin of cyber risk exposure for the world economy, followed by Europe, the United Kingdom, and China. In relative terms, Russia's main "destinations" for cyber risk exposure are the US (56%) followed by Europe and the UK (taken together they account for 21%) and then Asia (17%). China's main "destinations" are the US (56%), then Asia (22%) before Europe and the UK (16%).

[3] Table 6 of the Online Appendix shows all the numbers behind the flowmap.

[4] Our procedure can only identify associations, which in turn could carry multiple contexts. We do not claim to establish the identity of "attackers".

North Korea concentrates on the US (64%) and the UK (32%). In contrast, Israel's main "destinations" are Africa (60%, excluding Israel) and Europe and the UK (20%). Iran focuses overwhelmingly on the US (94%). The United States' main "destinations" are itself (62%) and Europe and the UK almost equally and totalling a 22% share. Europe concentrates on the US (50%) and itself (27%), while the United Kingdom's main "destinations" are the US and itself, both around 35% followed by Europe (21%). Table 7 of the Online Appendix provides the origin-destination decomposition at the industry level. Industries for which the "Country Names" topic is particularly prevalent are Manufacturing, IT, and Services. The financial sector as well as services are predominantly exposed to the US and Europe as origins of exposure. China and Russia as origins of exposure are concentrated in the manufacturing and IT sectors.

3.2 Time Series Trends

Figure 4 plots the time series of global unweighted average of cyber risk exposure $1/\mathrm{M}\sum_{\mathrm{i}} \mathrm{CyberExposure}_{\mathrm{it}}$, with M the number of transcripts per quarter. The series has been normalized by the standard deviation of the entire sample. We document that average $\mathrm{CyberExposure}_{\mathrm{it}}$ has multiplied by at least a factor of six from 2002q1 to 2020q4. The increase is particularly marked after 2013. This is the intensive margin. Furthermore, the percent of all conference calls with at least one mention of cyber-related terms has increased from roughly 1% in 2002 to 5% in 2020, with also a large increase after 2013. This is the extensive margin. Overall, both the intensive and extensive margins of global exposure to cyber risk have increased between 2002 and 2020 with a larger rate of growth after 2013.

Figure 4 also highlights notable reported cyber attacks and how their timing correlates with local spikes in our index. In 2004q3, service provider AOL reported to seek legal action as BuddyLinks - a type of spyware - penetrated users' computers through instant messaging programs, collected private data, and modified software on affected machines. In 2007q4, McAfee released a Virtual Criminology Report, in which experts warned that based on all emerging statistics and trends cyber risk would become the following decade's biggest security threat. To the best of our knowledge, this was one of the first documented recognitions of cyber risk as a new source of systemic risk. In 2010q4, Tencent reported a cyber attack from a malware called "Kou Kou Bodyguard", which was allegedly developed and distributed by China.

Starting from 2013, we begin to notice an increase in the rate of growth of many of our cyber risk measures. One 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 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.

The infamous Anthem medical data breach took place in 2015q1 in which criminal hackers reportedly stole identifiable information from 75+ million medical records. In 2016, the so-called "Petya" family of encripted malware was first discovered. The malware prevented Microsoft-based Windows systems from booting and subsequently demanded payments in Bitcoin for reinstating access to the machine. The creator of Petya was eventually arrested and fined. In 2017, a new variant of Petya, which was labelled "NotPetya" by the Kaspersky lab, was used in a series of international cyber attacks, including the June 2017 attacks on Ukraine. Multiple official sources in the U.S., Canada, and Australia attributed the attacks to Russia-linked entities (Kovacs, 2018).

In 2017 we additionally witnessed two of the most infamous cyber incidents in recent memory. First, in 2017q3 the American credit bureau Equifax reported that private records of about 150 million American and 15 million British citizens were stolen. Equifax eventually agreed in a settlement with the United States Federal Trade Commission to provide affected individuals with compensation and free credit score monitoring. To this day, the Equifax breach remains one of the biggest data compromises in history. Second, in May 2017 the world was shocked by the "WannaCry" global ransomware attack. One of the most affected agencies was the National Health Services (NHS) in the UK. Tens of thousands of devices, including MRI scans and computers, were compromised by the cryptoworm (Bodkin et al., 2017). Numerous non-critical emergencies had to be canceled and ambulance visits had to be reduced. The US and the UK formally blamed North Korea for the attacks (Corera, 2017).

Finally, the Covid-19 pandemic and the Great Lockdown have been associated with an unprecedented rise in cyber attacks. World Economic Forum reported that in 2020 the number of global cyber attacks had increased by 22% (Greenberg, 2021). Moreover, the healthcare industry experienced a 45% year-on-year rise in attacks, the highest across all sectors. It is also reported that phishing attacks were 600% more common in 2020 relative to the previous year. Attacks on the Microsoft Remote Desktop protocol have also increased. The average ransomware payment in the second half of 2020 was $170,000, a reported 60% year-on-year increase (Lallie et al., 2021). The total economic toll from Covid-related cybercrime for the global economy is thus likely to be in trillions of US dollars.

We continue the presentation of results with figure 5, which compares average $\text{CyberExposure}_{\text{it}}$ with other salient indices of uncertainty and volatility. Panel (a) plots our index alongside average PRisk_{it} - the firm-level political risk measure from Hassan et al. (2019). The Pearson correlation coefficient between the two series is 0.40. Hassan et al. (2019) also construct measures of political risk based on specific topics such as "technology", "trade", "tax policy", "health care", etc. Panels (c) and (d) of Figure 5 compare average $\text{CyberExposure}_{\text{it}}$ with average $\text{PRisk(Technology)}_{\text{it}}$ and $\text{PRisk(Trade)}_{\text{it}}$ indices, with the Pearson correlation coefficients of 0.35 and 0.61, respectively. Finally, panel (b) plots the Cboe Volatility Index (VIX) whose correlation with average $\text{CyberExposure}_{\text{it}}$ is a noisy -0.12 with a p-value of (0.32).

Based on the comparison with other indices of aggregate uncertainty, we document two main observations. First, at least in the minds of conference call participants, cyber risk is generally perceived differently from fundamental drivers of political risk or market volatility.

This bolsters the view that cyber risk is a unique and new source of risk for firms. Second, trade- and technology-based sub-measures of PRisk_{it} seem to pick up similar fundamentals as $\text{CyberExposure}_{\text{it}}$, as evidenced by their relatively high correlations. However, after around 2014q1 the series diverge significantly with a sustained growth of cyber risk.

We now analyze time-series properties of our topical indices. All time-series aggregates are unweighted averages, normalized by the standard deviation of their respective samples. Figure 6 plots the average of $\text{CyberUncertainty}_{\text{it}}$. The index has been trending up over the past decade, particularly post 2013. It has spikes in 2015-2017, potentially mirroring cyber attacks on the Democratic National Convention (June and July 2016), the WannaCry ransomware attack (May 2017), and the NotPetya virus attack (June 2017). It also spikes in 2020 where the Covid-19 lockdowns have created vulnerabilities in working environments with a large increase in the use of softwares for remote work purposes. These spikes can be interpreted as a persistent negative second moment shock in the spirit of Bloom (2009). In addition to the rise in the *frequency* of idiosyncratic and aggregate cyber incidents, conference call participants have an increasingly more uncertain view of the future with regards to cyber security.

Figure 7 presents the average of $\text{CyberPolitics}_{\text{it}}$. The plot suggests that the politics topic has a strong positive association with cyber risk.[5] This relationship potentially reflects the growth of occurrences and uncertainty about state-sponsored cyber incidents. In addition, this index seems to track the US political cycle, around which cyber incidents tend to cluster.

Figure 8 plots the averages of three sentiment indices: $\text{CyberPosSentiment}_{\text{it}}$, $\text{CyberNegSentiment}_{\text{it}}$, and $\text{CyberNetSentiment}_{\text{it}}$ for net sentiment. We observe that both positive and negative sentiment towards cyber risk have increased. However, *net* sentiment has fallen roughly fourfold since 2002q1. Our latest observation - 2020q4 - seems to be a local minimum for the index, potentially coinciding with the pandemic-related spike in cyber attacks. The large negative spike around 2017q4 could be related to several major incidents:

[5] The Pearson correlation coefficient of average $\text{CyberPolitics}_{\text{it}}$ with average $\text{CyberExposure}_{\text{it}}$ over the entire sample is 0.71.

NotPetya (2017q3), the Equifax data breach (2017q3), the Uber databreach (2017q4), and the admission by Yahoo (2017q3) that a previous data breach had in fact affected an astonishing 3 billion accounts. Meanwhile, on the political side, the hacking of Deep Root Analytics in 2017q3 led to the data of 198 million US voters being stolen.

Overall, a priori it is not obvious that the rise in cyber exposure is necessarily a detrimental, negative phenomenon. It is possible that there are just as many winners from the rise of cyber risk and uncertainty as there are losers. The winners in this situation could be firms with a negative cyber risk beta: IT consulting firms, cloud security providers, etc. However, the fact that net sentiment towards cyber risk is increasingly negative suggests that the fraction of losers is large and growing, and particularly fast over the past 3 years.

Figures 9 and 10 plot the averages of $\text{CyberInsurance}_{\text{it}}$ and $\text{CyberLoss}_{\text{it}}$, respectively. Both have increased considerably, suggesting that cyber risk exposure is becoming increasingly costly to firms; these costs manifest in insurance and legal claims, monetary, and non-pecuniary (reputational) losses. Figure 11 of the Online Appendix presents the time-series average of $\text{CyberCountry}_{\text{it}}$. Panel (a) shows the total across all countries/terms and panels (b)-(f) depict individual country-specific plots. Overall, occasions when specific country names are mentioned whenever cyber risk gets discussed are increasing.

Figure 11 reveals that association between discussions of cyber risk and the "Social Media" topic has been somewhat positive and rising over the past several years. Figure 12 shows that the average of $\text{CyberCrypto}_{\text{it}}$ spiked in 2017q4-2018q2 and 2019q2. Interestingly, these episodes coincide with the two local maxima in the market price of Bitcoin. A possible explanation for this relationship is that the frequency of malware attacks correlates with the value of cryptocurrencies, which are normally the currency in which attackers demand ransom for releasing data or information systems. While the sample size is not large enough for drawing more definitive conclusions, this association is nevertheless potentially interesting and worthy of further exploration.

Figure 13 shows that the joint search for cyber and global incidents peaked in 2017, the period that corresponds to two massive worldwide cyberattacks: Wannacry and Notpetya, as well as Yahoo's disclosure of a past massive breach. Finally, Figure 14 shows that pandemics-related cyber chatter only spikes in the four quarters of 2020 - as it should.

3.3 Heterogeneity by Geography and Industry

We now provide decompositions of global cyber risk exposure by region and industry. Figure 15 decomposes our total exposure index by region, defined as the location of firms headquarters. Panel (a) in the Figure depicts the percent of all transcripts, per year, with cyber-related discussions, i.e. the extensive margin. Panel (b) shows the regional composition of all cyber risk discussions, in percent. We observe that the vast majority of cyber chatter originates in US-based firms. However, this trend has been going through a structural change since about 2014. Since then, cyber risk is becoming an increasingly global phenomenon affecting all continents, in particular Europe and the UK and Asia.

Section E of the Online Appendix provides detailed country-level time-series data for the US, UK, France, Germany, Japan, China, India, Canada, Israel, Russia, Spain, and Italy.[6]

Figure 16 decomposes $\text{CyberExposure}_{\text{it}}$ by industry, proxied by two-digit NAICS codes. Panel (a) reports the fraction of all transcripts with non-zero cyber risk related discussions and panel (b) 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.

Table 17 offers a more granular look at the financial industry. The breakdown of global cyber exposure based on 4-digit NAICS codes appears to be broadly in terms of 40% for

[6] Additional countries are available upon request.

financial intermediaries, 40% for insurance companies, 10% for broker-dealers, and 10% for all the rest. Interestingly, the insurance sector has been increasingly exposed to cyber risk 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-bank intermediaries.

4 Validation

In this section we validate our text-based measures of cyber risk. It is important to confirm that our indices are informative about the dangers of actual cyber attacks. We therefore manually merge earnings call announcement data with the Privacy Rights Clearinghouse (PRC) database on reported cyber incidents. 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) between firm names in the two datasets. We impose a reasonable cutoff for good and bad matches.

We find that out of the roughly 9000 non-public, non-governmental, and non-medical cyber attacks in the PRC dataset (the latest available quarter was 2019q4), 2600+ unique firm-incident pairs can be matched to the earnings call data with the average cosine distance of 0.75, with 1 indicating a perfect match. As a final step, we identify some of the most salient realized cyberattacks in recent memory and tag them in both datasets. The goal of this exercise is to check whether our textual algorithms indeed pick up economically meaningful chatter when we know that an actual attack is taking place.

Section A of the Online Appendix presents earnings call text snippets and our commentary for 8 major known and reported past cyber incidents. For each event, we provide the company name, the date of the closest conference call, values for $\text{CyberExposure}_{\text{it}}$ and $\text{CyberNetSentiment}_{\text{it}}$, a precise excerpt from the conference call texts, and additional commentary. Overall, we find that these widely known realized cyber attacks are picked up very well by our indices. For example, the 2017 Equifax data breach has an $\text{CyberExposure}_{\text{it}}$ ($\text{CyberNetSentiment}_{\text{it}}$) score of 10 (-10), implying large exposure and very negative market sentiment. Another example is the 2014 Target data breach, when over 100 million individuals lost sensitive information including credit card account data. This event corresponds to an $\text{CyberExposure}_{\text{it}}$ ($\text{CyberNetSentiment}_{\text{it}}$) value of 19 (-19): both are outliers in the overall distribution.

Of course, not every spike in our measures must be associated with an actual cyber incident. Most of the time, firm executives and other call participants express concerns about *potential* events, which may or may not ever materialize.[7] However, this does not take away the fact that mere anticipation of an attack could significantly impact present-day franchise and market values and economic decision-making. In Section 5 we ask whether we *can* predict actual future cyber attacks using our text-based measures.

5 Predicting Future Cyberattacks

Can our text-based firm-level measures of cyber risk predict actual future cyberattacks on firms? In order to answer this question we use the merged datasets of PRC's reported cyber-attack data and quarterly earnings calls. Our left-hand-side is an indicator variable which takes on the value of 1 if a firm suffers a cyberattack at any time within the next 8 quarters. Our right-hand-side variables include a cyber risk measure and additional firm controls: intangible assets / total assets, capex spending / total assets, PP&E spending / total assets, cash flow / assets, long-term debt / total debt, log(total assets), total debt /

[7] In spirit, our paper is thus related to the literature on the impact of uncertainty shocks (Bloom, 2009; Bloom et al., 2018)

total assets, liquidity, log of age, ROA, equity issuance, turnover ratio, credit rating, Tobin's Q, book-to-market ratio, operational costs / assets, and the market beta.[8] In addition, we include year and industry fixed effects and cluster standard errors at the firm level. All right-hand-side variables except the cyber risk indices are normalized by their standard deviations. Because the PRC data covers predominantly cyberattacks that were reported on U.S. soil, we restrict our sample to U.S. firms.

Results from logit regressions are reported in Table 1. We see that if a firm has positive cyber risk exposure in quarter t, then it is much more likely to report a cyberattack within the next 8 quarters. In terms of economic significance, a one-unit increase in $\text{CyberExposure}_{\text{it}}$ raises the probability of a cyberattack by 6.6-8%. Relative to the baseline exposure index, an attack is *more* likely if cyber-related chatter is of negative net sentiment or includes terms from the $\text{CyberInsurance}_{\text{it}}$ and $\text{CyberLoss}_{\text{it}}$ topics.

Does the intensive margin of cyber risk exposure matter? Figure 18 shows the margins plot from the baseline logit regression of $\text{CyberExposure}_{\text{it}}$ on the cyberattack indicator, inclusive of all controls and fixed effects. We see that the intensity of cyber risk exposure does matter for forecasting future attacks. The probability of an attack ranges from roughly 1.3% for $<$5 cyber risk mentions to over 20% for ≥ 60 mentions. The relationship remains statistically significant at the 5% level for ≤ 20 mentions and at the 10% level for ≤ 30 mentions. Estimates are understandably noisier for the very right tail of the exposure distribution since the frequency of such incidents (transcripts with $\text{CyberExposure}_{\text{it}} \geq 30$) is low.

Overall, our results suggest that time-varying text-based measures of exposure to cyber risk can be viewed as robust early warning indicators for actual future cyber attacks.

[8] Table 3 of the Online Appendix provides a detailed definition and source for each variable used.

6 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. Table 3 in the Online Appendix provides detailed definitions of every variable that we use. Our main model is a probit regression of firm characteristics on indicators of cyber risk exposure. An indicator takes the value of 1 if a transcript records $\text{CyberExposure}_{\text{it}} > 0$. The same exercise is run on all of our topical indices.

Results are reported in Table 2. 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 ratio, and large size (as mentioned by total assets). For most of our topical indices, we see that these three firm characteristics are the most robust predictors of exposure. This implies that either large firms with lots of intangibles and liquidity *worry* about being cyber attacked (as seen from column (4) on CyberNegSentiment) and that the attackers target those firms more than the average firm; or that the business of a subset of these particular firms potentially benefits from cyber risk (since the CyberPosSentiment is also strongly related to those characteristics as seen in column (3)). In terms of explanatory power, we see that the pseudo-R^2 of our regressions is just about 0.2; a large fraction of cyber risk exposure is left puzzlingly unexplained.

Table 8 of the Online Appendix shows heterogeneity by region. Regions are defined based on the location of firm headquarters, taken from Compusat. For most areas, size or age are positively correlated with cyber risk exposure. The best predictors of exposure in the US seem to be intangible assets, size, liquidity and cash flow. In the Americas (ex US), the best predictors are size and Tobin's Q. In Europe: intangible assets and equity issuance. In the UK: debt maturity ratio, size, age, the market beta, and book-to-market ratio. In Asia: PP%E expenditures, age, and equity issuance. Finally, in Africa: CAPEX over assets, age, market beta, and debt to assets ratio. Some reasons for this heterogeneity can be traced back to sectoral composition, to which we now turn.

Table 9 of the Online Appendix shows heterogeneity by industry. Industries are defined based on 2-digit NAICS codes. The highest number of firms is in manufacturing where the liquidity ratio, size, leverage and age are the main determinants of cyber-risk exposure. The most exposed sectors to cyber risk tend to be IT and services. For those, exposure increases significantly with liquidity ratios and size and intangible assets (for services). For the financial sector, which is increasingly exposed to cyber risk, determinants are intangibles, size and Tobin's Q. In trade, size, cash flow, equity net issuance, return on assets and PP&E expenditures correlate with exposure.

Table 10 of the Online Appendix shows heterogeneity within the financial sector. Finance sub-sectors are defined based on 4-digit NAICS codes. The three largest categories are banks, insurance and broker dealers. For banks (regular depository financial intermediaries), the best predictors of cyber risk exposure are PP&E expenditure, size, and book-to-market equity; larger institutions seem more at risk. Insurance companies are important targets and their cyber exposure is positively correlated with intangibles, age, and Tobin's Q but negatively correlated with their CAPEX to assets ratio. Interestingly, for broker-dealers, some correlates are similar to banks and insurance (size, age) while others also appear but with an opposite sign (intangibles, PP&E). Two new variables seem very important: turnover (positively correlated) and operational costs over assets (negatively correlated).

7 Firm-Level Cyber Risk and Stock Returns

7.1 Direct Effects

In this section we explore asset pricing implications of our measures of cyber risk. We employ an event study approach and focus on stock market outcomes of affected, peer, and non-peer firms. First, we identify events in terms of calendar weeks w during which earnings call announcements recorded positive cyber risk exposure $\text{CyberExposure}_{\text{iw}}^{\text{T}}$. We discard weeks with no exposure. Second, in order to build our main shock variable, we take the sum of $\text{CyberExposure}_{\text{iw}}^{\text{T}}$ across firms in a given event week w, country (firm headquarters), and a tightly defined industry (6-digit NAICS code). Our first dependent variable is the weekly average of day-to-day stock returns for the affected firms during event weeks. Affected firms are defined as those whose earning call transcripts recorded $\text{CyberExposure}_{\text{iw}}^{\text{T}} > 0$. All our specifications include country, industry, and week fixed effects. We also control for firm size, proxied by stock market valuation.

We begin the discussion of results with Table 3. In columns (1)-(2) we focus on equally and value-weighted stock returns of affected firms for weeks with positive $\text{CyberExposure}_{\text{iw}}^{\text{T}}$. We find that cyber risk exposure carries negative and significant effects on returns of affected firms. Economic significance of the estimates is large: *each* additional mention of cyber risk terms reduces weekly returns by around 4.3 basis points. By extension, more than 23 mentions cause losses equalling roughly 1 percentage point.

7.2 Contagion Effects

We now move beyond estimating direct effects on affected firms and ask whether there are *contagion* effects from firm-level cyber risk exposure. To that end, we define *peer* firms as those which belong to the same industry and country as affected firms but recorded $\text{CyberExposure}_{\text{iw}}^{\text{T}} = 0$ during the event weeks. Industries are still defined at a very granular 6-digit NAICS level. Countries are still defined as firms' headquarter locations, taken from Compustat. Our new dependent variable is now the weekly average of day-to-day returns of peers for the weeks when affected firms experienced $\text{CyberExposure}_{\text{iw}}^{\text{T}} > 0$.

Results of contagion regressions are reported in columns (3)-(4) of Table 3. We find that cyber risk exposure has a statistically significant negative effect on unaffected peer firms. In other words, idiosyncratic exposure to cyber risk has the potential of propagating to corporate peers in financial markets and to cause "systemic" events. Economically, one additional cyber term decreases returns of peer firms by 2.76 bps. Roughly 36 mentions or more cause losses equalling 1 percentage point. These results are also robust to the presence of country, industry, and week fixed effects. We include valuations of both affected and peer firms as additional controls.

In order to nail down the precision of our definition of peer firms, we now check whether cyber risk exposure has any impact on non-peers. We define non-peers as firms that are from the same location but not from the same industry as affected firms. We continue to define industries and countries as 6-digit NAICS codes and headquarter locations, respectively. Our left-hand-side variable is now weekly averages of day-to-day stock returns of all non-peer firms for the same event weeks as before. Results are shown in columns (5)-(6) of Table 3. Reassuringly, we find a noisy zero effect of cyber risk exposure on non-peers, suggesting that belonging to the same tightly defined industry cluster is a good proxy for financial market connectedness.

So far our analysis has included only our baseline measure $\text{CyberExposure}^{\text{T}}_{\text{iw}}$ as the main covariate. We now look at our topical indices as potential causes of direct and contagion stock market effects. We focus on the $\text{CyberUncertainty}_{\text{iw}}$, $\text{CyberNegSentiment}_{\text{iw}}$, $\text{CyberInsurance}_{\text{iw}}$, and $\text{CyberLoss}_{\text{iw}}$ topics. For each topical index, we compute weekly totals for weeks surrounding earnings calls with positive topical exposure. As before, our dependent variables are weekly averages of day-to-day stock returns for affected, peer, and non-affected firms, each defined accordingly.

Table 4 reports the results. Each topical index has a negative and significant effect on affected firms' performance, as can be seen from columns (1)-(4). The economic effect is more than 50% higher for the "Negative Sentiment", "Insurance and Legal" and "Monetary Loss" topics than for the baseline exposure measure. In other words, when cyber risk chatter is pessimistic or also includes discussions of potential material economic loss or insurance and legal claims, the stock market impact becomes more severe. Columns (5)-(8) present the results for peer firms. All coefficients are negative and statistically significant. Relative to the baseline total exposure measure, they are also economically more potent. For example, discussions of insurance and legal topics increase the impact on peers by 100% (-5.4 bps) relative to the baseline estimate in Table 3 (-2.7 bps). Finally, columns (9)-(12) report results for the non-peers. There are no economically or statistically significant effects of topical cyber risk exposure on non-connected, non-peer firms.

The average ratio of the number of peers to the number of affected firms, per country and industry, is 23.17 with a standard deviation of 37 and the 1st and 99th percentiles equal to 1 and 234, respectively. This implies that for every one firm with $\text{CyberExposure}_{\text{iw}} > 0$ roughly 23 unaffected peer firms get exposed to cyber risk through financial market linkages. Because the distribution of peer-to-affected firm ratios is very right-skewed, the scope for downside risk is potentially significant.

To sum up, results of this section suggest that "cyber risk shocks" can cause significant stock market disruptions for the affected firms. Moreover, idiosyncratic cyber risk can propagate through the network of peers in stock markets and cause ripple effects across the financial system. We consider this finding as prima facie evidence for the *systemic* nature of firm-level cyber risk.

8 Factor Structure in Cyber Risk and the Cross-Section of Stock Returns

In this section we show that there is a factor structure in firm-level measures of cyber risk exposure. Common cyber risk exposure is a *factor* that can help price the cross-section of asset prices. We conjecture that firms that covary more negatively with the cyber risk latent factor are more likely to experience higher excess returns, on average. That is, firms that earn less in states of the world where aggregate cyber risk is high (i.e. low cyber risk beta stocks) must demand higher excess returns in equilibrium. The economic cause of this mechanism is that (a) some firms are fundamentally more reliant on business technologies that are more prone to data breaches and malware attacks and (b) some firms are individually less able to insure against idiosyncratic or aggregate cyber attacks with operational risk capital.

8.1 CyberE Pricing Factor

We begin by constructing our main pricing factor. We focus on cyber risk total exposure, which is our main baseline measure. First, we compute the monthly average of $\text{CyberExposure}^{\text{T}}_{\text{it}}$.[9] Second, we fit an AR(1) model onto the time series and extract residuals. This ensures that we capture *shocks* to cyber risk. Finally, we standardize these residuals across the full sample, i.e. we subtract the mean and divide by the standard deviation. Figure 19 plots the resulting time series of our factor which we label CyberE (Cyber Exposure).

8.2 Portfolio Sorting

Our first asset pricing test involves first running trailing-window 30-month time-series regressions of firm-level excess returns on the CyberE factor and a constant. We require at least 30 observations in these regressions. We extract the distribution of firm-level CyberE betas and truncate it at the 1% and 99% levels. For each month, we perform a one-way beta sort with either equal- or value-weighted average excess returns. Five portfolios are formed and held for one month. Table 5 reports the results. We see that average portfolio returns are decreasing in CyberE betas. Stocks in the first CyberE quantile have negative CyberE betas and thus on average lose value when aggregate cyber risk exposure is high. In contrast, stocks in the highest quantile hedge CyberE growth, paying off precisely when high cyber risk states realize. The long-short portfolio built on CyberE beta sorted stocks pays an average of -3.3% in returns per year. This excess returns spread cannot be readily explained away by the market, size and book-to-market (Fama-French), and momentum pricing factors, as evidenced by model alphas that are large and significantly different from zero.

[9] Whether we normalize by transcript length or not doesn't affect the results.

8.3 Fama-MacBeth and the Price of Cyber Risk

Our second asset pricing test is the Fama and MacBeth (1973) exercise. We formally test whether the CyberE factor can explain abnormal returns of CyberE-beta sorted portfolios. We repeat the time-series step from before and now form 10 cyber beta single-sorted portfolios. In the second stage, we run cross-sectional regressions of average excess returns of the 10 test assets on factor betas plus a constant. A good model features a sizable R^2, small constants, and small pricing errors (mean average pricing error, or MAPE).

Table 6 reports the results from the cross-sectional estimation step. Columns (1) and (5) show that the CAPM cannot price CyberE-beta sorted portfolios. R^2 are low and MAPEs are very large. When we introduce our factor in columns (2) and (6) we see that R^2 increases by 65+ percentage points. Pricing errors fall substantially. For example, the value-weighted CyberE specification exhibits MAPE of just 50 basis points per year when just CyberE and market factors are used in the estimation. The R^2 of that specification is 0.74. We also highlight that point estimates for the price of risk are always positive and statistically significant at the 5% level.

In columns (3) and (7) we add the Fama-French size and book-to-market factors and observe that our factor is still economically and statistically significant. MAPEs drop to as low as 25.5bps per year and the R-squared is 94%. Finally, in columns (4) and (8) we add the momentum factor.[10] The price of cyber risk is still positive and statistically significant. The R^2 reaches 95.8% and the MAPE drops to 22 bps per year.

We conclude this section by emphasizing that there appears to be a strong factor structure in firm-level exposure to cyber risk. In addition to the findings of Section 7.2 on contagion effects, this is our second evidence in favor of cyber risk being a source of *systemic* risk for financial markets.

[10] We obtain the time-series for the momentum factor from the Kenneth French online data repository.

9 Determinants of Cybersecurity ETF Prices

Our last exercise is to try and explain monthly movements in returns on some of the most popular cybersecurity ETFs. We focus on the ETFMG Prime Cyber Security ETF (HACK) and the First Trust NASDAQ Cybersecurity ETF (CIBR). We construct monthly returns on the full available time-series of both indices and correlate them with our measures of firm-level cyber risk exposure and the Fama-French three factors. Our goal is to understand if the market price of cybersecurity ETFs reflects the *fundamentals* of cyber risk. Furthermore, we wonder whether the two instruments could act as a hedge against cyber risk shocks.

Figure 12 in the Online Appendix plots monthly returns on the S&P500 and on the two cybersecurity ETFs. The three series are hardly quantitatively distinguishable; the Pearson correlation coefficients between the HACK and CIBR ETFs with the market are 0.63 and 0.64 with p-values of (0.00) and (0.00), respectively. Table 11 in the Online Appendix presents the full correlation heatmatrix. The two ETFs are correlated with the size factor but, surprisingly, not with our text-based measures. Among our topical indices, the best correlates are $\text{CyberPolitics}_{\text{it}}$, $\text{CyberInsurance}_{\text{it}}$, and $\text{CyberCrypto}_{\text{it}}$. But the absolute values of these correlation coefficients are low and they are all smaller than 25% (as a comparison point, the coefficient with the size factor is about 48%[11]. Interestingly, HACK and CIBR ETFs do not appear to price the information contained in our baseline $\text{CyberExposure}_{\text{it}}$ measure.

We conclude that neither HACK nor CIBR cybersecurity ETF price movements can be readily explained by microeconomic, text-based fundamental measures of cyber risk exposure. This finding suggests that these instruments mostly reflects aggregate market returns, are potentially mispriced and cannot not credibly serve as a hedge against cyber risk shocks.

[11] Recall that in our firm level panel regressions, the size of firms is a positive correlate of our cyber exposure variable albeit one with a low explanatory power.

10 Discussion: Why Economics of Cybersecurity?

In this section we supplement our empirical analysis with a general-interest discussion on the importance of research on cyber security going forward.

Among the Most Challenging Risks for Firms

In recent surveys of financial market participants, firms consistently rank cyber risk as second only to political or climate risk. According to the "Systemic Risk Survey" (SRS) of the Bank of England, cyber threat is the second most cited risk in the UK financial system (BoE, 2020). The SRS is conducted by the BoE on a biannual basis to estimate and track market participants' views of risks to financial stability and resilience. The 2020H1 survey cites cyber uncertainty as the fastest rising form of risk. Roughly 50% of respondents currently view cyberattacks as one of the most challenging risks for the management of a firm.

A New Type of Disasters

The growing number of cyber breaches could complement the empirical disaster risk literature. An important feature of cyber attacks is targeted ill will and presence of a malignant attacker/criminal. This feature is shared with other forms of acts of terror such as physical terrorism and related disasters (Kashyap and Wetherilt, 2019). In this sense, cyber risk could be treated as a new type of disaster risk (Rietz (1988) and Barro (2006)) - aggregate or sectoral. Equilibrium models with disasters help to rationalize some financial and macroeconomic puzzles and facts.[12] An important difficulty of the empirical disaster risk literature is the relatively low number of actual international disasters. The direct effect of disasters, as a result, potentially underestimates its expectational impact on risk premia. Cyber risk exposure and realized incidents are likely to continue their rapid growth over the next decades. To the extent that cyber risk exposure causes direct economic and financial

[12] Gabaix (2012) and Gourio (2012) are important contributions to disasters in business cycle research. See Wachter and Tsai (2015) for a review of recent work on disaster risk in financial economics.

harm, a new generation of models with disaster risk could be calibrated with estimates of cyber disaster risk and uncertainty.

Both Idiosyncratic and Aggregate Shocks

Cyber attacks can be viewed as proxies of negative idiosyncratic or aggregate shocks to financial net worth. In February 2016, it was reported that 30+ fraudulent instructions were issued by cyber criminals via the SWIFT network to transfer over $1 billion from the Federal Reserve Bank of New York accounts belonging to the Central Bank of Bangladesh (Gopalakrishnan and Mogato, 2016). Over $50 million were successfully transferred and never recovered, while the rest was either recovered or prevented via inspections and monitoring. The "Bangladesh Bank Robbery" is a peculiar form of a negative shock to bank net worth. Because it involved only a single institutional victim, was essentially unforecastable, and did not trigger mass second-order spillover effects on the Bangladeshi economy, we can label this event as an uninsurable idiosyncratic negative net worth shock.

In May 2017, a worldwide cyberattack took place that targeted hundreds of thousands of computers in 150 countries (Bodkin et al., 2017). Attackers paralyzed computers with the WannaCry ransomware cryptoworm that took private data hostage and demanded a timely ransom for data release. Hundreds of millions of dollars in damages were lost due to this event. Because the attack was international in nature, affected multiple institutions simultaneously, this incident can be categorized as an aggregate negative net worth shock.

In May of 2018, the Federal Bureau of Investigation (FBI) apparently warned banks of an imminent large-scale operation attempting to empty ATMs of their holdings, a coordinated cyber crime that would cause large losses for financial institutions (Kirk, 2018). Once penetrated into banks' financial systems, attackers can install malware that removes limits on payment card accounts and modifies internal ATM systems. Cyber criminals proceed with payment card cloning using data from compromised point-of-sales. Stolen data is then encoded into magnetic stripes on the backs of credit cards. Massive international coordinated cashouts can trigger systemic bank runs. Small-scale attacks have already taken place in Japan, South Africa, and Turkey.

Going forward, the frequency of both idiosyncratic and aggregate cyber shocks to net worth may well outnumber the more traditional firm-level or global operational risks like factory malfunctions, accounting scandals, weather disasters, etc.

An Automation-Cyber Risk Trade-Off

If we believe in automation as a trend, then the size and persistence of negative cyber shocks is likely to grow in the future. Autor et al. (2003), Acemoglu and Restrepo (2017), Martinez (2021) among others study the intricate role that automation will play in the evoluation of aggregate factor shares and labour income in the 21st century. Trend growth of automation and its net impact on the number and nature of distinct labour tasks as well as on productivity is likely to be very substantial.

On one hand, trend growth in automation can be viewed as a persistent positive productivity shock. On the other hand, advances in digital technology create new avenues for malicious attackers, thieves, and cyber criminals to exploit the digital architecture and disrupt proper functioning of financial transactions. For example, the emergence of cloud technologies enables firms to store and access huge amounts of data that is crucial for their operation. However, cyber attacks on security protocols of firms that manage such cloud technologies (internally or through outsourcing) is a source of systemic risk - all firms that are linked to the same cloud provider could be affected. Quantitatively, the cost of cyber risk should be taken into count in models of automation-driven growth. In addition, Moll et al. (2021) argue that new technologies accrue to the owners of capital in the form of higher capital incomes, which feeds inequality. Disruptions to automation, caused by cyber risk exposure or actual cyber attacks, could thus have immediate implications for income and wealth inequality.

Cyber Networks

Digital innovation and the emergence of Fintech as a form of payment creates a new channel for the escalation and propagation of cyber attacks. On June 20, 2019 the Bank of England announced that it intended to allow financial technology companies access to bilateral payment systems on a level playing field with major commercial banks (Giles and Binham, 2019). Technically, new payment providers such as Facebook would be allowed to store funds overnight in interest-bearing accounts at the Central Bank. This would facilitate the spread and adoption of Diem (formerly Libra), Facebook's digital currency, as a novel form of payment. The end goal of this move would be to enable the BoE to regulate the fintech sector more efficiently.

This, of course, would make fintech firms a target for cyber criminals. Traditionally, commercial banks have been a routine target of cyber attacks and even some Central Banks have been hit. In the Bank of Bangladesh heist of 2016, criminals stole over $50+ million via fraudulent transfers from the Federal Reserve Bank of New York accounts belonging to BoB (Gopalakrishnan and Mogato, 2016). But commercial banks are now more heavily invested in IT security and are, generally, better prepared for such cyber incidents both in terms of security management and regulatory liquidity buffers (Duffie and Younger, 2019). It remains to be seen how fintech companies will withstand future cyber threats, because access to central banking vaults would put a bounty target on their backs.

11 Conclusion

We build novel text-based firm-level measures of cyber risk exposure. We classify content of cyber risk discussions based on various topics of interest such as monetary loss, legal and insurance claims, sentiment, etc. Our approach leverages state-of-the-art techniques from computational linguistics and identifies cyber risk related textual bigrams in the texts of quarterly corporate earnings call announcements. We document an important increase of exposure to cyber risk around the world, in particular post 2013, affecting in relative terms more and more Europe and Asia. There are interesting correlates, worthy of further investigation, between cyber risk intensity and political cycles or the value of crypto currencies. While still mainly hitting the IT and services sectors, our evidence suggests that cyber risk is spreading towards the financial sector, in particular insurance companies. We also find that firms that are large or are older, have a high ratio of intangible assets, and lots of liquidity are more likely to be exposed to cyber risk than the median firm.

We validate our measure by cross-referencing with databases containing reported cyber attacks and show that our text-based measures can predict future realized cyber attacks. Exposure to cyber risk has an economically and statistically significant negative effect on stock market performance of affected firms. Moreover, we find strong evidence of contagion effects - we trace out the impact on firms that did not discuss anything related to cyber risk but are in the same country and industry as the affected firm. Idiosyncratic firm-level cyber risk thus has the potential of spreading through interconnected financial markets.

There is a factor structure in firm-level discussions of and references to cyber risk. We construct a new pricing factor - CyberE - which is based on our text-based cyber risk exposure measure. Our factor can help price the cross-section of stock returns. Firms that are more sensitive to spikes in aggregate cyber risk, proxied by our factor, require equilibrium compensation via higher excess returns. Finally we find that existing cyber ETFs reflect much more market risk and the conventional size factor than exposure to cyber risk.

References

Acemoglu, D. and P. Restrepo (2017): “Robots and Jobs: Evidence from US Labor Markets,” *NBER Working Paper*, 23285.

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

Autor, D., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 118(4).

Barro, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3).

Biener, C., M. Eing, and J. H. Wirfs (2015): "Insurability of Cyber Risk: An Empirical Analysis," *Working Paper*, 151.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, 86(3).

Bodkin, H., B. Henderson, L. Donnelly, and R. Mendick (2017): "Government under pressure after NHS crippled in global cyber attack as weekend of chaos looms," *The Telegraph*.

BoE (2020): "Bank of England Systemic Risk Survey," .

Corera, G. (2017): "Cyber-attack: US and UK blame North Korea for WannaCry," *BBC*, December 19.

Crosignani, M., M. Macchiavelli, and A. F. Silva (2020): "Pirates without Borders: The Propagation of Cyberattacks through Firms' Supply Chains," *Federal Reserve Bank of New York staff report No 937*.

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

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

Eisenbach, T., A. Kovner, and M. J. Lee (2020): "Cyber Risk and the U.S. Financial System: A Pre-Mortem Analysis," *Federal Reserve Bank of New York Staff report*, 909.

ESRB (2020): "Systemic Cyber Risk," *Report*, February.

Fama, E. and K. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1).

Fama, E. and J. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3).

Florakis, C., C. Louca, R. Michaely, and M. Weber (2020): "Cybersecurity Risk," *NBER Working Paper 28196.*

Gabaix, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127(2).

Gentzkow, M., B. T. Kelly, and M. Taddy (2021): "Text as Data," *Journal of Economic Literature*, Forthcoming.

Giles, C. and C. Binham (2019): "BoE to grant tech companies access to overnight accounts," *Financial Times.*

Gopalakrishnan, R. and M. Mogato (2016): "Bangladesh Bank official's computer was hacked to carry out $81 million heist," *Reuters.*

Gourio, F. (2012): "Disaster Risk and Business Cycles," *American Economic Review*, 102(6).

Greenberg, I. (2021): "Fifth-generation cyberattacks are here. How can the IT industry adapt?" *WEF*, February 12.

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

——— (2020a): "Firm-Level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1," *Working Paper.*

——— (2020b): "The Global Impact of Brexit Uncertainty," *Working Paper.*

Kamiya, S., J. Kang, J. Kim, A. Milidonis, and R. Stulz (2020): “Risk Management, Firm Reputation, and the Impact of Successful Cyberattacks on Target Firms,” *Journal of Financial Economics*, Forthcoming.

Kashyap, A. and A. Wetherilt (2019): “Some Principles for Regulating Cyber Risk,” *AEA Papers and Proceedings*, 109, 482–487.

Kirk, J. (2018): “FBI Warns Of Pending Large Scale ATM Cashout Strike,” *Bankinfosecurity.*

Kovacs, E. (2018): “U.S., Canada, Australia Attribute NotPetya Attack to Russia,” *Security Week*, February 16.

Lallie, H. S., L. A. Shepherd, J. R. Nurse, A. Erola, G. Epiphaniou, C. Maple, and X. Bellekens (2021): “Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic,” *Computers Security*, 105.

Martinez, J. (2021): “Putty Clay Automation,” *CEPR Discussion Paper 16022.*

Moll, B., L. Rachel, and P. Restrepo (2021): “Uneven Growth: Automation’s Impact on Income and Wealth Inequality,” 28440.

ORX (2020): “2020 Annual Banking and Insurance Operational Loss Reports,” .

Rietz, T. (1988): “The equity risk premium: a solution,” *Journal of Monetary Economics*, 22(1).

Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang (2020): “Firm-level climate change exposure,” *SSRN Working Paper*, 3642508.

Tosun, O. K. (2019): “Cyber Attacks and Stock Market Activity,” *Working Paper.*

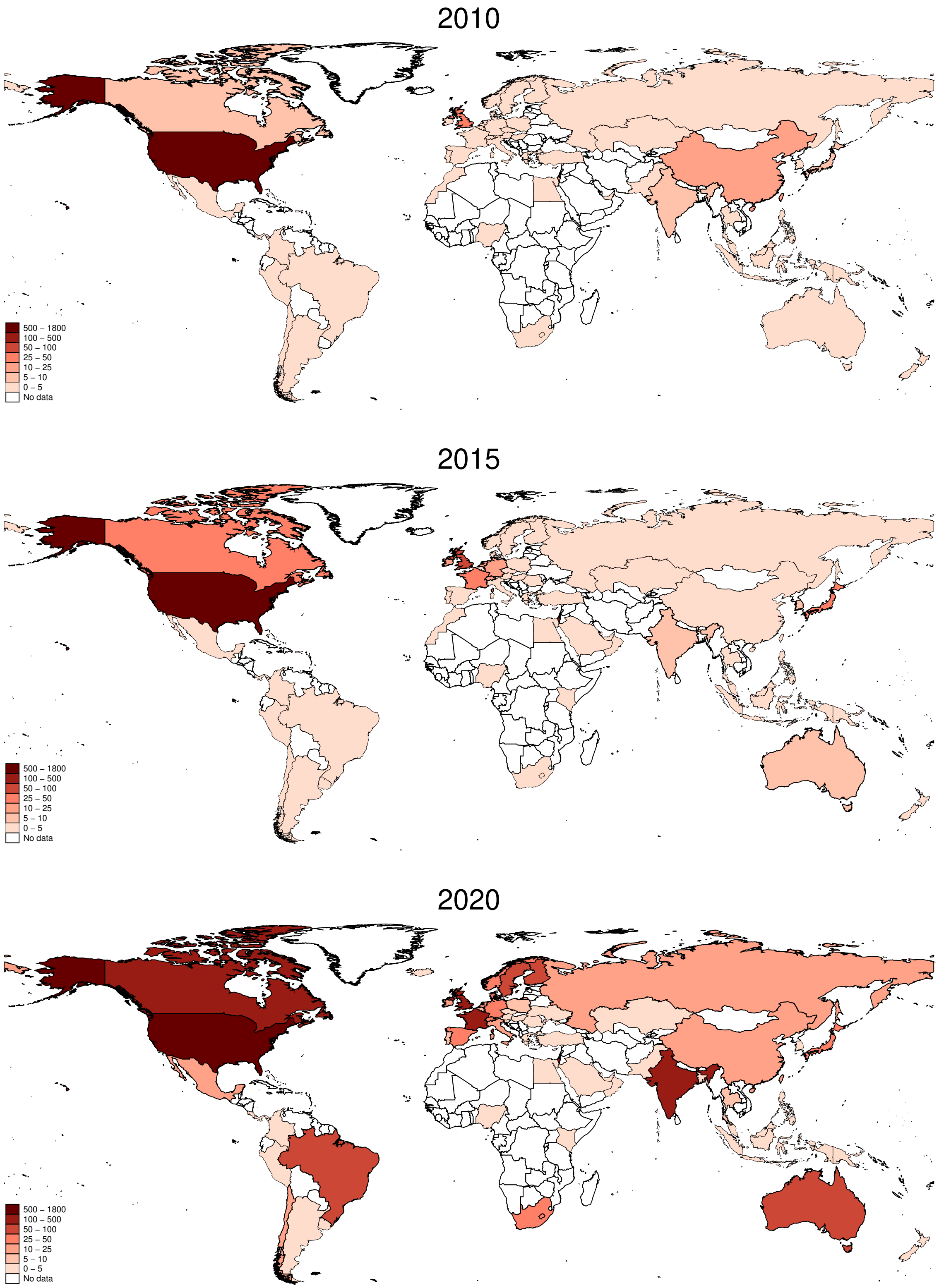
Wachter, J. and J. Tsai (2015): "Disaster Risk and Its Implications for Asset Pricing," *Annual Review of Financial Economics*, 7.

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

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

Figures and Tables

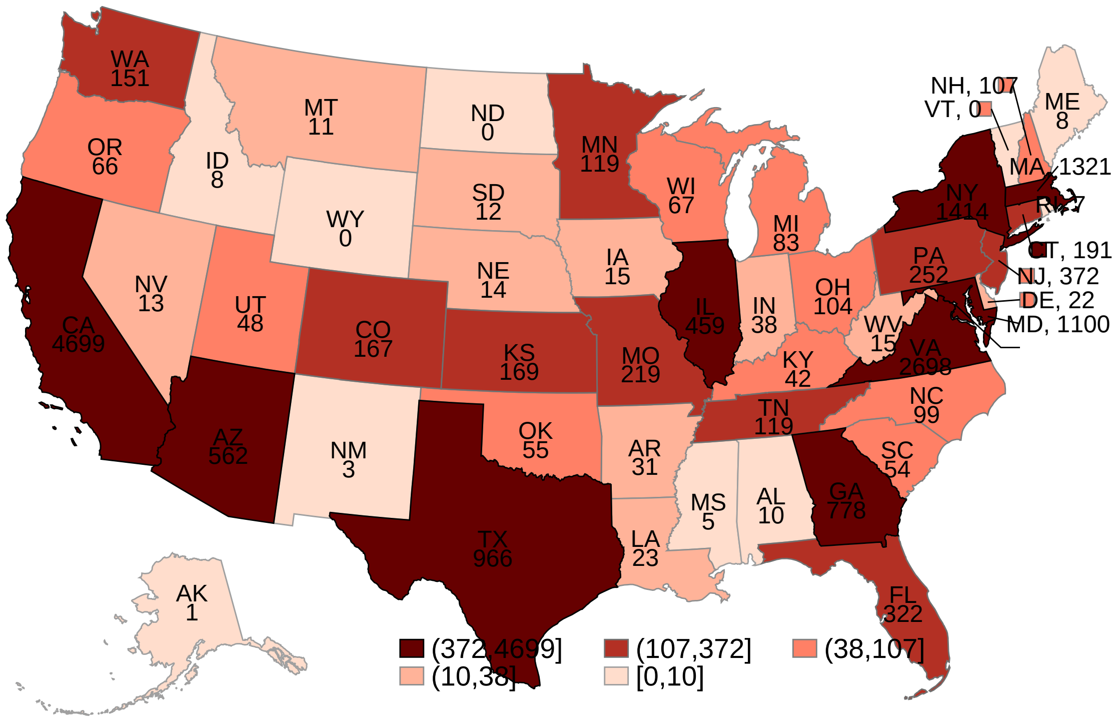
Figure 1: **Global Heatmap of Cyber Risk Exposure**



Notes: Country-level totals of CyberExposure$^{\mathrm{T}}_{\mathrm{it}}$ for different years. Higher values correspond to darker shades of brown. The legends show five ranges for values of CyberExposure$^{\mathrm{T}}_{\mathrm{it}}$ and their corresponding colors on the map.

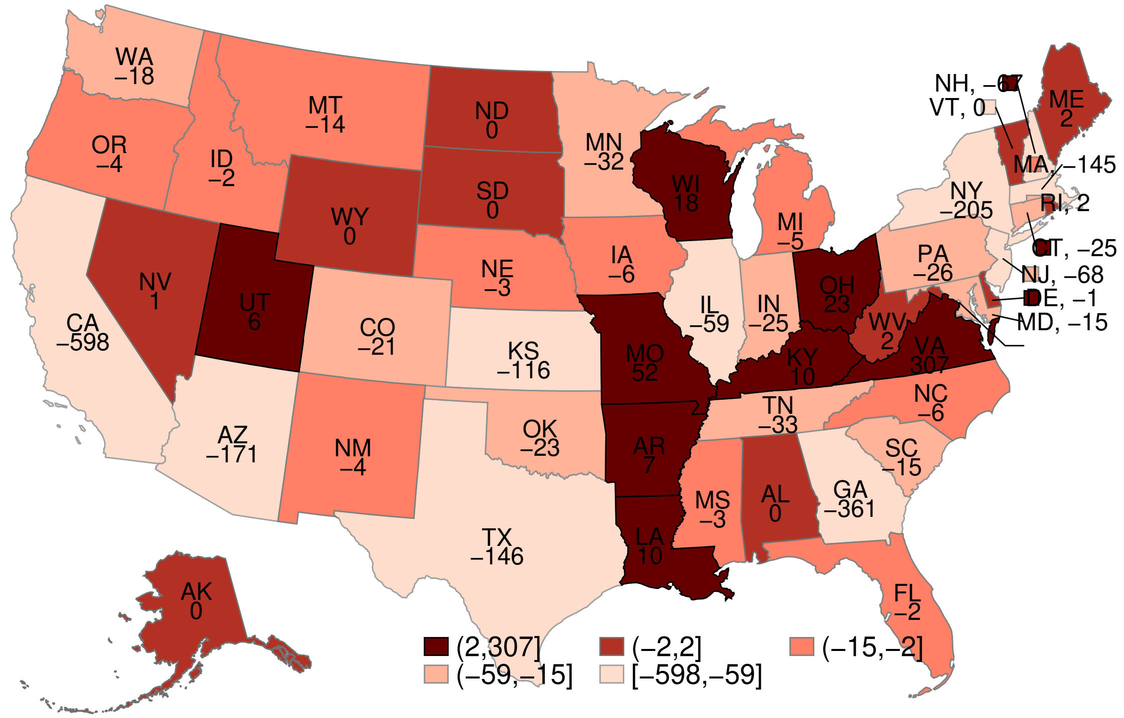
Figure 2: **US Heatmaps of Cyber Risk Exposure and Sentiment**

(a) **Total Exposure**



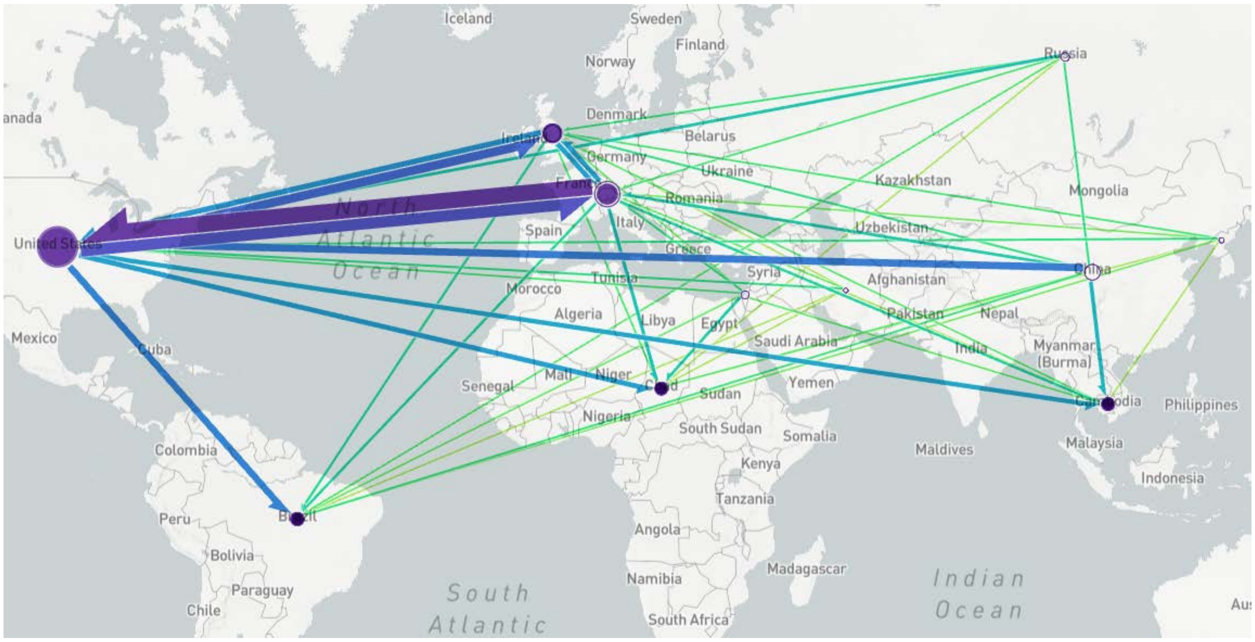
Notes: State-level totals of $\text{CyberExposure}_{\text{it}}^{\text{T}}$ over 2002q1-2020q4. Higher values correspond to darker shades of brown. The legend shows five ranges for values of $\text{CyberExposure}_{\text{it}}^{\text{T}}$ and their corresponding colors on the map.

(b) **Net Sentiment**



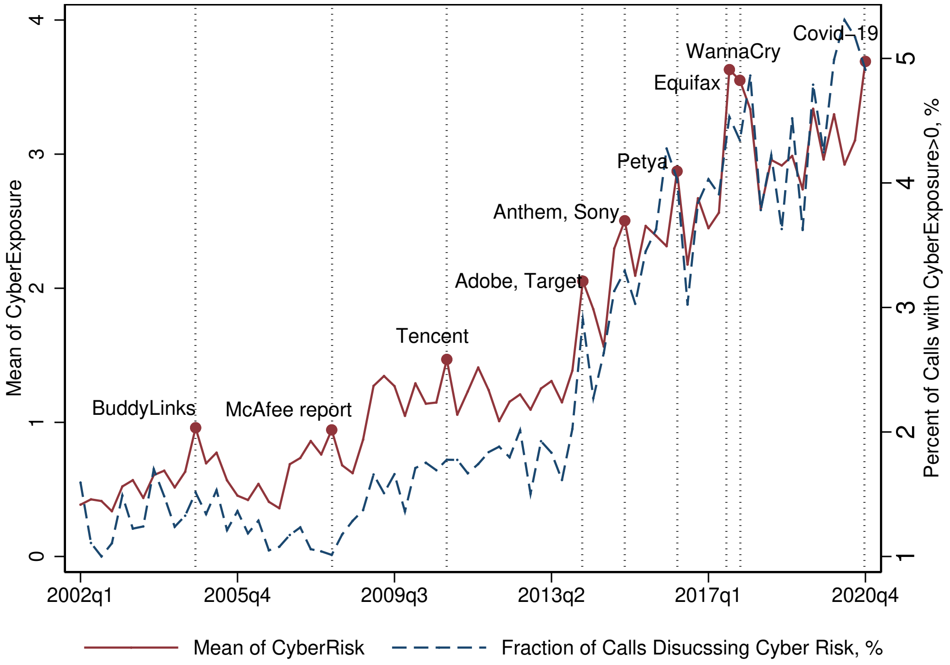
Notes: State-level totals of $\text{CyberNetSentiment}_{\text{it}}^{\text{T}}$ over 2002q1-2020q4. Higher values correspond to darker shades of brown. The legend shows five ranges for values of $\text{CyberNetSentiment}_{\text{it}}^{\text{T}}$ and their corresponding colors on the map.

Figure 3: **Global Flowmap of Major Cyber Risk Exposure**



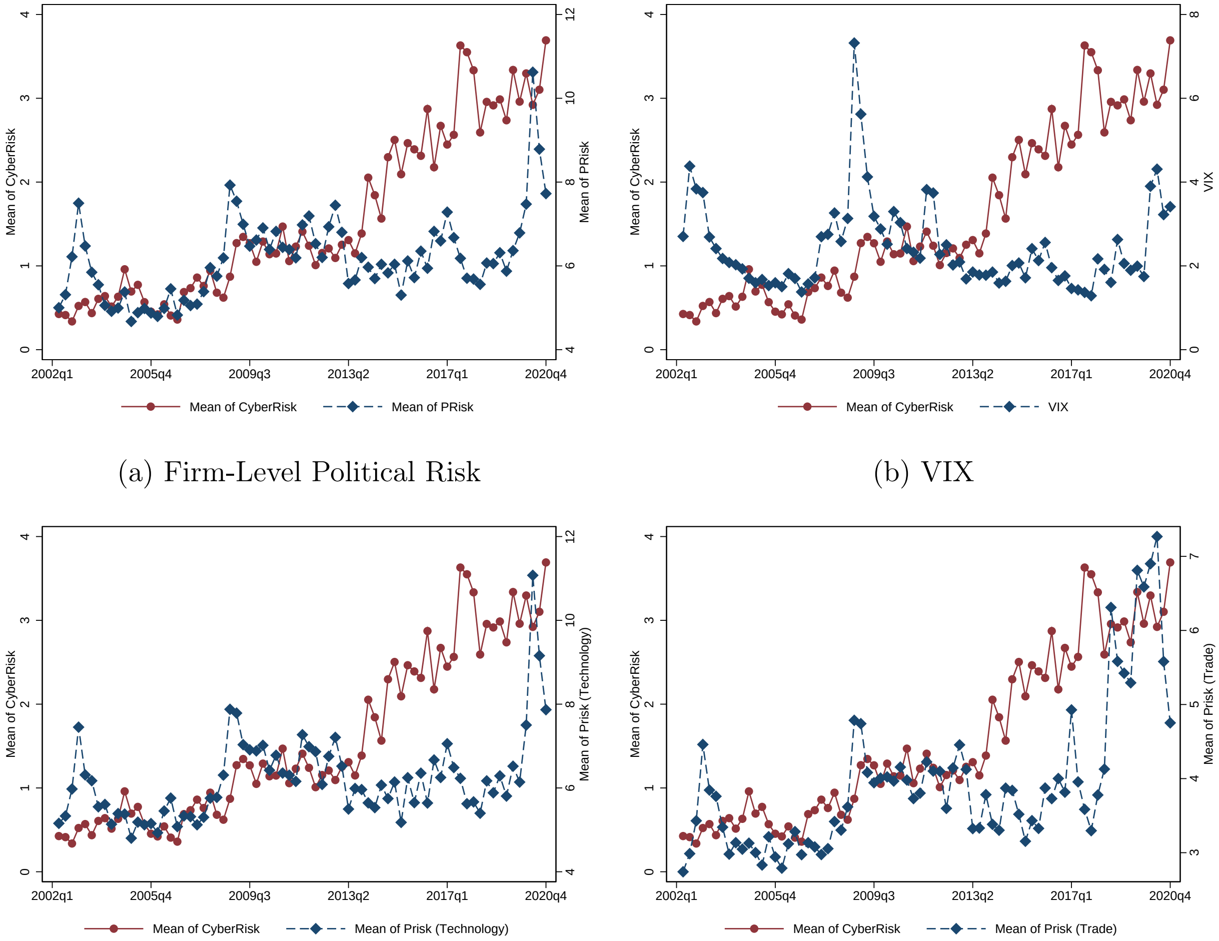
Notes: This figure visualizes the flow of $\text{CyberExposure}^{\text{T}}_{\text{it}}$ by major origin and destination. Origins (destinations) are unfilled (filled) circles. The six destination regions are United States, United Kingdom, Europe, Americas (ex US), Africa, and Asia. Destination regions are defined as headquarter locations of firms with $\text{CyberCountry}^{\text{T}}_{\text{it}} > 0$. Origin countries are the underlying terms/countries behind $\text{CyberCountry}^{\text{T}}_{\text{it}}$. The numbers behind this flowmap are summarized in Table 6 of the Online Appendix. This flowmap was created on the open source platform Flowmap.blue which was developed by Ilya Bolodin and is freely available under the MIT license.

Figure 4: **Firm-Level Cyber Risk - Global Exposure**



Notes: Time series average of CyberExposure$_{it}$ over all firms in each quarter, normalized by the standard deviation of the entire sample, and the time-varying percentage of all earnings calls with CyberExposure$_{it}$ > 0. All underlying cyber risk terms are listed in Table 1 of the Online Appendix.

Figure 5: **Correlations with other Indices of Risk and Volatility**



(a) Firm-Level Political Risk

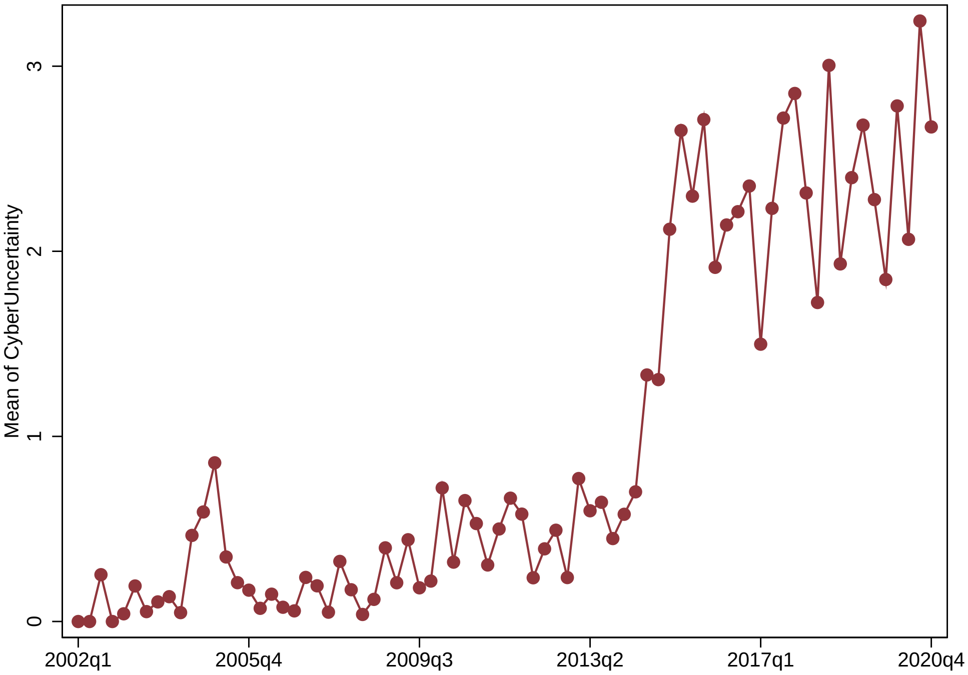
(b) VIX

(c) Firm-Level Political Risk (Technology)

(d) Firm-Level Political Risk (Trade)

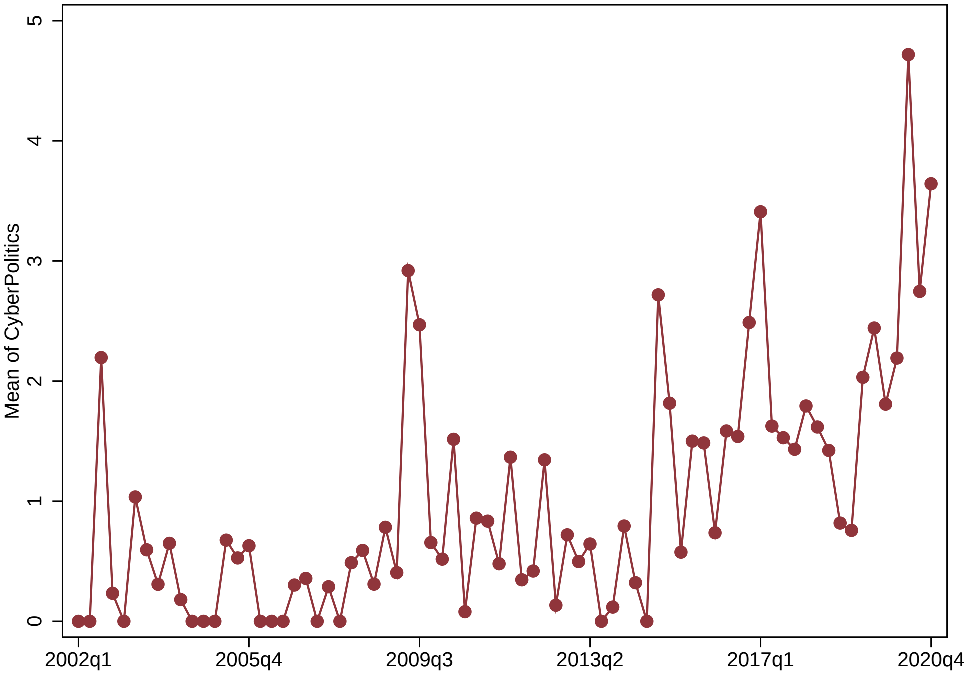
Notes: Time series average of CyberExposure$_{it}$, normalized by the standard deviation of the entire sample, and three external indices of uncertainty. Panel (a) plots the firm-level political risk (PRisk$_t$) index from Hassan et al. (2019). The Pearson correlation coefficient with the cyber risk measure is 0.4040 with the p-value of (0.003). Panel (b) plots the Cboe Volatility Index (VIX). The Pearson correlation coefficient with the cyber risk measure is -0.1159 with the p-value of (0.3221). Panel (c) plots the technology topic of firm-level political risk from Hassan et al. (2019). The Pearson correlation coefficient with the cyber risk measure is 0.3492 with the p-value of (0.003). Panel (d) plots the trade topic of firm-level political risk from Hassan et al. (2019). The Pearson correlation coefficient with the cyber risk measure is 0.6096 with the p-value of (0.000).

Figure 6: **Risk and Uncertainty**



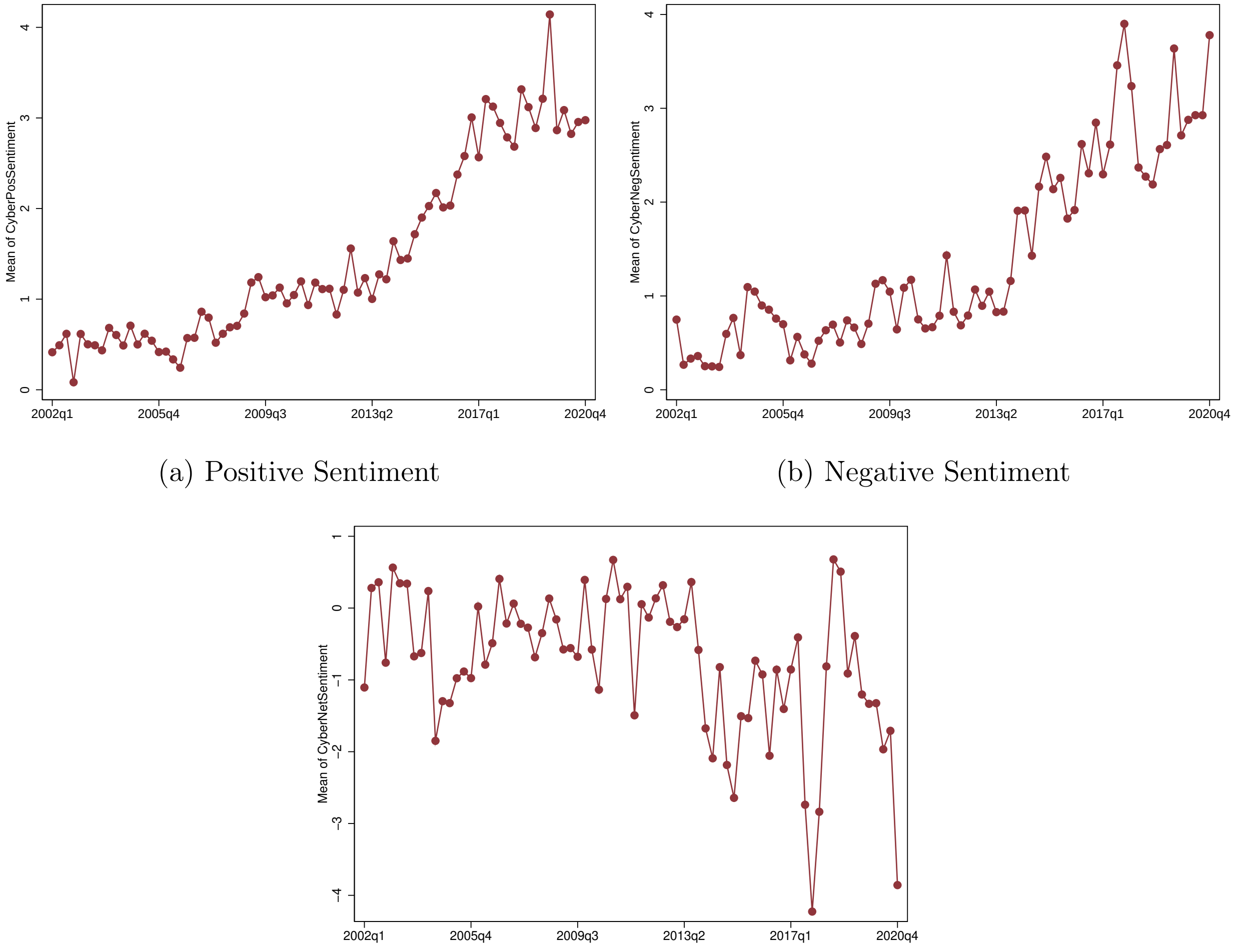
Notes: Time series average of $\text{CyberUncertainty}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Risk and Uncertainty" topic are listed in Table 2 of the Online Appendix.

Figure 7: **Politics**



Notes: Time series average of $\text{CyberPolitics}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Politics" topic are listed in Table 2 of the Online Appendix.

Figure 8: **Sentiment**



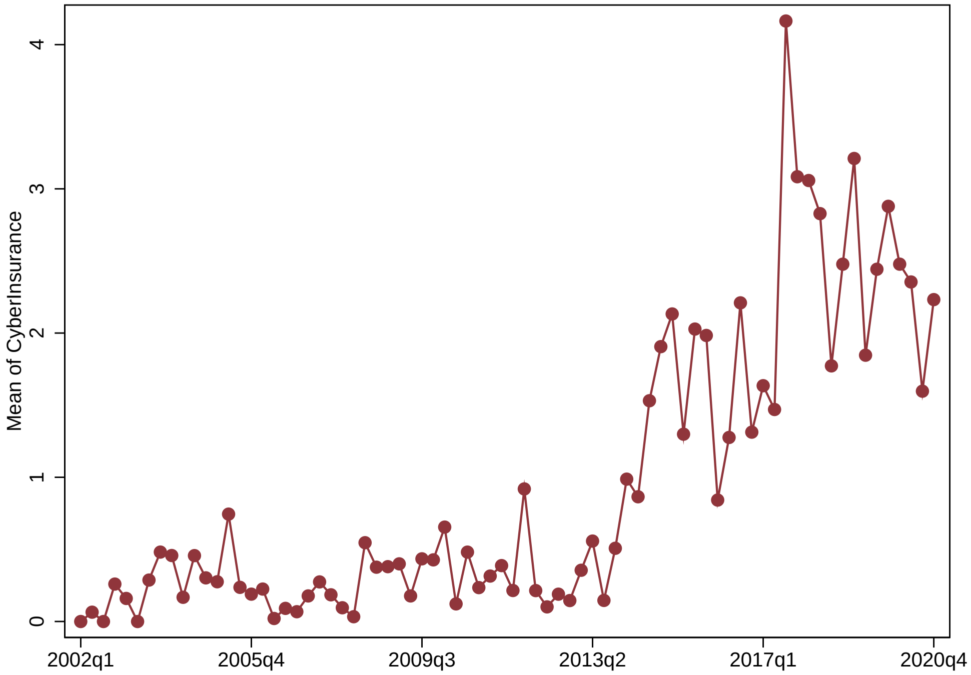
(a) Positive Sentiment

(b) Negative Sentiment

(c) Net Sentiment

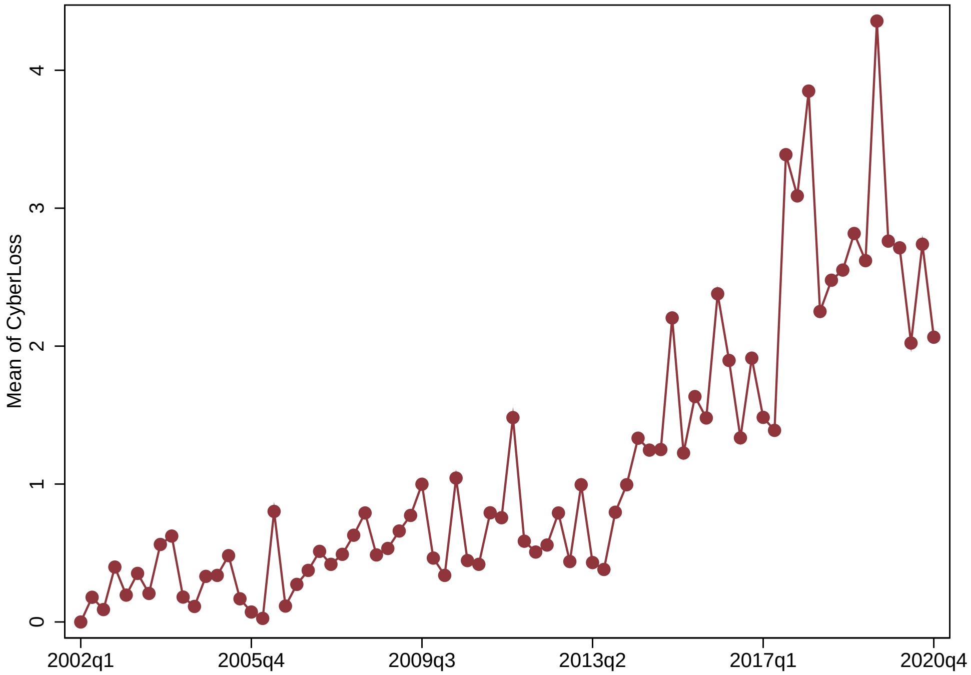
Notes: Time series average of $\text{CyberPosSentiment}_{\text{it}}$, $\text{CyberNegSentiment}_{\text{it}}$, and $\text{CyberNetSentiment}_{\text{it}}$, each normalized by the standard deviation of the respective sample. All terms behind the three sentiment topics are listed in Table 2 of the Online Appendix.

Figure 9: **Insurance and Legal**



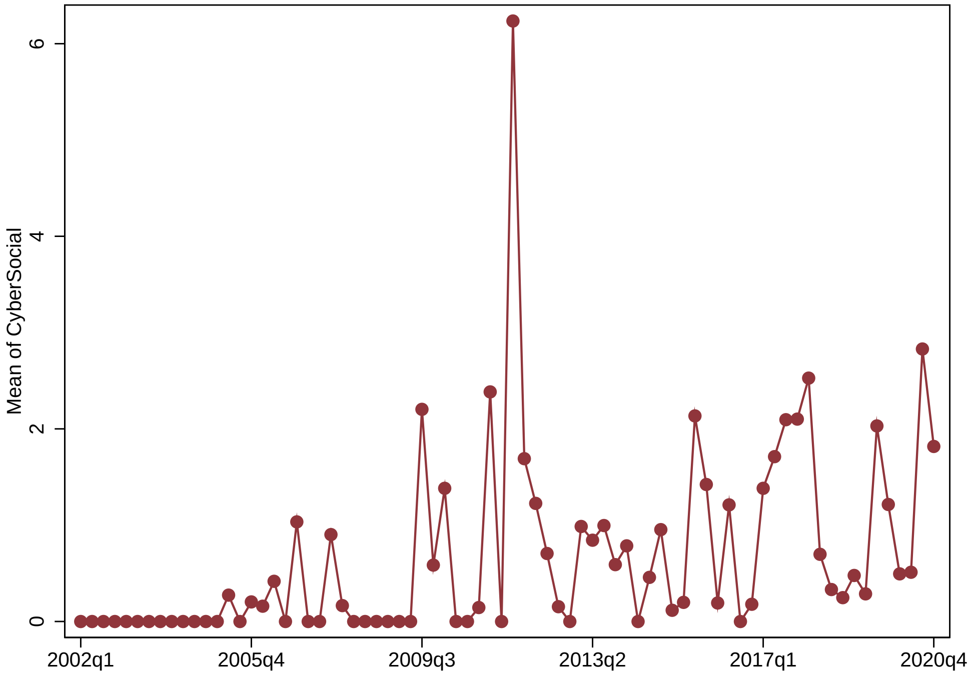
Notes: Time series average of $\text{CyberInsurance}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Insurance and Legal" topic are listed in Table 2 of the Online Appendix.

Figure 10: **Monetary Loss**



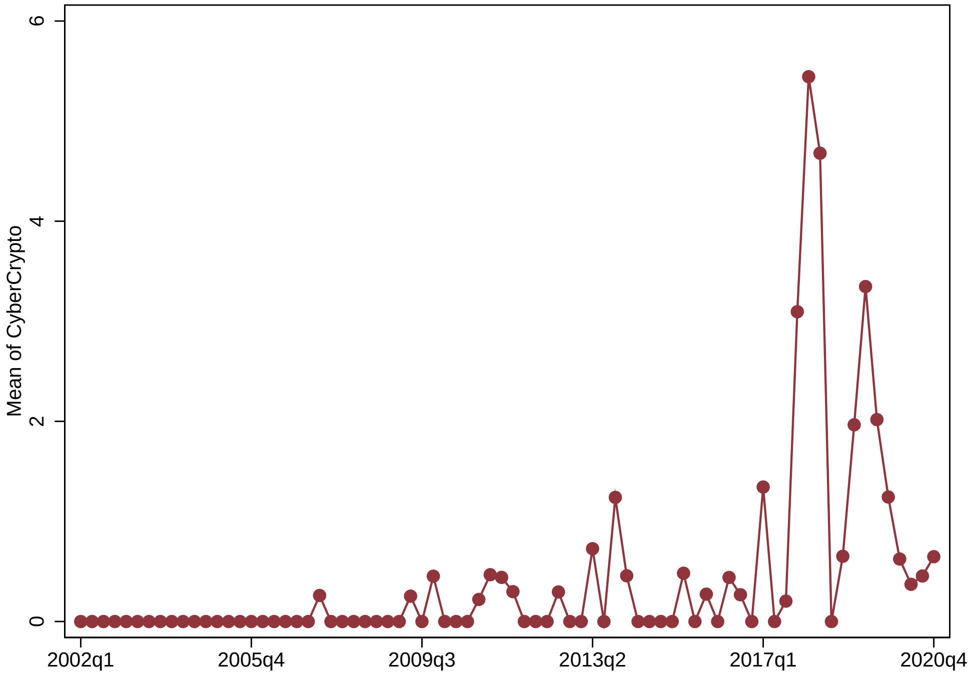
Notes: Time series average of $\text{CyberLoss}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Monetary Loss" topic are listed in Table 2 of the Online Appendix.

Figure 11: **Social Media**



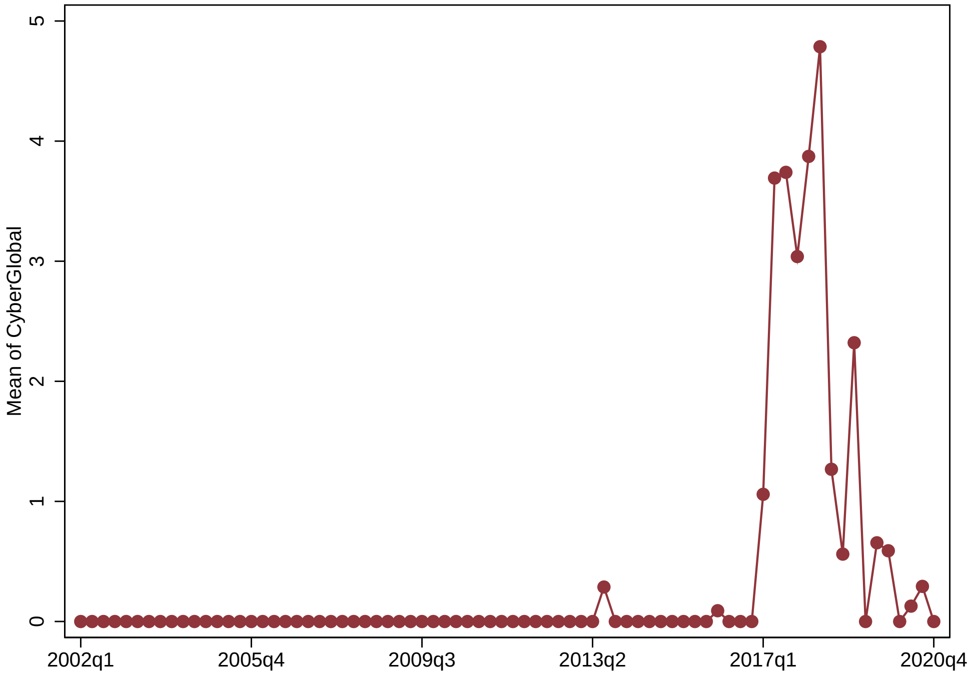
Notes: Time series average of CyberSocial$_{it}$, normalized by the standard deviation of the entire sample. All terms behind the "Social Media" topic are listed in Table 2 of the Online Appendix.

Figure 12: **Crypto**



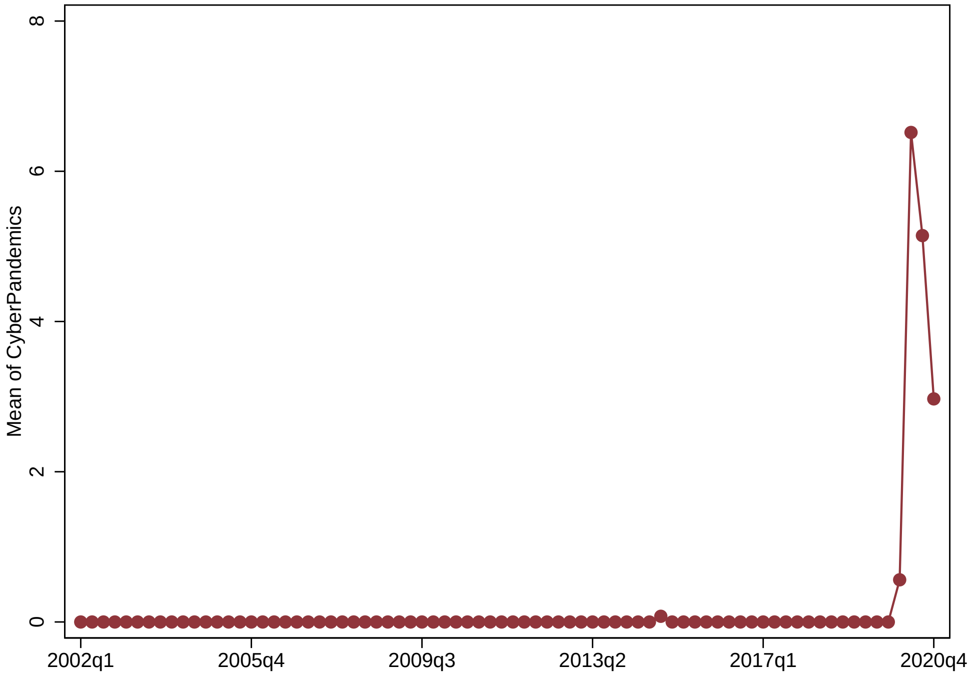
Notes: Time series average of CyberCrypto$_{it}$, normalized by the standard deviation of the entire sample. All terms behind the "Crypto" topic are listed in Table 2 of the Online Appendix.

Figure 13: **Global Events**



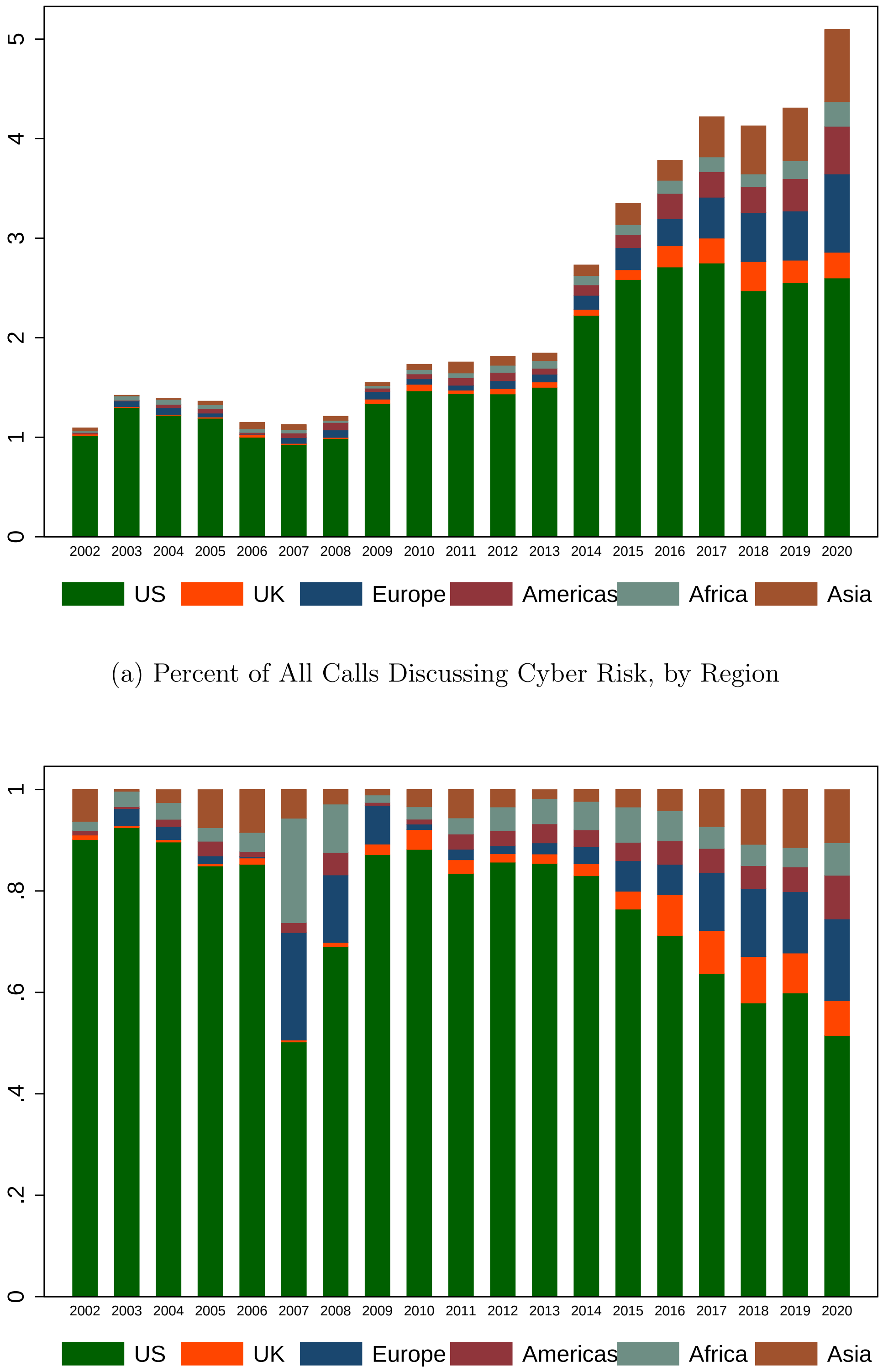
Notes: Time series average of $\text{CyberGlobal}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Global Events" topic are listed in Table 2 of the Online Appendix.

Figure 14: **Pandemics**



Notes: Time series average of $\text{CyberPandemics}_{\text{it}}$, normalized by the standard deviation of the entire sample. All terms behind the "Pandemics' topic are listed in Table 2 of the Online Appendix.

Figure 15: **Global Cyber Risk Exposure - Decomposition by Region**

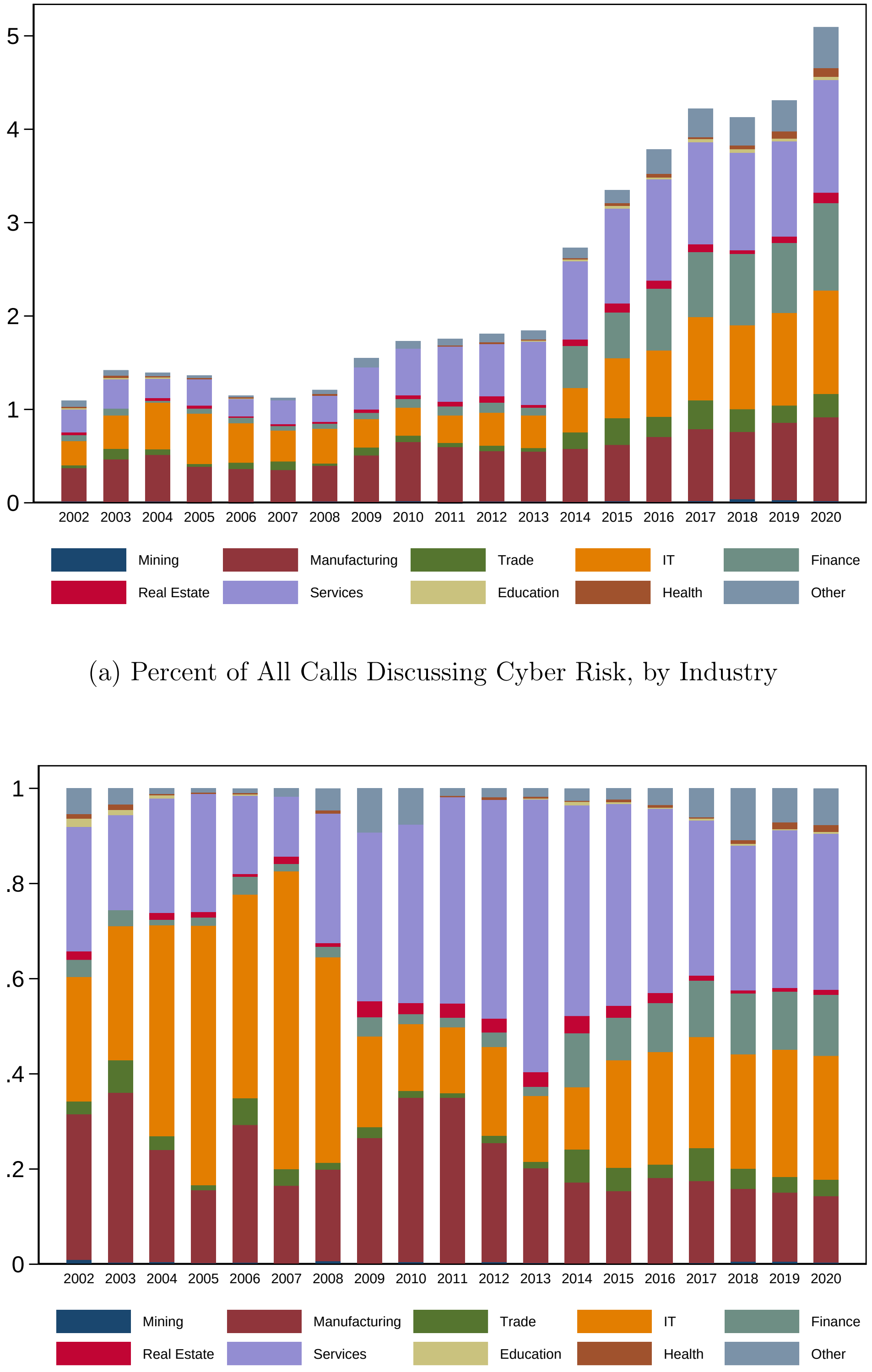


(a) Percent of All Calls Discussing Cyber Risk, by Region

(b) % of Global Cyber Risk Discussions, by Region

Notes: Region-specific totals of CyberExposure$^{T}_{it}$. Regions are defined by the location of firm headquarters from Compustat. On each bar, regional categories are in the following order from bottom to top: US, UK, Europe, Americas (ex US), Africa, Asia.

Figure 16: **Global Cyber Risk Exposure - Decomposition by Industry**

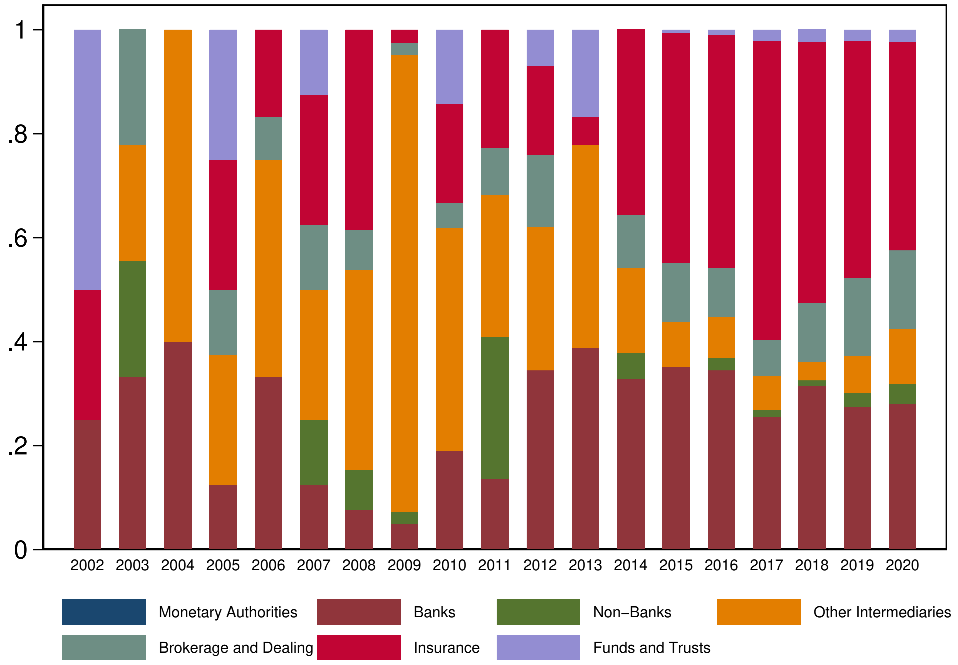


(a) Percent of All Calls Discussing Cyber Risk, by Industry

(b) % of Global Cyber Risk Discussions, by Industry

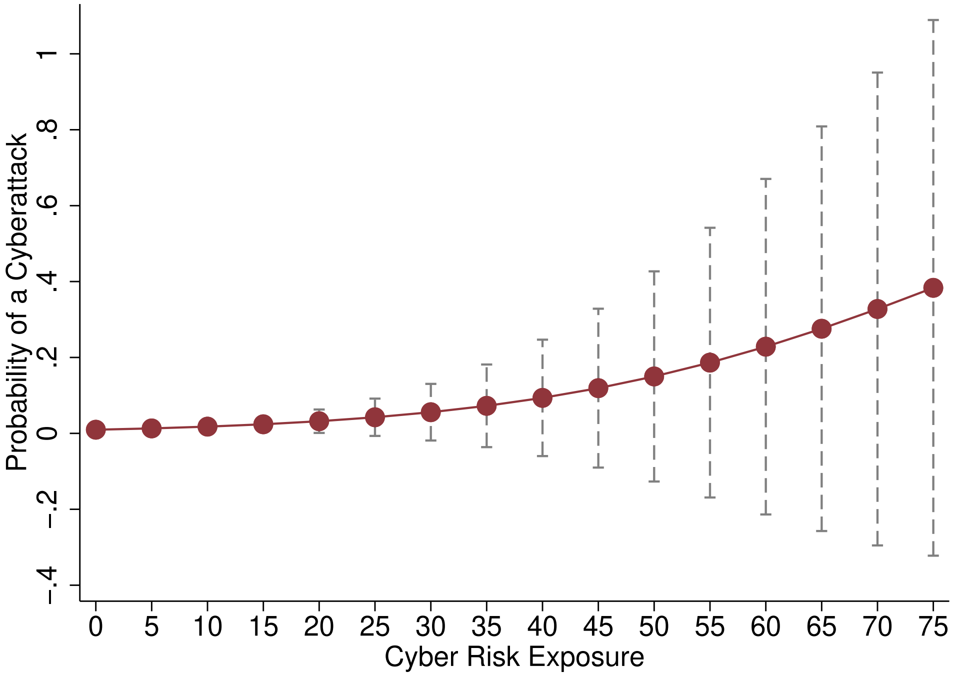
Notes: Industry-specific totals of $\text{CyberExposure}_{it}^{T}$. Industries are defined based on the 2-code NAICS classification. On each bar, industry categories are in the following order from bottom to top: Mining, Manufacturing, Trade, IT, Finance, Real Estate, Services, Education, Health, Other.

Figure 17: **Global Cyber Risk Exposure - Decomposition by Finance Sub-Industries**



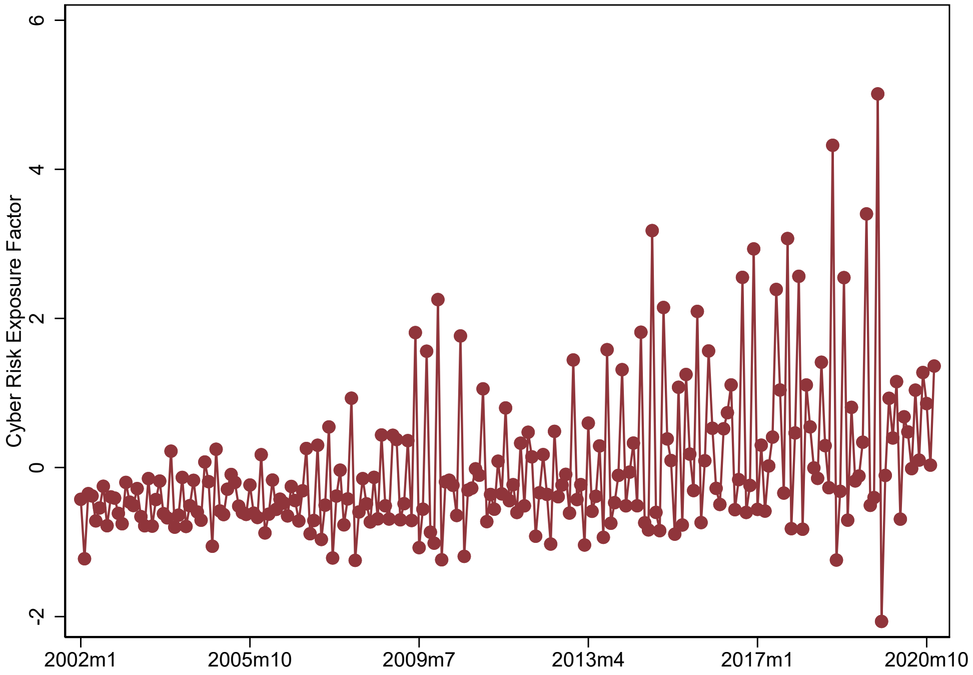
Notes: Financial industry-specific totals of CyberExposure$_{it}^{T}$. Finance industries are defined based on the 4-code NAICS classification. On each bar, industry categories are in the following order from bottom to top: Monetary Authorities, Banks, Non-Banks, Other Intermediaries, Brokerage and Dealing, Insurance, Funds and Trusts.

Figure 18: **Predicting Realized Cyberattacks**



Notes: This figure plots the margins plot from a firm-level logit regression of the realized cyberattack indicator on our text-based cyber risk total exposure measure, firm-level controls, and fixed effects. The cyberattack indicator captures whether a firm reports an actual cyberattack at least once within the next 8 quarters. Dashed lines indicate 95% confidence intervals.

Figure 19: **Cyber Risk Exposure (CyberE) - Monthly Pricing Factor**



Notes: The monthly pricing factor is defined as residuals from an AR(1) process which is fit onto the monthly sum of $\text{CyberExposure}_{\text{it}}^{\text{T}}$. The series is then standardized over the entire 2002:m1-2020m12 sample.

Table 1: **Predicting Future Cyberattacks with Firm-Level Cyber Risk Exposure**

Logit Model	Dependent Variable: Firm-Level Indicator of Reporting a Cyberattack within t+8 Quarters							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CyberExposure	0.0802*** (0.0272)	0.0660** (0.0275)						
CyberUncertainty			-0.2001 (0.4068)					
CyberPosSentiment				0.0984 (0.0996)				
CyberNegSentiment					0.0624** (0.0271)			
CyberNetSentiment						-0.0808** (0.0353)		
CyberInsurance							0.0844** (0.0358)	
CyberLoss								0.0759** (0.0368)
Firm Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	151706	60610	60610	60610	60610	60610	60610	60610
pseudo R2	0.083	0.171	0.170	0.170	0.170	0.171	0.170	0.170

Notes: Probability of reporting a cyberattack anytime over the next 8 quarters as a function of our text-based cyber risk measures. The quarterly sample is 2002:q1-2019q4, ending on the last quarter for which realized cyberattack data was available as of March 2021. Realized cyberattack data is from PRC. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 2: **Determinants of Firm-Level Cyber Risk Exposure**

Probit Model	Dependent Variable: Firm-level Indicator of Cyber Risk Exposure					
	(1)	(2)	(3)	(4)	(5)	(6)
	CyberExposure	CyberUncertainty	CyberPosSentiment	CyberNegSentiment	CyberNetSentiment	CyberCountry
PP&E / Assets	0.3170	-0.0148	0.7723**	0.3048	0.5219	-0.1575
	(0.2731)	(0.5662)	(0.3099)	(0.2884)	(0.3217)	(0.3297)
Intangibles / Assets	1.0656***	1.0501***	1.2684***	0.9479***	1.1552***	0.9374***
	(0.1960)	(0.3505)	(0.2328)	(0.2444)	(0.2308)	(0.2439)
CAPEX / Assets	-0.3971*	-0.1583	-0.5134**	-0.3653	-0.4926**	-0.6993**
	(0.2365)	(0.4226)	(0.2617)	(0.3090)	(0.2406)	(0.3205)
Cash Flow / Assets	-2.4574*	-1.9694	-1.0709	-0.9215	-0.5686	-0.7299
	(1.2874)	(1.7505)	(2.1063)	(1.8691)	(1.8384)	(1.5214)
Long-Term Debt / Assets	0.0646	0.1342	0.0816	0.0772	0.0662	0.2175
	(0.0837)	(0.2098)	(0.1035)	(0.1093)	(0.1045)	(0.1411)
Liquidity Ratio	0.6482***	0.6850	0.8028***	0.8670***	0.6274**	0.5598*
	(0.2412)	(0.4432)	(0.2672)	(0.2706)	(0.2728)	(0.3133)
Log (Size)	0.1261***	0.1013***	0.1204***	0.1101***	0.1070***	0.1245***
	(0.0181)	(0.0326)	(0.0209)	(0.0229)	(0.0191)	(0.0252)
Debt / Assets	-0.1336	0.3481	-0.0741	0.1261	-0.1327	0.1720
	(0.1525)	(0.3101)	(0.1980)	(0.1953)	(0.2161)	(0.2207)
Log (Age)	-0.0058	-0.0622	-0.0647	-0.0793	-0.0303	0.1129
	(0.0826)	(0.1510)	(0.0974)	(0.1087)	(0.0856)	(0.1012)
Equity Net Issuance	0.0825	0.0068	-0.1659	0.2240	-0.1375	0.7099*
	(0.3656)	(0.7952)	(0.4778)	(0.4809)	(0.4862)	(0.4118)
ROA	1.2804	2.0515	-0.0949	0.7725	-0.8002	2.2110*
	(1.2491)	(1.7329)	(2.0688)	(1.8494)	(1.7801)	(1.1620)
S&P Rating	0.0122	0.0198	0.0176	0.0139	0.0082	0.0510**
	(0.0177)	(0.0368)	(0.0207)	(0.0206)	(0.0217)	(0.0259)
Sales / Assets	1.4588*	-0.8862	0.9613	1.0599	1.2531	0.9551
	(0.8074)	(1.9782)	(0.9506)	(1.1672)	(0.9534)	(1.2560)
Tobin's Q	0.0107	0.0219	0.0050	-0.0260	0.0087	0.0095
	(0.0201)	(0.0372)	(0.0248)	(0.0329)	(0.0238)	(0.0265)
Book to Market Equity	-0.0733	-0.1492	-0.0979	-0.1021	-0.1049	-0.0079
	(0.0597)	(0.1685)	(0.0705)	(0.0800)	(0.0760)	(0.0863)
Operational Costs / Assets	-1.0813	1.0824	-0.4043	-0.7853	-0.7199	-0.6086
	(0.8197)	(1.9928)	(0.9541)	(1.1812)	(0.9522)	(1.2630)
Market Beta	-0.0082	0.0630	-0.0131	0.0137	0.0190	0.0247
	(0.0354)	(0.0677)	(0.0462)	(0.0446)	(0.0501)	(0.0494)
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74341	34651	68414	69344	66589	59617
pseudo R2	0.229	0.234	0.219	0.228	0.187	0.208

Notes: Probit regression of firm-level cyber risk measures on various balance sheet and income statement characteristics. All variables are defined in Table 3 of the Online Appendix. The quarterly sample is 2002:q1-2020q4. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 2: **Determinants of Firm-Level Cyber Risk Exposure (Continued)**

Probit Model	Dependent Variable: Firm-level Indicator of Cyber Risk Exposure						
	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	CyberPandemics	CyberCrypto	CyberInsurance	CyberLoss	CyberSocial	CyberGlobal	CyberPolitics
PP&E / Assets	0.2755	1.6551	1.1885***	0.4799	1.4708	-1.7795	0.0550
	(0.7028)	(1.0169)	(0.3892)	(0.3617)	(1.3305)	(4.8683)	(0.5169)
Intangibles / Assets	0.5889	0.9188*	1.4566***	1.1284***	1.4681	-1.7261	1.3470***
	(0.5979)	(0.4695)	(0.2817)	(0.2844)	(0.9225)	(1.6316)	(0.4020)
CAPEX / Assets	-0.2783	-0.9455*	-0.1366	-0.3968	-0.3297	1.5874	-0.2781
	(0.5882)	(0.5663)	(0.4055)	(0.3231)	(0.8617)	(1.9690)	(0.4097)
Cash Flow / Assets	2.9098	-13.9126***	-3.1388	-3.6773**	-7.9161***	91.6411	1.3686
	(2.1704)	(4.4872)	(2.5667)	(1.5217)	(2.9609)	(67.0790)	(1.9407)
Long-Term Debt / Assets	-0.5250	0.5117	0.0758	0.0522	2.9201***	7.0916*	0.4958**
	(0.3461)	(0.3293)	(0.1656)	(0.1350)	(0.9385)	(3.7475)	(0.2056)
Liquidity Ratio	1.2783*	-0.8107	0.4828	0.7086**	0.9902	-2.1304	1.6391**
	(0.7083)	(1.0592)	(0.3591)	(0.3139)	(1.0153)	(2.5588)	(0.7106)
Log (Size)	0.0803*	0.1619	0.1243***	0.0922***	0.1504***	0.3710	0.0617
	(0.0441)	(0.1118)	(0.0272)	(0.0238)	(0.0583)	(0.2311)	(0.0382)
Debt / Assets	0.4522	0.4846	-0.0380	-0.2552	-0.1622	1.9382	-0.0658
	(0.5498)	(0.5493)	(0.3033)	(0.2286)	(0.5277)	(2.3499)	(0.3257)
Log (Age)	0.4325	-0.3044	0.0267	0.0784	-0.1356	1.1454	0.1190
	(0.3042)	(0.2670)	(0.1490)	(0.1072)	(0.2218)	(0.7845)	(0.1964)
Equity Net Issuance	-1.7183	1.2072	-1.4871	-0.7007	0.2979	-45.4788*	0.1151
	(1.0660)	(2.0502)	(1.1426)	(0.6512)	(1.1943)	(23.8438)	(0.6612)
ROA	-1.2907	5.1689	3.5552*	2.5753*	7.0065**	-94.5802	1.3733
	(1.9882)	(3.7912)	(1.9534)	(1.4506)	(2.8152)	(72.6328)	(1.4708)
S&P Rating	0.0534	-0.0905	0.0503*	0.0505**	-0.0217	0.1260	-0.0134
	(0.0533)	(0.0742)	(0.0263)	(0.0247)	(0.0607)	(0.1269)	(0.0303)
Sales / Assets	-3.1632	9.4180**	1.8274	0.5807	-0.9744	5.2041	0.5860
	(3.0236)	(4.5974)	(1.8713)	(1.5519)	(4.1410)	(12.0847)	(2.2870)
Tobin's Q	-0.0765	-0.3513**	0.0096	-0.0350	-0.0818	-0.4446**	-0.0019
	(0.0484)	(0.1712)	(0.0376)	(0.0447)	(0.0781)	(0.1840)	(0.0452)
Book to Market Equity	-0.3716	-1.1101***	-0.1218	-0.1272	-0.6619	-15.4518***	0.1086**
	(0.2283)	(0.3843)	(0.1422)	(0.1088)	(0.4416)	(4.0433)	(0.0528)
Operational Costs / Assets	3.9248	-8.7336*	-1.4798	-0.4025	2.4465	5.7803	-0.1161
	(3.0691)	(4.7470)	(1.8901)	(1.5581)	(4.1919)	(14.0201)	(2.3677)
Market Beta	0.0854	-0.1035	0.0196	-0.0247	0.0113	0.5231*	0.1042
	(0.1596)	(0.1699)	(0.0775)	(0.0503)	(0.1229)	(0.3112)	(0.0723)
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3430	2612	38005	58847	3897	714	22198
pseudo R2	0.221	0.195	0.206	0.189	0.193	0.482	0.165

Notes: Probit regression of firm-level cyber risk measures on various balance sheet and income statement characteristics. All variables are defined in Table 3 of the Online Appendix. The quarterly sample is 2002:q1-2020q4. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 3: **Cyber Risk and Stock Market Effects**

	Dependent Variable: Average Daily Returns for Weeks Surrounding Earnings Calls					
	Affected Firms		Peers		Non-Peers	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
	(1)	(2)	(3)	(4)	(5)	(6)
CyberExposure	-0.0355***	-0.0430***	-0.0239*	-0.0276***	-0.00385	-0.000468
	(0.0107)	(0.0102)	(0.0119)	-0.00903	-0.00323	-0.00135
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2782	2782	2782	2782	2782	2782
R2	0.113	0.108	0.346	0.316	0.93	0.858

Notes: Regressions of average daily stock returns for calendar weeks surrounding earnings calls with positive $\text{CyberExposure}_{\text{it}}$. Affected firms are those who record $\text{CyberExposure}_{\text{it}} > 0$. Peers are defined as firms with $\text{CyberExposure}_{\text{it}} = 0$ but which are from the same country and industry (defined by the 6-digit NAICS classification) as the affected firms. Non-peers are defined as firms with $\text{CyberExposure}_{\text{it}} = 0$ and which are from the same country but not the same industry (defined by the 6-digit NAICS classification) as the affected firms. Columns (2), (4), and (6) report results for value-weighted returns. Standard errors are clustered at the 2-digit NAICS level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. The sample is over 2002w4-2020w52.

Table 4: **Cyber Risk Topics and Stock Market Effects**

	Dependent Variable: Average Daily Returns for Weeks Surrounding Earnings Calls											
	Affected Firms				Peers				Non-Peers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CyberUncertainty	-0.041***				-0.045***				0.017			
	(0.010)				(0.006)				(0.017)			
CyberNegSentiment		-0.066***				-0.054***				0.001		
		(0.008)				(0.009)				(0.002)		
CyberInsurance			-0.074*				-0.054***				0.003	
			(0.033)	-0.064**			(0.010)	-0.046**			(0.008)	-0.017
CyberLoss				(0.028)				(0.016)				(0.015)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	338	1200	285	481	338	1719	285	481	338	1719	285	481
R2	0.106	0.108	0.046	0.214	0.460	0.325	0.414	0.420	0.818	0.853	0.874	0.910

Notes: Regressions of average daily stock returns for calendar weeks surrounding earnings calls with a positive cyber risk topical exposure. Affected firms are those who with positive exposure. Peers are defined as firms with zero exposure but which are from the same country and industry (defined by the 6-digit NAICS classification) as the affected firms. Non-peers are defined as firms with zero cyber risk exposure and which are not from the same country and industry (defined by the 6-digit NAICS classification) as the affected firms. All results are for value-weighted returns. Standard errors are clustered at the 2-digit NAICS. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. The sample is over 2002w4-2020w52.

Table 5: **Cyber Risk Sorted Portfolios**

Panel A: CyberE-Beta-Sorted Value Weighted Portfolios

	L (1)	(2)	(3)	(4)	H (5)	H-L
Average Excess Returns (%)	0.978	0.850	0.801	0.755	0.708	-0.269**
Volatility (%)	5.349	4.304	4.112	4.309	5.267	1.892
Alpha CAPM	-0.174	0.018	0.099	0.115	0.088	-0.262**
Alpha Fama-French	-0.060	0.101	0.166**	0.191**	0.192	-0.251**
Alpha Fama-French, Momentum	-0.0493	0.113*	0.196***	0.212***	0.231**	-0.280**
Average Market Cap ($bn)	20.734	21.182	21.043	20.844	20.330	
Cyber Exposure beta	-3.523	-1.188	-0.034	1.119	3.579	
Number of Months	186	186	186	186	186	186

Panel B: CyberE-Beta-Sorted Equal Weighted Portfolios

	L (1)	(2)	(3)	(4)	H (5)	H-L
Average Excess Returns (%)	0.987	0.852	0.799	0.759	0.712	-0.275**
Volatility (%)	5.355	4.299	4.097	4.279	5.260	1.897
Alpha CAPM	-0.168	0.029	0.103	0.121	0.100	-0.268**
Alpha Fama-French	-0.051	0.115	0.172**	0.201**	0.209*	-0.261**
Alpha Fama-French, Momentum	-0.0402	0.127*	0.204***	0.224***	0.249**	-0.290**
Average Market Cap ($bn)	20.734	21.182	21.043	20.844	20.330	
Cyber Exposure beta	-3.523	-1.188	-0.034	1.119	3.579	
Number of Months	186	186	186	186	186	186

Notes: CyberE-sorted stock portfolios. Firm-level monthly excess returns are regressed on the CyberE pricing factor in 30-month rolling regressions. Five value- and equal-weighted tradeable portfolios are formed based on the cyber beta. The H-L portfolio is long high- and short low-CyberE beta stocks. Price data is from CRSP. Betas are scaled by 100 for readability. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Estimations are run on the 2002:m1-2019m12 sample.

Table 6: **Fama-MacBeth Regressions - Pricing 10 Cyber Risk Sorted Portfolios**

	Unweighted				Weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Market	0.228	0.164	0.390	0.698	0.203	0.111	0.394	0.724
CyberE		0.952***	0.543**	0.706**		0.989***	0.530**	0.705**
HML			-1.541**	-2.280**			-1.540**	-2.260**
SMB			0.0537	-0.0612			0.0151	-0.105
MOM				1.922				1.969*
Constant	0.577	0.484***	0.558	0.492	0.599	0.538***	0.561*	0.471
Observations	10	10	10	10	10	10	10	10
R2	0.080	0.733	0.922	0.934	0.064	0.742	0.943	0.958
MAPE	0.915	0.502	0.301	0.275	0.924	0.500	0.255	0.221

Notes: The table reports results from Fama-MacBeth cross-sectional regressions. Firm-level monthly excess returns are regressed on the CyberE pricing factor in 30-month rolling regressions. Ten value- and equal-weighted tradeable portfolios are formed based on the cyber beta. For each portfolio, average returns are computed. Cross-sectional regressions of average returns on the factor betas are then run. HML and SMB refer to the book-to-market and size factors from Fama and French (1993). MOM is the momentum factor from the Kenneth French data repository. MAPE is the mean average pricing error, in annualized percentage terms. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Estimations are run on the 2002:m1-2019m12 sample.

Online Appendix for “The Anatomy of Cyber Risk”

Rustam Jamilov   Hélène Rey   Ahmed Tahoun

Contents

A Snippets from Earnings Calls Transcripts

Table 1: **Earnings Calls Snippets around Reported Cyberattacks**

Company Name	Call Date	CyberExposure	CyberNetSentiment	Excerpt of Discussion	Call summary
Equifax	10.11.2017	10	-10	I have to admit given the impact of –cyber– it is a little more ((difficult)) right now to give . . . of the cybersecurity incident These costs were generally for legal –cyber– forensic investigations and other professional services million in accrued expenses . . . do we have insurance to cover costs in connection with the –databreach– ((breach)) ((incidents)) with limits in excess of the current amount . . . as ((opposed)) to being . . . the type of cost that we've incurred	In March 2017, personally identifying data belonging to 147+ million of people was stolen. Incident lead to the company agreeing to a global settlement with the Federal Trade Commission, the Consumer Financial Protection Bureau, and 50 U.S. states and territories. By most accounts, the largest data breach in U.S. commercial history.
State Bank of India	19.05.2017	4	-3	the SBI is doing to check any impact on recent –cyberattack– attack in view of reports about of ATMs in India . . . is putting together the technology to prevent any kind of –malware– or cyberhacking to come inside our system to either ((disrupt)) . . . ((difficult)) to say that we are ready to prevent of –cyber– attacks So if an ((incident)) happens how do we manage . . . apart from reporting to the regulators and to the governments –cyber– emergency response team we also share amongst each other to	It was reported that 3.2 million debit cards were compromised in late 2016, affecting major Indian banks including the SBI which was among the worst hit. The breach underwent for several months. SBI announced the blocking and replacement of 500,000+ debit cards
Bank of Montreal	30.05.2018	5	-1	continued (good) performance Now on the topic of the recent –cyber– ((incident)) As Darryl said we are focused on our customers . . . for industries like ours Within this changing landscape information and –cybersecurity– security has been an ongoing priority for some time and . . . nexus between this accelerated digital transformation and the ((breach)) the –databreach– ((breach)) Ive always thought of this digital transformation as (enhancing) . . . is As the technology people will describe it or the –cyber– people describe it as the attack surface increases so is . . . we got to be (better) prepared for it Now this –databreach– ((breach)) as far as I can tell and whatever it	Canada's 4th largest financial services institution reported that more than 50,000 accounts across the country were in the hands of hackers. Following the incident disclosure, shares traded down by 1%, as per reports.
Capital One	24.10.2019	6	-5	And finally we recognized million of charges associated with the –cyber– ((incident)) that we announced at the end of July These . . . to million in certain incremental direct costs associated with the –cyber– ((incident)) response and that we expected to record these costs . . . expect to make incremental investments in cybersecurity related to the –cyber– ((incident)) and we expect to absorb the estimated incremental investments . . . Sanjay with respect to the public cloud and then the –cyber– ((incident)) while the event occurred in the cloud the ((vulnerability))	An outside individual gained unauthorized access and obtained certain types of personal information about Capital One credit card customers and individuals who had applied for our credit card products. The event affected approximately 100 million individuals in the United States and approximately 6 million in Canada.

Table 2: **Earnings Calls Snippets around Reported Cyberattacks (Continued)**

Company Name	Call Date	CyberExposure	CyberNetSentiment	Excerpt of Discussion	Call summary
Target	26.02.2014	19	-19	performance along with costs related to our recent ((restructuring)) and –databreach– ((breach)) along with small accounting and tax matters As weve . . . of the recent ((slowdown)) in growth weve seen following the –databreach– ((breach)) In Canada in we generated just over billion in . . . headwind and we continue to see the impact of the –databreach– ((breach)) on guest sentiment and traffic We believe that well . . . (beneficial) interest asset and any potential costs related to the –databreach– ((breach)) While this has been a ((challenging)) year we are	Over 100 million individuals were exposed in the attack. Target reported that the information compromised in the attack included mailing addresses, names, email addresses, phone numbers, and credit and debit card account data.
Maersk	07.11.2017	31	5	third quarter of As youre well aware we had a –cyberattack– attack that impacted the business ((severely)) in July and into August mainly in Maersk Line and Damco This –cyberattack– attack caused volume and revenue ((loss)) as well as additional . . . a quarter with solid global demand growth And adjusting for –cyber– ((loss)) we would have had a flat development in volumes . . . is of course temporary working capital elements related to a –cyberattack– attack We were somewhat ((slower)) on invoicing and therefore our operations were significantly ((hampered)) in the third quarter by the –cyberattack– attack and were certainly not (pleased) with	The NotPetya ransomware attack had a "devastating" effect on Maersk. As per Adam Banks - head of technology and global transport - all end-user devices including 50,000 laptops were destroyed. 1,200 applications were destroyed or inaccessible. More than 50% of company servers were destroyed. Any recovered data or devices got immediately reinfected.
Merck	28.07.2017	18	5	on our second quarter results Overall full recovery from the –cyberattack– attack will take some time but we are making steady . . . As Ken has outlined the –malware– that infected our computational environment had a very substantial effect . . . resources behind our ongoing launches remediation expenses related to the –cyberattack– attack as well as additional RD costs associated with our gears for a moment Let me speak briefly about the –cyberattack– attack on June which as you know affected Merck along	In June 2017, the NotPetya ransomware attack crippled more than 30,000 computers at Merck, as well as 7,500 servers (per Bloomberg). Affected screens froze and the following message appeared: "Ooops, your important files are encrypted. . . . We guarantee that you can recover all your files safely and easily. All you need to do is submit the payment". The cost was 300 bitcoins or roughly $600,000.
Tencent	10.11.2010	17	2	attack by a program which we believe to be a –malware– and its called Kou Kou Bodyguard The malware was created Qihoo which also operates the most (popular) Internet security software Security Guard in China The –malware– (encouraged) users to install it by offering functions such as . . . the fact that an Internet security software company actually developed –malware– targeting and affecting an application software We intend to seek . . . government authorities However the ensuing ((investigation)) took time while the –malware– spread quickly with s (strong) promotion support In a matter	More than 70 victims dating back to 2006 were affected, allegedly, by Chinese hacking operations that lasted up until 2011. The series of attacks was orchestrated by Elderwood Group based in Beijing, China. The codeword for the incidents was Operation Aurora. Google was the first company to publically disclose the cyberattack, followed by a series of disclosures across dozens of organizations. (source: Economist.com, Forbes.com, Wired.com)

B Variable Definitions

Table 1: **Cyber Risk Terms**

Sub-Category	Terms
Cyber	"cyber", "cybersecurity", "network security", "cyberattack", "cybercrime", "cyber threat", "cyber incident", "cyber event"
Data	"data loss", "data integrity", "data security", "information theft", "data breach", "data theft", "data leak", "data compromise", "data fraud"
Malware	"worm", "spyware", "phishing", "trojan", "malware", "ddos attack", "ransomware"
Fraud	"hacker", "hack", "hacked", "card fraud", "card breach", "system outage", "email compromise"

Notes: All cyber risk related terms that we search for in the quarterly earnings calls transcripts.

Table 2: **Topics of Joint Searches**

Topic	Terms	Source
Country Names	Russia, Russian, China, Chinese, North Korea, North Korean, Israel, Israelian, Iran, Iranian, United States, the US, America, American, Europe, European, the EU, United Kingdom, Great Britain, British, the UK	This paper
Crypto	crypto, crypto currency, cryptocurrency, ledger, cryptography, blockchain, bitcoin, altcoin,token, ethereum, rupple, litecoin, tether, libra, monero, diem	This paper
Insurance and Legal	insurance, liability, coverage, cover, policy, legal, law, settle, settlement	This paper
Monetary Loss	loss, cost, income, reputation, monetary, damage, recover	This paper
Pandemics	corona, coronavirus, corona virus, covid, sars	This paper
Social Media	Zoom, webex, hangouts, Facebook, Google, Twitter, Bing, Snapchat, Linkedin, MailChip,Baidu, Tencent, Weibo, Yandex, Rambler, Line, whatsapp	This paper
Politics	election, state, sponsor, state sponsored, state-sponsored, espionage, democratic national committee, DNC, Trump, Clinton, Assange	This paper
Global Events	GDPR, Cambridge, Cambridge Analytica, Notpetya, Wannacry, Wikileaks, Wiki leaks, Panama papers	This paper
Risk and Uncertainty	risk, risks, uncertainty, variable, chance, possibility, pending, uncertainties, uncertain, doubt, prospect, bet, variability, exposed, likelihood, threat, probability, unknown, varying, unclear, unpredictable, speculative, fear, reservation, hesitant, gamble, risky, instability, doubtful, hazard, tricky, sticky, dangerous, tentative, hazardous, queries, danger, fluctuating, unstable, vague, erratic, query, jeopardize, unsettled, unpredictability, dilemma, skepticism, hesitancy, riskier, unresolved, unsure, irregular, jeopardy, suspicion, risking, peril, hesitating, risked, unreliable, unsafe, hazy, apprehension, unforeseeable, halting, wager, torn, precarious, undetermined, insecurity, debatable, undecided, dicey, indecision, wavering, iffy, faltering, endanger, quandary, insecure, changeable, riskiest, hairy, ambivalent, dubious, riskiness, treacherous, oscillating, perilous, tentativeness, unreliability, wariness, vagueness, dodgy, equivocation, indecisive, chancy, menace, qualm, vacillating, gnarly, disquiet, ambivalence, imperil, vacillation, incalculable, untrustworthy, equivocating, diffident, fickleness, misgiving, changeability, undependable, incertitude, fitful, parlous, unconfident, defenseless, unsureness, fluctuant, niggle, diffidence, precariousness, doubtfulness,	Hassan et al. (2019)
Positive Sentiment	good, strong, great, better, opportunities, able, positive, progress, opportunity, best, improvement, benefit, improve, pleased, improved, improving, success, effective, profitability, successful, greater, stronger, strength, advantage, leadership, achieve, despite, confident, improvements, achieved, excited, favorable, stable, leading, efficiency, gain, happy, optimistic, gains, profitable, innovation, excellent, encouraged, attractive, win, efficient, benefited, highest, tremendous, enhance, exciting, achieving, enable, successfully, efficiencies, easy, strengthen, enhanced, encouraging, strengthening, innovative, stability, excellence, satisfaction, pleasure, winning, superior, gaining, perfect, easier, alliance, collaboration, enabled, advantages, exceptional, stabilize, gained, strongest, accomplished, enhancing, enables, valuable, impressive, progressing, strengthened, enjoy, positively, efficiently, exclusive, achievement, strengths, enabling, easily, stabilized, satisfied, accomplish, benefiting, accomplishments, transparency, diligently	Hassan et al. (2019)
Negative Sentiment	loss, decline, difficult, against, negative, restructuring, challenges, force, late, closing, declined, losses, critical, challenging, weak, closed, problem, claims, break, slow, recall, challenge, delay, concerned, bad, cut, concern, problems, litigation, weakness, volatility, difficulty, lost, crisis, concerns, declines, weaker, delays, impairment, opposed, recession, slowdown, downturn, slower, closure, lack, unfortunately, missed, declining, adverse, negatively, unemployment, worse, lag, wrong, bridge, delayed, severe, dropped, volatile, lose, disclosed, shut, complicated, breakdown, slowing, serious, difficulties, disclose, losing, slowed, stress, caution, disruption, discontinued, failure, challenged, downward, poor, deficit, suspect, slowly, nonperforming, unfavorable, deterioration, opportunistic, termination, miss, investigation, breaking, shortage, attrition, damage, chargeoffs, worst, drag, hurt, disappointed, bankruptcy, shutdown	Hassan et al. (2019)

Notes: All topic-specific terms that we search for in the quarterly earnings calls transcripts.

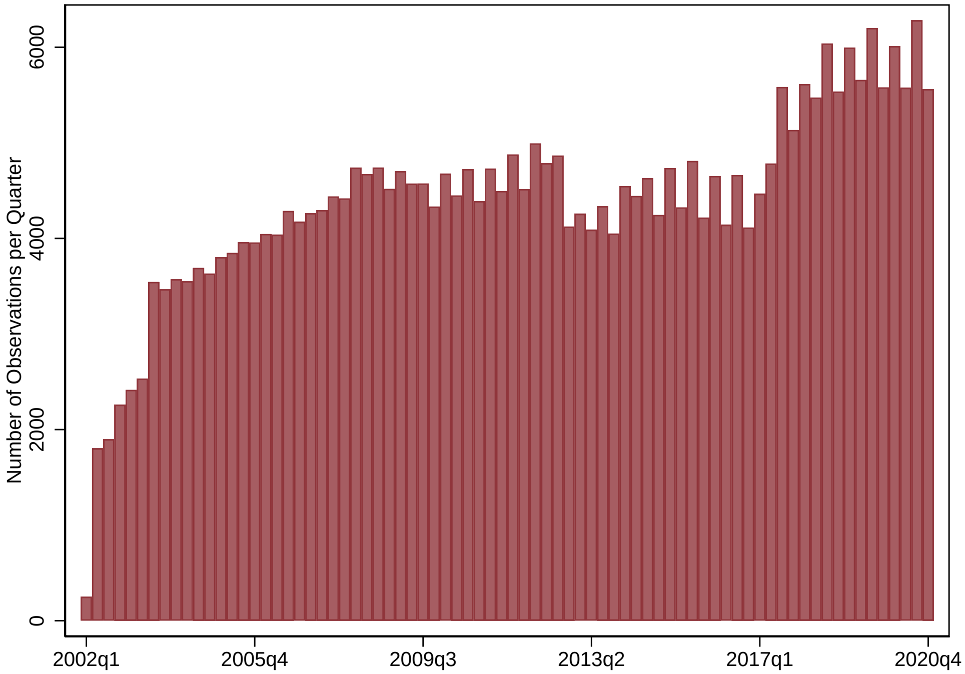
Table 3: **Variable Definitions**

Variable	Description	Source
PP&E	Property, planet, and equipment expenditures (ppent)	Compustat
Intangibles	Intangible assets (intan)	Compustat
CAPEX	Total invested capital (icapt)	Compustat
Cash flow	Income before extraordinatory items (ib) + depreciation and amortization (dp) / assets	Compustat
Debt	Long-term debt (dltt) + debt in current liabilities (dlc)	Compustat
Liquidity	Cash and short-term investments / assets	Compustat
Firm size	Log (total assets)	Compustat
Equity net issuance	Common shares issued (cshi)	Compustat
Total equity	Stockholders equity (seq)	Compustat
ROA	Net income (ni) / assets	Compustat
S&P Rating	S&P quality ranking (spcsrc)	Compustat
Tobin's Q	Total assets - common equity + market equity (at-ce+prcc*csho) / at	Compustat
Book to market ratio	Common equity (ceq) / market equity (prcc*csho)	Compustat
Operational expenses	Operating expense (xopr)	Compustat
Market beta	Market beta of stocks estimated using monthly returns over rolling 30 months	CRSP

Notes: Firm variables used throughout the paper.

C Summary Statistics

Figure 20: **Observations per Quarter**



Notes: Number of observations per quarter in the earnings calls dataset.

Table 4: **Statistics by Industry**

	Cyber Risk Measure											
Industry	Cyber Exposure	Cyber Uncertainty	Cyber Pos Sentiment	Cyber Neg Sentiment	Cyber Country	Cybre Pandemics	Cyber Crypto	Cyber Insurance	Cyber Loss	Cyber Social	Cyber Global	Cyber Politics
Mining	88	6	26	24	14	0	0	2	11	0	0	3
Manufacturing	4770	152	1754	1447	984	69	19	199	604	20	3	49
Trade	959	13	312	586	189	25	4	143	276	39	2	16
IT	6210	536	1893	2949	773	96	24	447	798	116	87	130
Finance	2268	399	659	916	346	63	25	859	635	16	27	31
Real Estate	427	16	118	106	36	2	1	28	23	3	1	9
Services	8984	498	2899	3276	1100	154	31	487	987	68	50	147
Education	76	1	26	11	13	0	0	14	6	0	0	0
health	164	13	62	65	6	7	0	30	52	2	0	2
Other	1416	77	441	897	259	42	5	457	449	20	6	63
Total	25362	1711	8190	10277	3720	458	109	2666	3841	284	176	450

Notes: Total count of each cyber risk exposure measure by industry, pooled over the entire sample. Industries are defined by the 2-digit NAICS classification.

Table 5: **Statistics by Country**

	Cyber Risk Measure											
	Cyber Exposure	Cyber Uncertainty	Cyber Pos Sentiment	Cyber Neg Sentiment	Cyber Country	Cybre Pandemics	Cyber Crypto	Cyber Insurance	Cyber Loss	Cyber Social	Cyber Global	Cyber Politics
UAE	1	0	0	0	0	0	0	0	0	0	0	0
Argentina	7	0	2	0	2	2	0	0	1	0	0	0
Austria	7	0	2	1	0	0	0	1	4	0	0	0
Australia	206	21	77	66	14	12	0	13	40	4	0	0
Belgium	22	1	3	4	7	1	0	1	5	0	0	1
Bermuda	158	31	50	41	54	2	1	92	25	0	0	0
Brazil	63	2	10	42	6	16	0	3	6	0	0	4
Canada	738	42	277	217	140	19	3	66	99	4	2	17
Switzerland	221	12	70	39	67	3	4	47	63	5	2	2
Chile	72	13	31	29	9	0	0	16	31	0	0	0
China	121	8	35	37	37	0	1	6	24	14	0	0
Colombia	5	0	2	1	2	0	0	0	2	0	0	0
Cyprus	3	0	0	2	0	0	0	0	0	0	0	0
Czechia	12	0	3	3	5	2	0	0	1	0	0	0
Germany	283	25	63	95	86	8	3	105	82	0	2	1
Denmark	152	8	43	70	16	23	3	31	60	35	0	0
Egypt	1	0	0	0	0	0	0	0	1	0	0	0
Spain	64	0	19	10	4	7	0	7	1	3	0	0
Finland	198	5	78	61	15	15	0	9	16	0	1	4
France	507	21	139	134	158	9	1	38	77	9	9	4
United Kingdom	1368	165	488	528	437	23	4	206	287	9	32	7
Guernsey	6	0	2	3	3	2	0	0	0	0	1	0
Gibraltar	1	0	0	0	0	0	0	0	0	0	0	0
Grece	4	4	3	1	0	0	0	0	0	0	0	0
Hong Kong	53	9	15	30	12	0	2	9	16	1	1	0
Indonesia	4	0	2	0	0	0	0	0	0	0	0	0
Ireland	101	7	42	46	40	2	4	3	16	2	0	1
Israel	1225	90	363	543	192	24	5	41	113	19	12	13
Isle of Man	4	0	0	2	1	0	0	0	0	0	0	0
India	430	34	135	87	35	19	1	32	36	7	3	4
Italy	69	6	24	18	26	4	0	2	12	0	0	3
Jersey	2	0	0	0	0	0	0	0	0	0	0	0
Japan	360	24	94	129	47	3	13	45	54	7	1	1
South Korea	31	1	2	36	2	0	0	4	32	1	0	0
Kuwait	7	0	0	0	0	0	0	0	3	0	0	0
Cayman Islands	6	1	3	1	1	1	0	0	2	0	0	0
Kazakhstan	1	0	0	0	0	0	0	0	0	0	0	0
Luxemburg	48	0	18	30	3	4	0	8	22	0	0	0
Malta	5	0	2	1	0	0	0	0	0	0	1	0
Mexico	46	4	7	22	0	2	1	0	13	0	0	0
Malaysia	8	2	1	3	0	0	0	0	3	0	0	0
Nigeria	1	0	0	0	0	0	0	0	2	0	0	0
Netherlands	120	5	49	43	39	4	4	6	12	10	0	1
Norway	268	9	58	101	89	10	0	42	24	0	1	1
New Zealand	18	3	8	1	1	1	0	0	5	2	0	0
Oman	3	4	0	9	0	0	0	0	0	0	0	0
Panama	1	1	0	0	0	0	0	0	0	0	0	0
Peru	2	0	0	2	0	0	0	0	0	0	0	0
Philippines	34	2	6	8	2	1	0	0	0	1	0	0
Pakistan	1	0	0	0	0	0	0	0	0	0	0	0
Poland	15	0	4	1	0	0	0	4	1	0	0	0
Puerto Rico	4	0	3	1	1	0	0	0	1	0	0	0
Portugal	26	0	9	0	11	0	0	0	0	0	0	2
Qatar	3	0	3	2	0	1	0	0	0	0	0	0
Russia	36	0	9	4	9	2	0	0	3	0	0	0
Saudi Arabia	3	0	0	1	0	0	0	1	9	0	0	0
Sweden	153	9	45	79	30	14	0	28	20	1	0	3
Singapore	351	3	98	54	27	2	3	6	79	2	0	0
Thailand	13	0	2	1	0	0	0	1	6	0	0	0
Turkey	10	0	1	2	0	1	0	1	0	0	0	0
Taiwan	25	1	19	17	5	0	0	2	9	0	0	2
United States	17602	1119	5756	7587	2083	211	56	1788	2502	148	107	379
Virgin Islands	1	0	0	0	0	0	0	1	0	0	0	0
South Africa	52	19	15	32	2	8	0	1	21	0	1	0
Total	25362	1711	8190	10277	3720	458	109	2666	3841	284	176	450

Notes: Total count of each cyber risk exposure measure by country, pooled over the entire sample.

D Decomposition by Origin and Destination

Table 6: **CyberExposure by Country of Origin and Regional Destination**

	Origin								
Destination	Russia	China	North Korea	Israel	Iran	United States	Europe	United Kingdom	Total
Africa	0	4	0	36	0	99	49	6	194
Americas	4	10	1	2	0	148	27	22	214
Asia	13	62	0	3	0	81	28	4	191
Europe	6	31	0	10	1	256	224	68	596
United Kingdom	10	14	7	2	0	207	95	107	442
United States	42	158	14	7	15	1318	419	110	2083
Total	75	279	22	60	16	2109	842	317	3720

Notes: Decomposition of CyberExposure$^{\mathrm{T}}$ by country of origin and regional destination. Destination regions are defined as headquarter locations of firms with a positive CyberCountry$^{\mathrm{T}}$ realization. Origin countries are the underlying terms/countries behind the CyberCountry$^{\mathrm{T}}$ topical index.

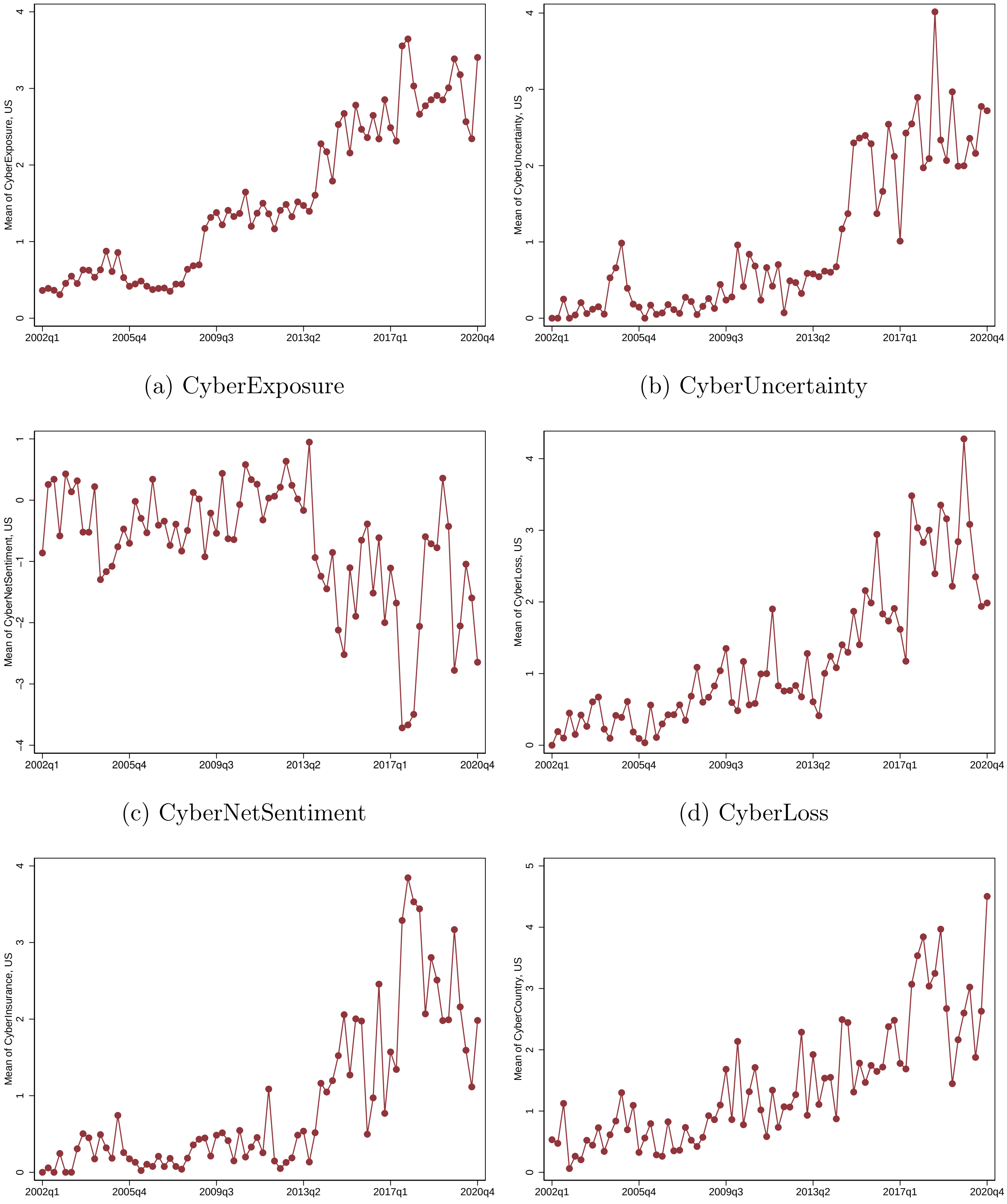
Table 7: **Cyber Risk by Country of Origin and Industry Destination**

	Origin								
Destination	Russia	China	North Korea	Israel	Iran	United States	Europe	United Kingdom	Total
Mining	0	3	0	0	0	9	2	0	14
Manufacturing	23	102	5	16	2	554	183	99	984
Trade	6	16	5	4	0	89	47	22	189
IT	21	89	3	29	2	423	161	45	773
Finance	0	12	3	1	1	182	116	31	346
Real Estate	0	1	0	0	0	35	0	0	36
Services	13	49	6	10	11	621	306	84	1100
Education	0	0	0	0	0	13	0	0	13
Health	0	0	0	0	0	4	0	2	6
Other	12	7	0	0	0	179	27	34	259
Total	75	279	22	60	16	2109	842	317	3720

Notes: Decomposition of CyberExposure$^{\text{T}}$ by country of origin and sectoral destination. Destination industries are defined by the 2-digit NAICS code of firms with a positive CyberCountry$^{\text{T}}$ realization. Origin countries are the underlying terms/countries behind the CyberCountry$^{\text{T}}$ topical index.

E Additional Time-Series Plots

Figure 1: **United States**



(a) CyberExposure

(b) CyberUncertainty

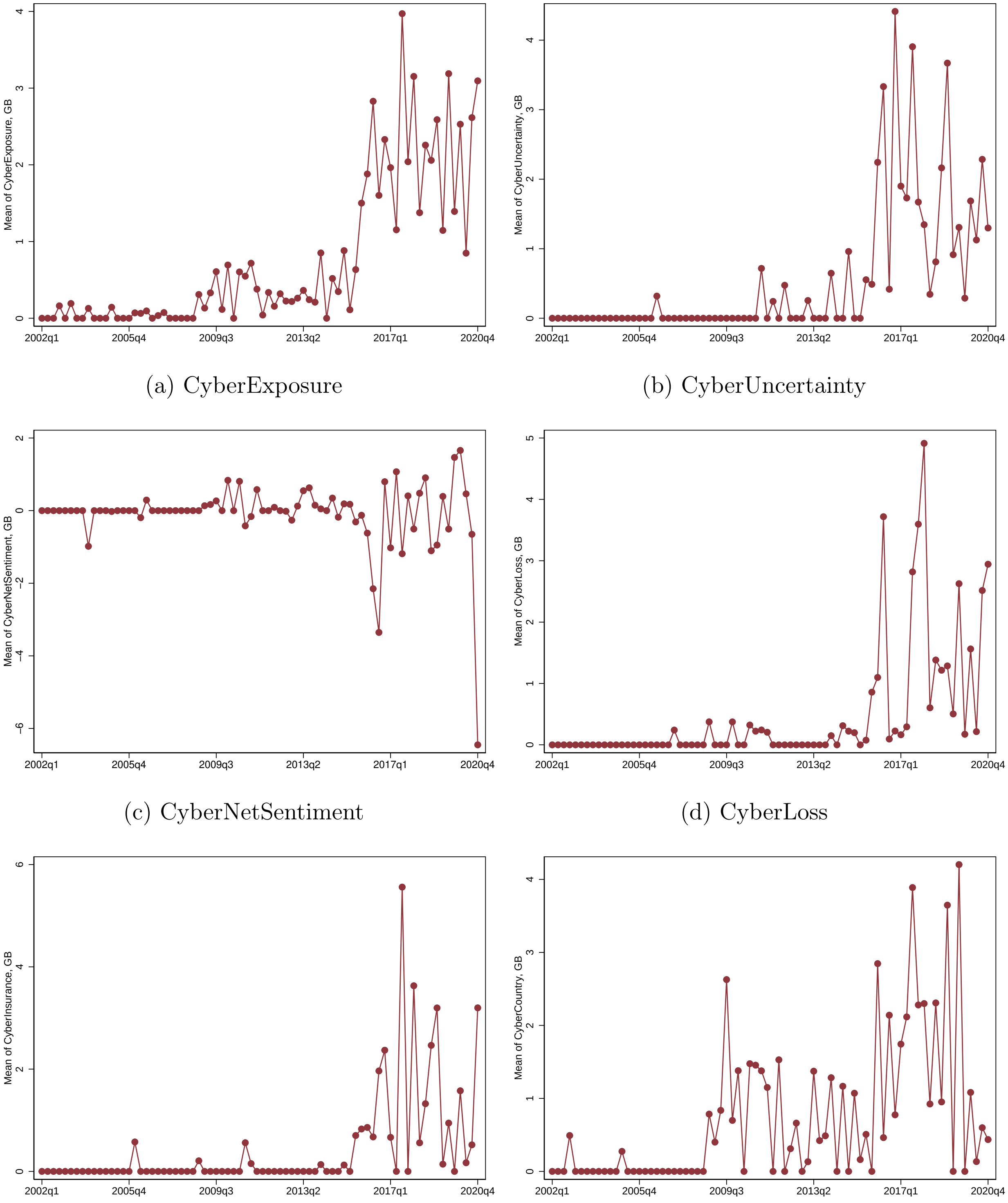
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 2: **United Kingdom**



(a) CyberExposure

(b) CyberUncertainty

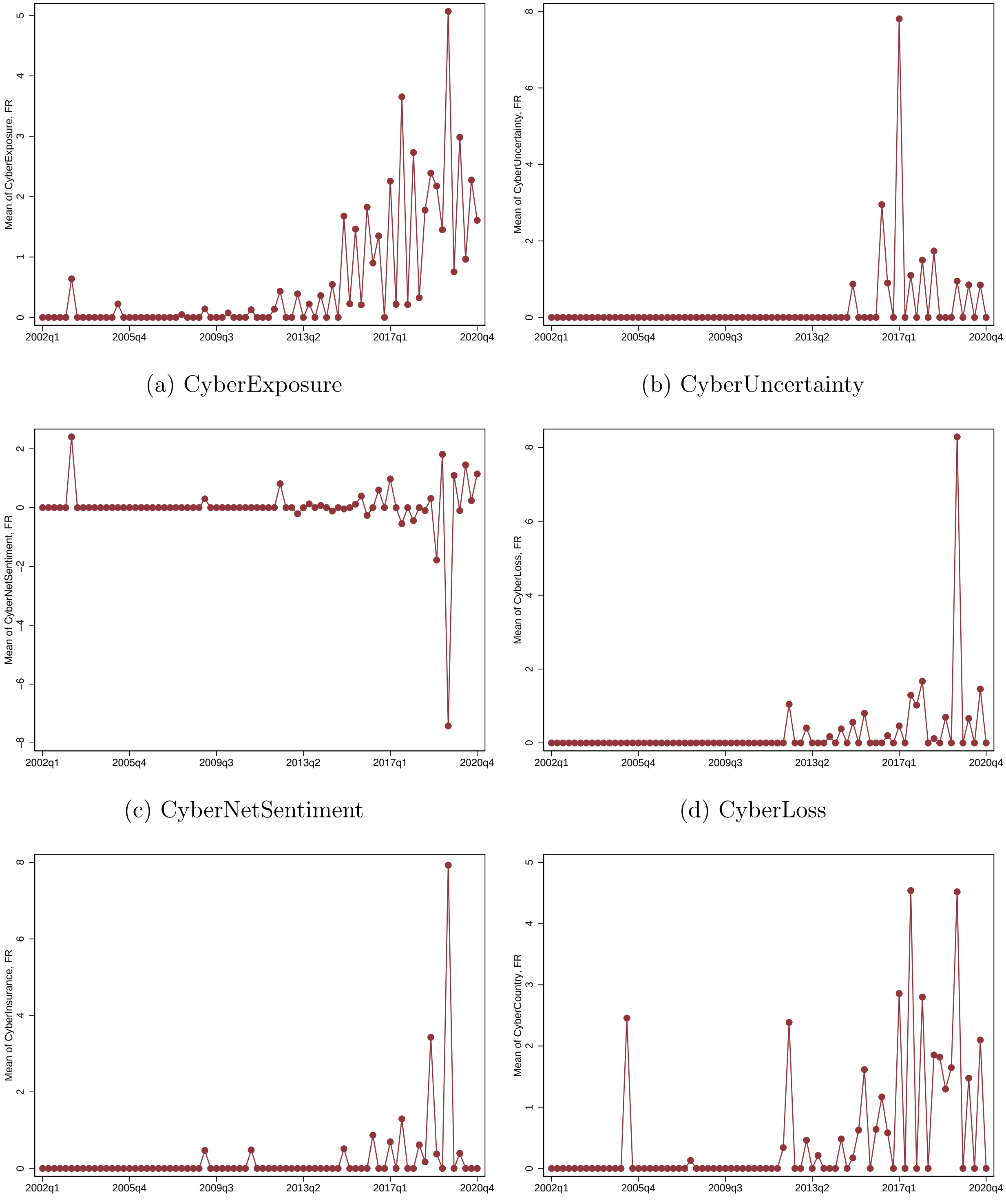
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 3: **France**



(a) CyberExposure

(b) CyberUncertainty

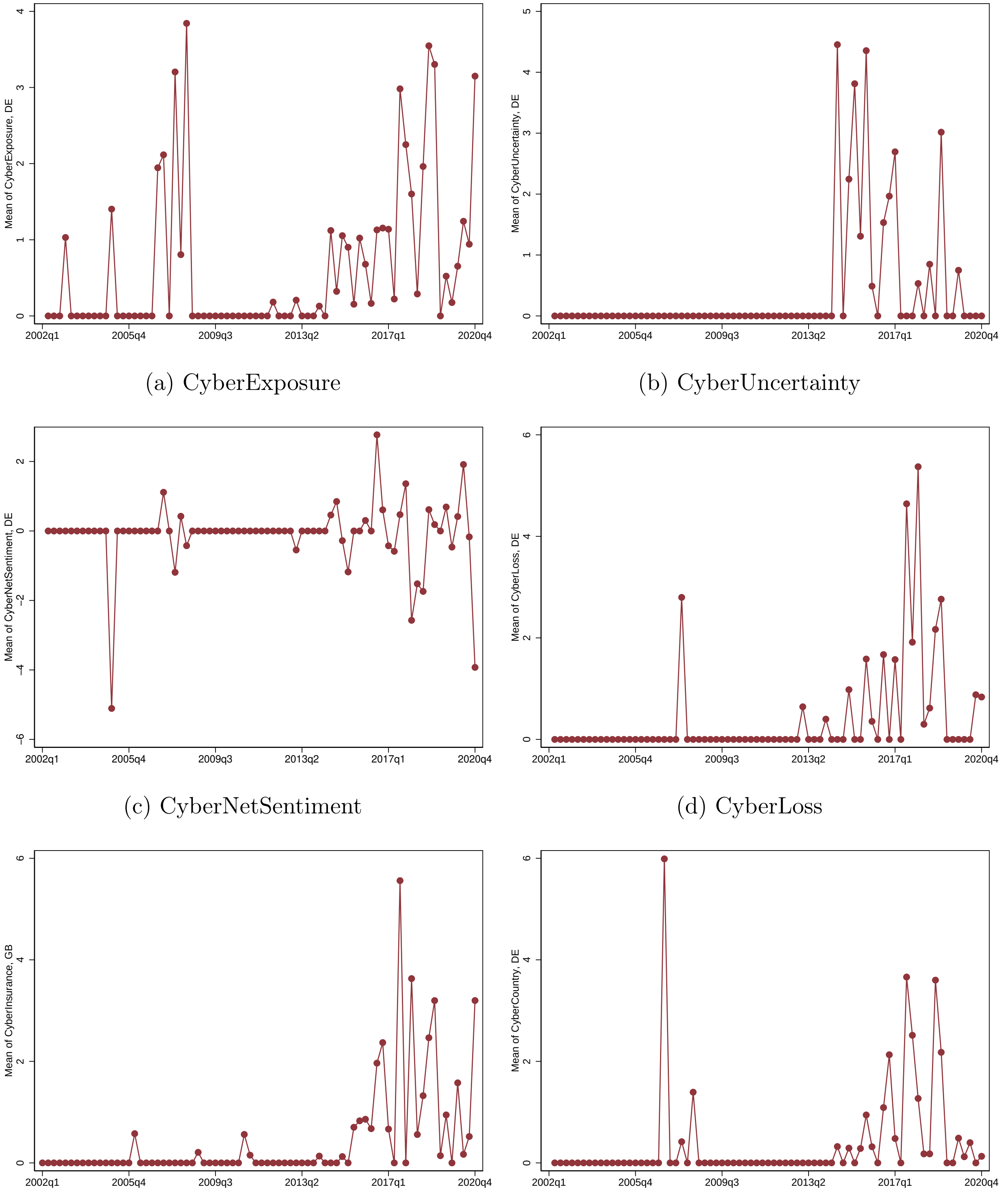
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 4: **Germany**



(a) CyberExposure

(b) CyberUncertainty

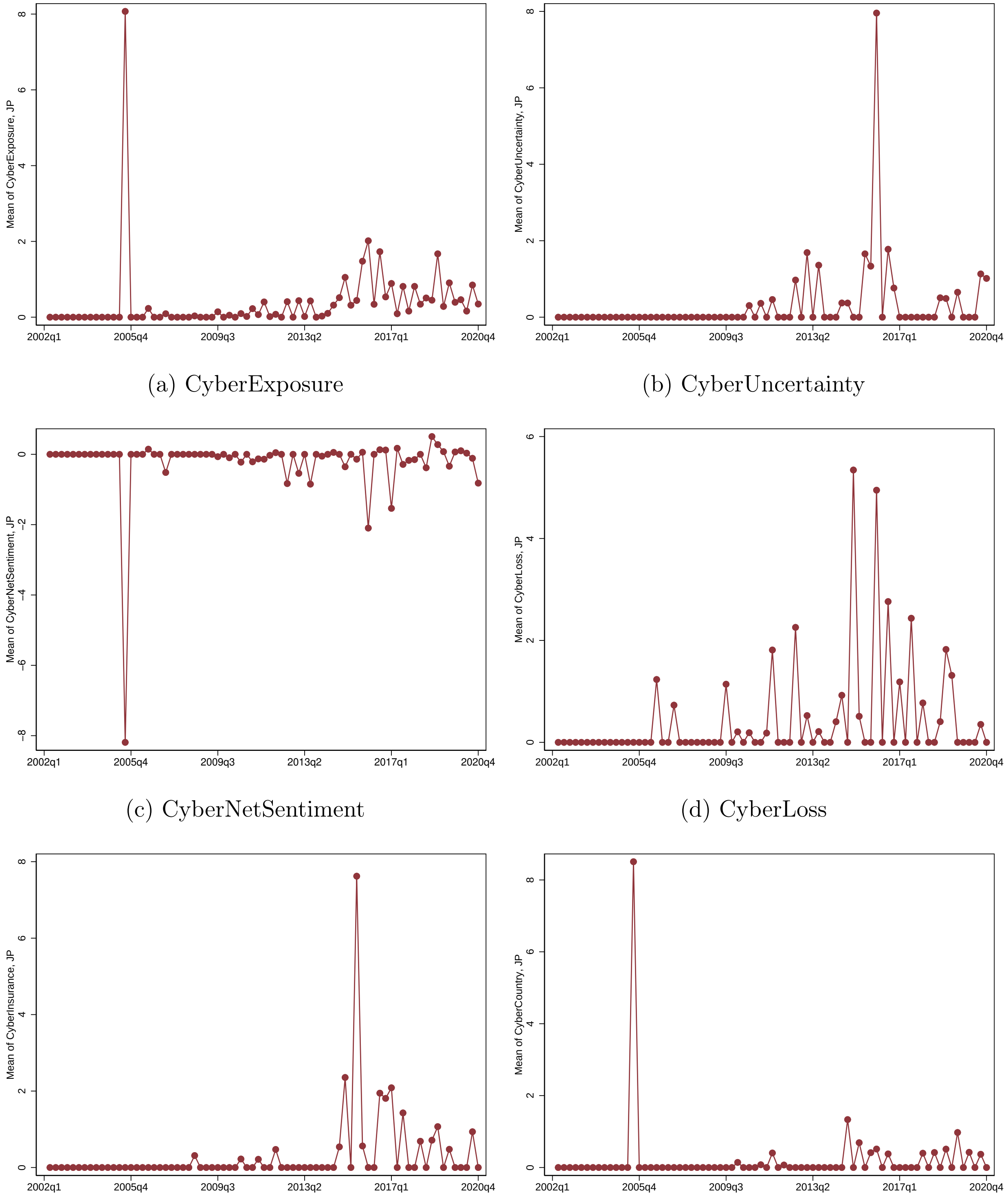
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 5: **Japan**



(a) CyberExposure

(b) CyberUncertainty

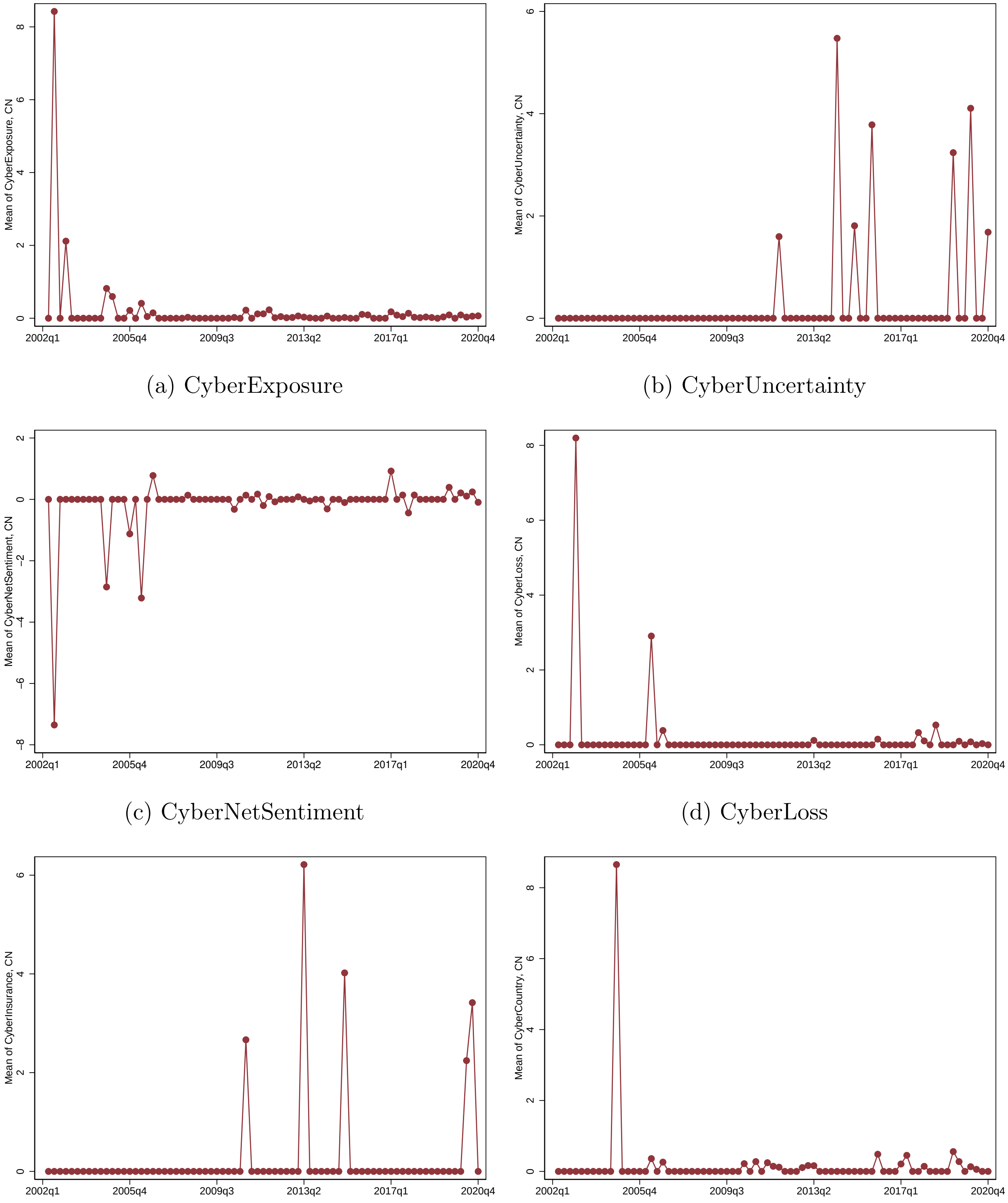
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 6: **China**



(a) CyberExposure

(b) CyberUncertainty

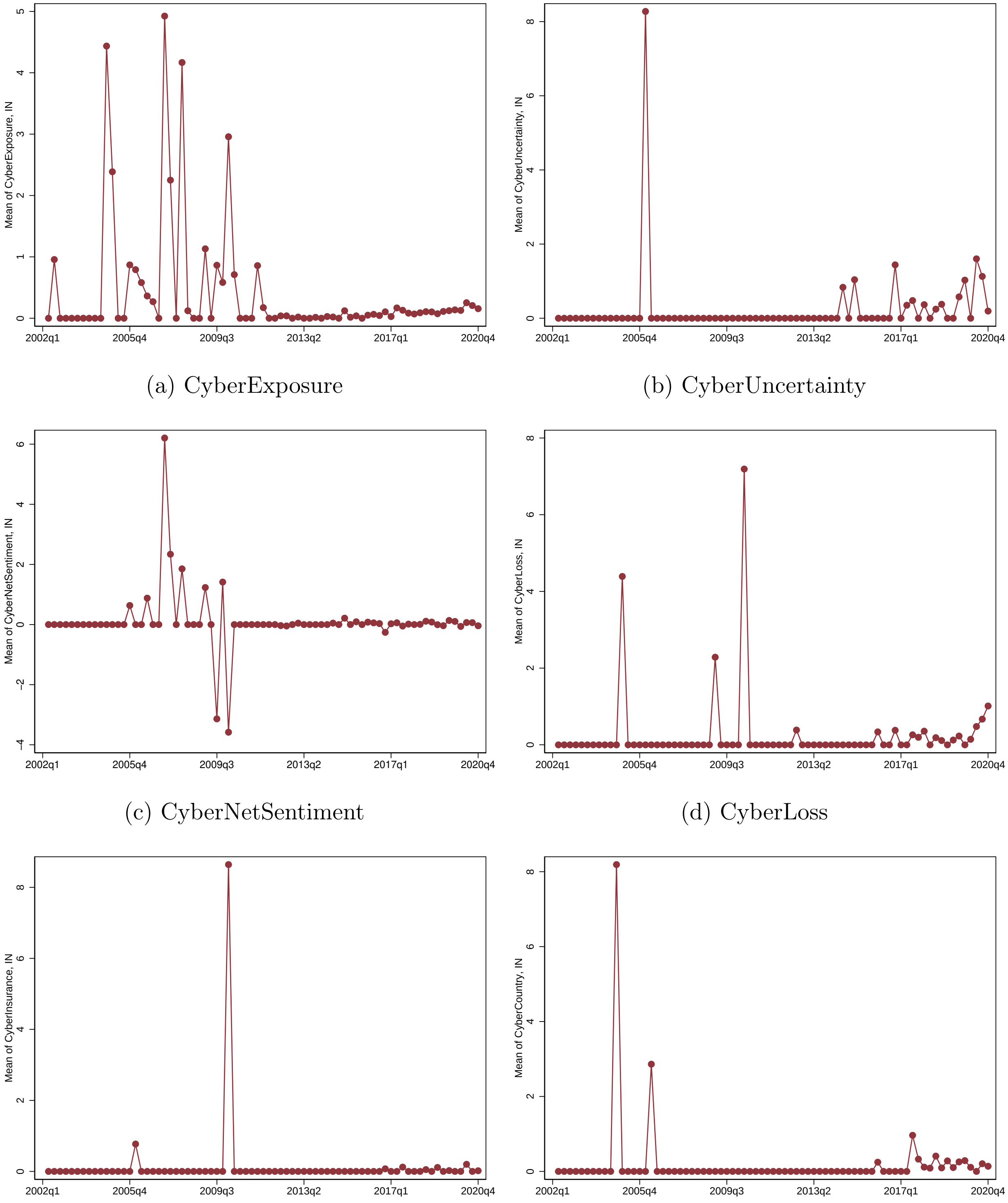
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 7: **India**



(a) CyberExposure

(b) CyberUncertainty

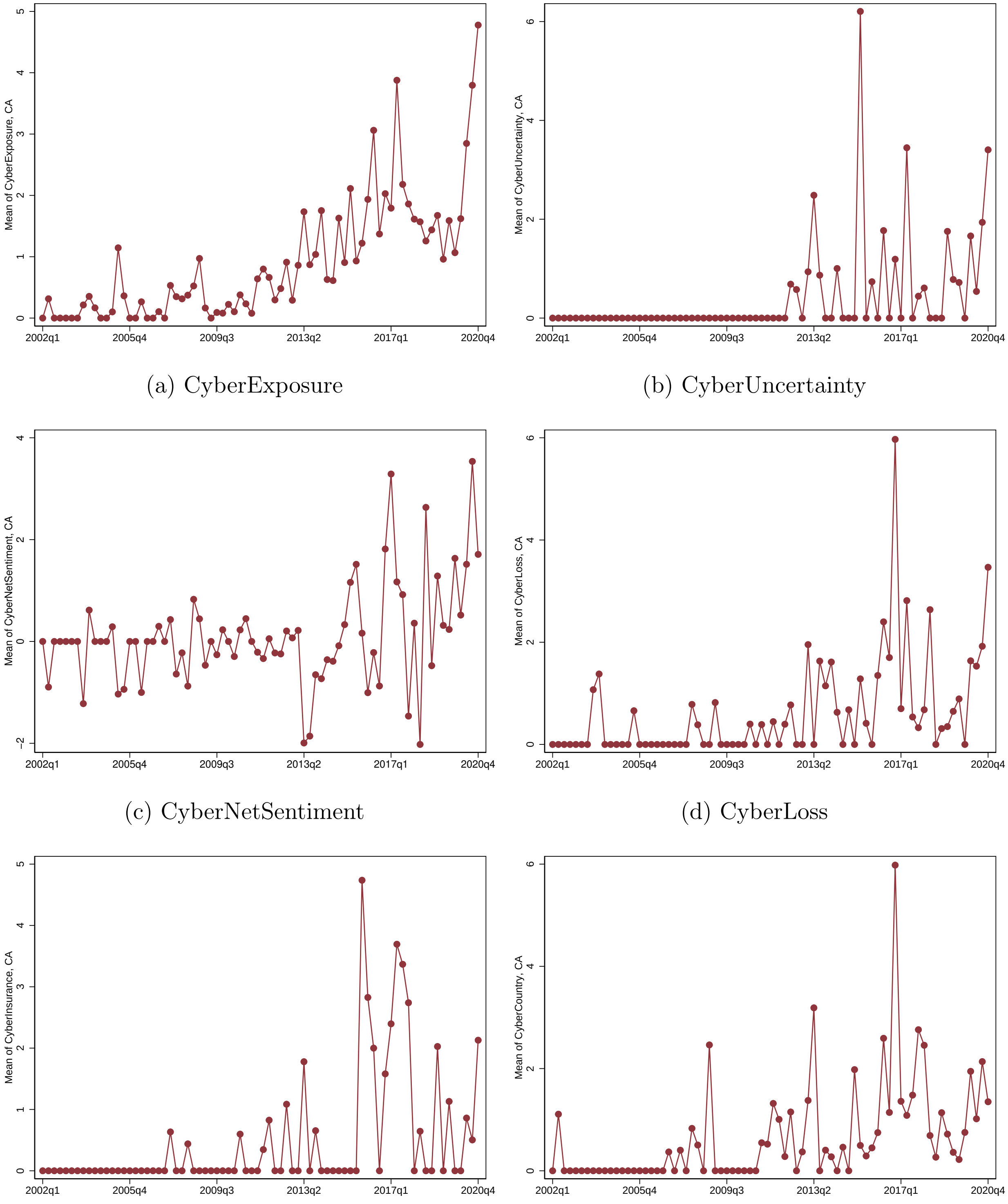
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 8: **Canada**



(a) CyberExposure

(b) CyberUncertainty

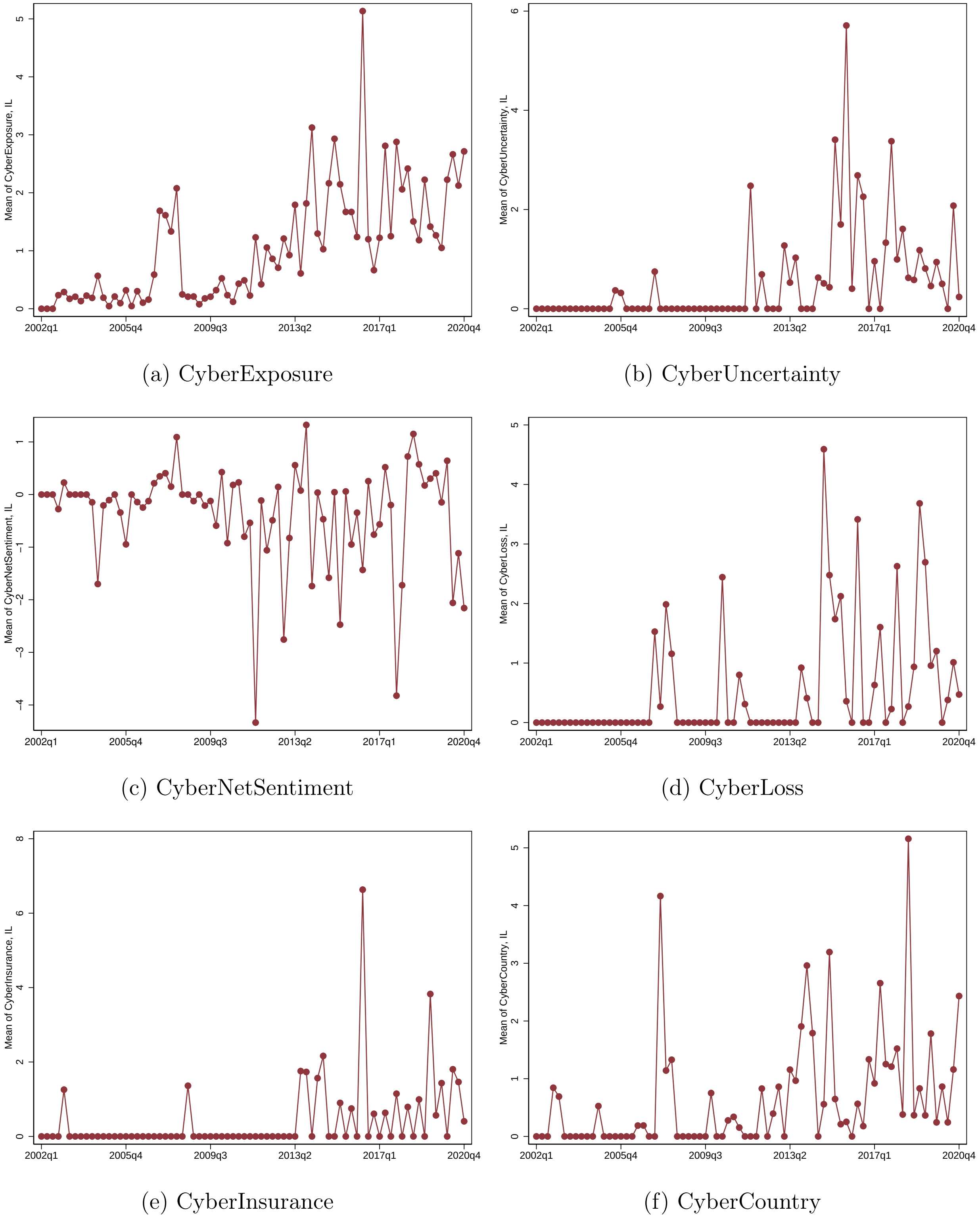
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 9: **Israel**



(a) CyberExposure

(b) CyberUncertainty

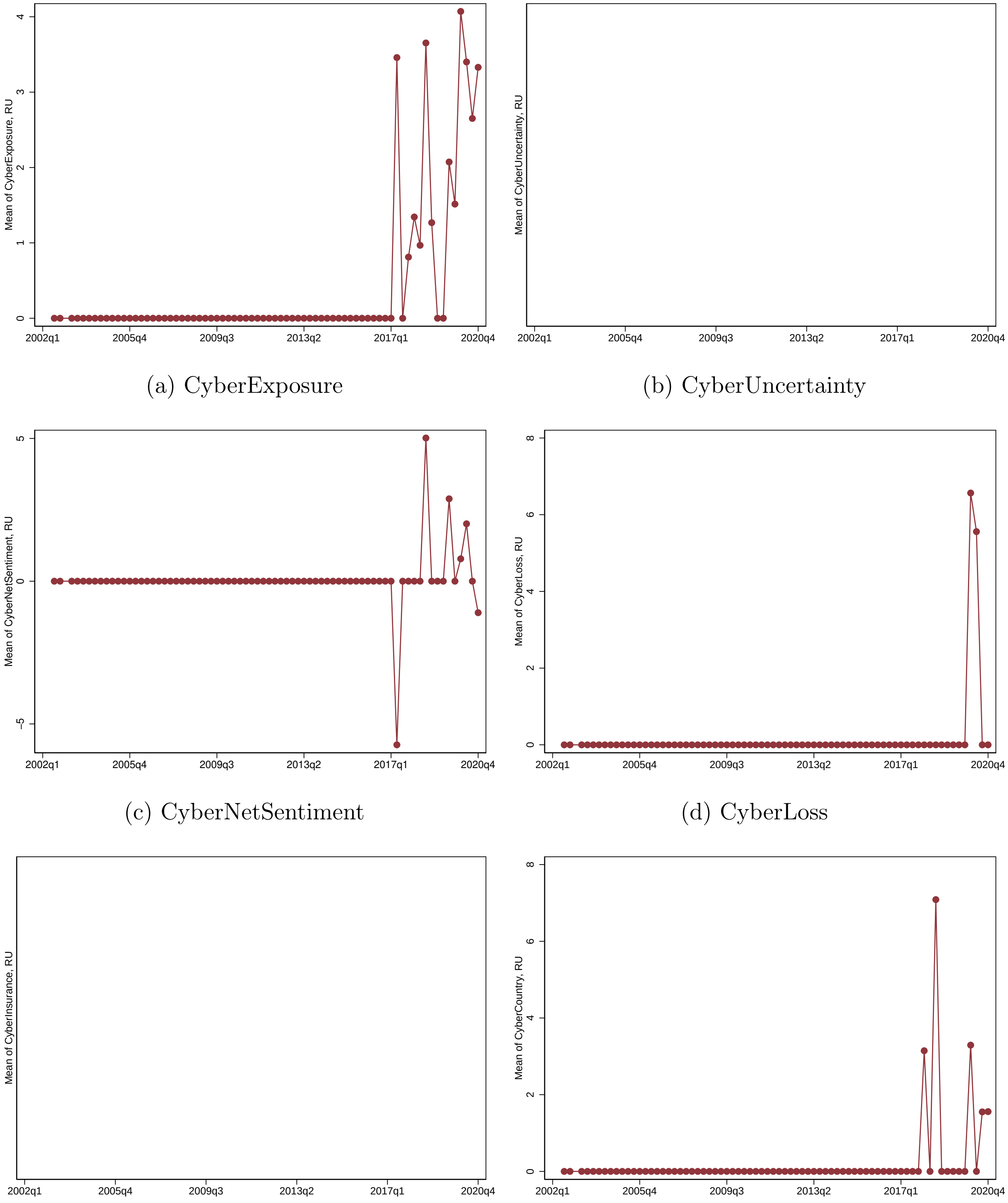
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 10: **Russia**



(a) CyberExposure

(b) CyberUncertainty

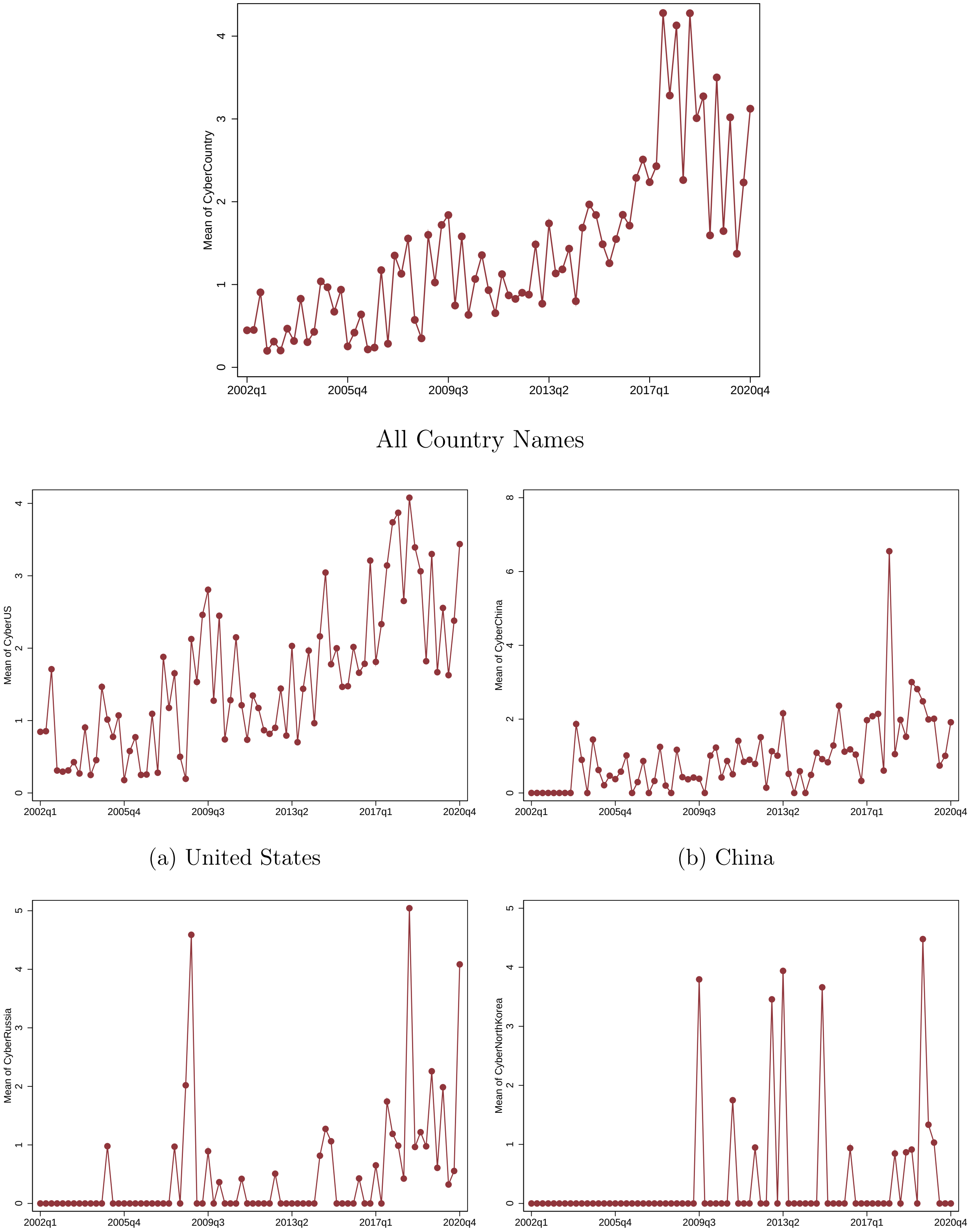
(c) CyberNetSentiment

(d) CyberLoss

(e) CyberInsurance

(f) CyberCountry

Figure 11: **Country Names Topic**



All Country Names

(a) United States

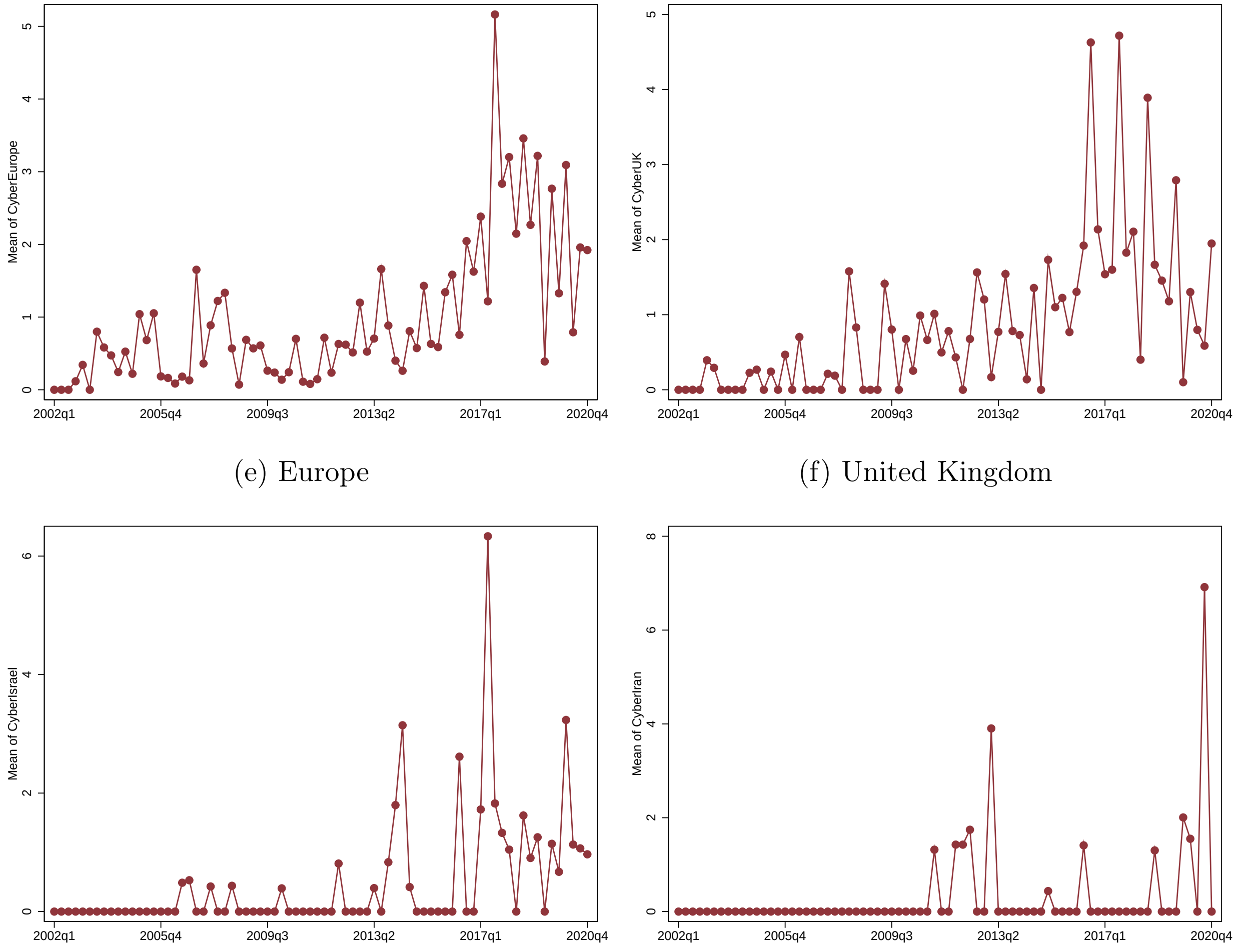
(b) China

(c) Russia

(d) North Korea

Notes: Time-series averages of CyberCountry_{it}, CyberUS_{it}, CyberChina_{it}, CyberRussia_{it}, and $\text{CyberNorthKorea}_{it}$, each normalized by the standard deviation of the respective sample. All terms behind the "Country Names" topic are listed in Table 2 of the Online Appendix.

Figure 11: **Country Names (Continued)**



(e) Europe

(f) United Kingdom

(g) Israel

(h) Iran

Notes: Time-series averages of CyberEurope_{it}, CyberUK_{it}, CyberIsrael_{it}, and CyberIran_{it}, each normalized by the standard deviation of the respective sample. All terms behind the "Country Names" topic are listed in Table 2 of the Online Appendix.

F Additional Firm-Level Results by Industry and Region

Table 8: **Determinants of Firm-Level Cyber Risk Exposure: Regions**

Probit Model	Dependent Variable: Firm-level Indicator of Cyber Risk Exposure					
	(1)	(2)	(3)	(4)	(5)	(6)
	United States	Americas Ex-US	Europe	United Kingdom	Asia	Africa
PP&E / Assets	0.2693	-0.0179	0.5791	-0.2006	-1.6575***	-10.0078
	(0.2527)	(0.8350)	(1.4275)	(1.9750)	(0.6008)	(6.6569)
Intangibles / Assets	0.8568***	-0.3332	2.5381**	-0.4393	0.0636	-4.4473
	(0.1729)	(0.8208)	(1.2149)	(1.7011)	(0.9247)	(4.2783)
CAPEX / Assets	-0.3800*	2.5689***	-2.3027***	-1.8759	-1.2944	9.6864***
	(0.2142)	(0.9540)	(0.8268)	(1.5130)	(0.7886)	(3.6581)
Cash Flow / Assets	-1.6122***	-4.4656	4.5874	8.0047	9.3740	14.4234
	(0.6204)	(6.5324)	(7.7241)	(12.9419)	(11.4642)	(29.7561)
Long-Term Debt / Assets	0.0501	-0.2102	0.3873	1.7834***	0.7845**	0.1472
	(0.0823)	(0.3044)	(0.3651)	(0.6655)	(0.3753)	(1.2680)
Liquidity Ratio	0.5881***	0.6337	1.1695	-1.0207	-0.6485	3.2307
	(0.2057)	(0.6837)	(1.4689)	(2.4572)	(0.8527)	(3.1007)
Log (Size)	0.1106***	0.2240***	0.1253	-0.5650***	-0.1117	0.1816
	(0.0174)	(0.0739)	(0.0952)	(0.2018)	(0.0897)	(0.1855)
Debt / Assets	-0.0188	-3.0907***	1.5974	0.3946	-0.2939	-7.9700*
	(0.1441)	(0.8114)	(1.0172)	(0.8592)	(0.8180)	(4.2668)
Log (Age)	0.0006	0.4917**	-0.1952	1.0450***	0.8430***	1.8205***
	(0.0479)	(0.2127)	(0.2375)	(0.3503)	(0.2731)	(0.4466)
Equity Net Issuance	0.0454	-0.8501	2.5557***	-10.1795*	4.0236***	2.5552*
	(0.3317)	(1.7632)	(0.9726)	(5.6462)	(1.0263)	(1.4864)
ROA	0.6658	1.5203	-4.4939	-3.2787	-6.0291	-27.2869
	(0.4716)	(6.0834)	(7.1255)	(13.7385)	(10.1332)	(34.9933)
Sales / Assets	0.9709	3.8150	2.3760	-4.7598	-3.0906	12.7727
	(0.7419)	(4.2300)	(4.5949)	(8.7368)	(4.9551)	(9.5821)
Tobin's Q	0.0154	-0.2282***	-0.0911*	0.1670*	0.0541	-0.4571
	(0.0161)	(0.0757)	(0.0546)	(0.0919)	(0.0524)	(0.9531)
Book to Market Equity	-0.0723	-0.2536	-0.2241	-2.8033**	-0.3983*	-8.5383
	(0.0602)	(0.2111)	(0.2790)	(1.2463)	(0.2149)	(6.2104)
Operational Costs / Assets	-0.6881	-3.1742	0.6274	-1.3108	1.6506	-27.3944**
	(0.7529)	(4.4633)	(5.1788)	(9.8665)	(4.8935)	(13.5214)
Market Beta	-0.0302	-0.1808	0.0458	-0.4378**	-0.3113	-2.9950***
	(0.0326)	(0.1773)	(0.1059)	(0.2191)	(0.1918)	(0.8183)
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	74844	2535	991	875	2153	564
pseudo R2	0.227	0.365	0.294	0.291	0.265	0.642

Notes: Probit regression of firm-level indicators of positive cyber risk total exposure on various balance sheet and income statement characteristics. Each column corresponds to a case where all but the relevant region are dropped. Regions are defined by the locations of firm headquarters. All variables are defined in Table 3 of the Online Appendix. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 9: **Determinants of Firm-Level Cyber Risk Exposure: Industries**

Probit Model	Dependent Variable: Firm-level Indicator of Cyber Risk Exposure								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Mining	Manufacturing	Trade	IT	Finance	Real Estate	Services	Health	Other
PP&E / Assets	0.4261	-0.9766	-0.9012**	-0.7443	-0.7800	-1.2688	0.5156	-1.4666	0.0773
	(1.1174)	(0.7408)	(0.3582)	(0.5222)	(0.8379)	(1.1550)	(0.8465)	(1.2345)	(0.2836)
Intangibles / Assets	3.4610**	0.4925	-0.2132	0.3066	0.8724***	0.4205	2.7902***	-0.0502	0.5116
	(1.6515)	(0.3347)	(0.5405)	(0.4061)	(0.2888)	(0.7122)	(0.6483)	(1.3208)	(0.4965)
CAPEX / Assets	-2.8259**	-1.4222***	-0.5881	-1.5391***	0.0574	-0.1351	-0.6924	-2.5569**	-0.7312*
	(1.1464)	(0.4074)	(0.5090)	(0.5144)	(0.2538)	(1.7931)	(0.6153)	(1.1545)	(0.4205)
Cash Flow / Assets	15.9067**	1.2269	15.3783**	4.3228	8.5459	3.4542	-1.9119*	4.6469	-17.2684***
	(6.9316)	(2.9654)	(6.5651)	(3.9409)	(9.4167)	(10.2867)	(0.9776)	(4.9785)	(5.6825)
Long-Term Debt / Assets	1.7727*	0.2017	0.1711	0.3063	0.0534	2.4799	-0.1473	1.4109***	0.0833
	(0.9071)	(0.1465)	(0.2198)	(0.2357)	(0.1609)	(1.8060)	(0.1606)	(0.5056)	(0.3074)
Liquidity Ratio	0.2437	0.7886**	-0.6375	1.2806**	-0.4919	1.0079	1.7040***	-1.7507	-0.8004
	(1.6840)	(0.3770)	(0.6746)	(0.5028)	(0.4235)	(2.0852)	(0.6388)	(1.3316)	(0.7177)
Log (Size)	-0.6957***	0.0966**	0.1891***	0.0376	0.1303***	0.2819	0.2432***	0.0897	0.1482***
	(0.2598)	(0.0379)	(0.0588)	(0.0303)	(0.0293)	(0.2153)	(0.0608)	(0.0948)	(0.0438)
Debt / Assets	-0.9360	-0.5540**	-0.3011	0.0706	-0.1695	-0.4701	-0.2812	-0.0442	-0.7059*
	(0.7688)	(0.2685)	(0.3759)	(0.2599)	(0.2258)	(1.3578)	(0.3872)	(0.8555)	(0.4056)
Log (Age)	0.2701	0.1865*	0.3603*	0.0082	0.1219	0.1354	0.0569	0.0616	0.0369
	(0.2012)	(0.0977)	(0.2040)	(0.1069)	(0.0807)	(0.4757)	(0.1463)	(0.2160)	(0.1221)
Equity Net Issuance	-4.0706***	0.1035	1.8419**	0.4730	-0.0362	-3.5114	-1.0812	3.3807**	0.7668
	(1.4240)	(0.4752)	(0.9036)	(0.4988)	(0.9786)	(7.2626)	(0.8604)	(1.5486)	(0.8052)
ROA	-20.4156***	-0.9962	-14.5807**	-5.1689	-7.8394	-2.5313	0.8468*	1.7939	15.9682***
	(7.5097)	(2.8670)	(6.5285)	(3.6397)	(9.3878)	(10.0717)	(0.4620)	(3.6183)	(4.8366)
Sales / Assets	3.6349	0.0467	2.7496	0.1309	1.5701	1.3771	2.0246	-8.0008	3.7970
	(2.2894)	(1.2660)	(1.6946)	(2.0726)	(2.9560)	(10.2076)	(1.3741)	(7.2250)	(3.1351)
Tobin's Q	0.3706*	-0.0525	0.0105	-0.0103	0.1831***	0.0200	0.0389	-0.0847	0.0063
	(0.2107)	(0.0319)	(0.0641)	(0.0285)	(0.0467)	(0.2757)	(0.0271)	(0.1441)	(0.0584)
Book to Market Equity	-0.3077	-0.0637	0.0832	0.0547	-0.1099	-0.1146	0.1394	-1.0706***	-0.3355*
	(0.3341)	(0.1178)	(0.1218)	(0.0811)	(0.1040)	(0.2673)	(0.1794)	(0.3754)	(0.1993)
Operational Costs / Assets	-7.9272***	-0.0325	-2.9209*	-0.8500	-1.9283	2.0938	0.0519	5.9477	-3.7146
	(2.3918)	(1.0790)	(1.7108)	(1.9146)	(2.9729)	(12.0442)	(1.4640)	(7.1584)	(3.0948)
Market Beta	0.0279	0.0021	0.0001	0.0208	-0.1185	-0.4013	-0.1266	0.0880	0.1206
	(0.1377)	(0.0653)	(0.0900)	(0.0916)	(0.0819)	(0.2604)	(0.0807)	(0.2158)	(0.1068)
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1473	32374	8777	5382	8904	310	5991	416	7104
pseudo R2	0.405	0.098	0.108	0.148	0.159	0.199	0.204	0.149	0.121

Notes: Probit regression of firm-level indicators of positive cyber risk total exposure on various balance sheet and income statement characteristics. Each column corresponds to a case where all but the relevant industry are dropped. Industries are defined by the NAICS 2-digit classification. All variables are defined in Table 3 of the Online Appendix. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

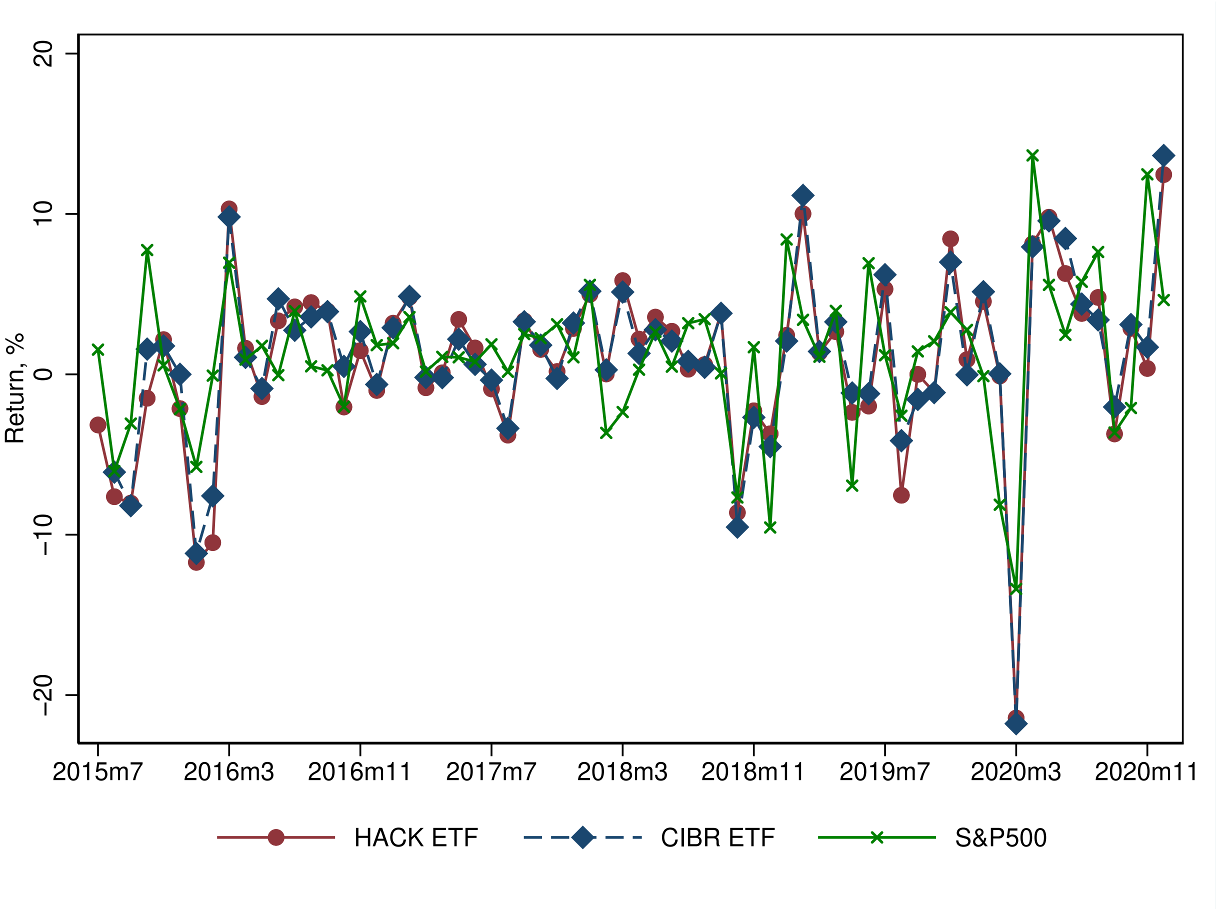
Table 10: **Determinants of Firm-Level Cyber Risk Exposure: Finance Sub-Industries**

Probit Model	Dependent Variable: Firm-level Indicator of Cyber Risk Exposure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Banks	Non-Banks	Other Intermediaries	Broker-Dealers	Insurance	Funds and Trusts
PP&E / Assets	18.9672***	-0.8489	3.8343	-8.4711**	-0.4905	3.5857
	(5.3082)	(1.5323)	(2.9196)	(3.5930)	(3.0650)	(4.3517)
Intangibles / Assets	-4.6806	-2.5572	1.5396***	-0.6894*	2.3110***	0.5445
	(3.2251)	(1.9411)	(0.4135)	(0.3547)	(0.4669)	(2.0341)
CAPEX / Assets	0.9850	-0.2973	0.1091	0.3106	-1.7334**	-0.5995
	(1.6204)	(1.8141)	(0.7489)	(0.4574)	(0.7243)	(1.0531)
Cash Flow / Assets	-18.3614	5.8184	24.7099	21.7942	18.8505	-456.0154
	(42.4829)	(18.1018)	(15.7052)	(23.1973)	(15.2930)	(285.2190)
Long-Term Debt / Assets	0.1564	1.2387	0.0974	0.0388	-0.2629	0.4733
	(0.2331)	(1.7844)	(0.3604)	(0.3423)	(0.3843)	(0.6530)
Liquidity Ratio	0.9609	0.0017	2.3114**	-0.7863	-0.2236	-6.2763*
	(0.6115)	(0.8209)	(0.9160)	(0.7560)	(0.6618)	(3.3330)
Log (Size)	0.1486***	-0.0392	0.2374***	0.1027**	-0.0470	0.2457***
	(0.0422)	(0.1519)	(0.0889)	(0.0437)	(0.0746)	(0.0946)
Debt / Assets	-0.0078	1.6538	1.2113**	-1.1188**	2.6198***	-2.6467**
	(1.2667)	(1.2835)	(0.5567)	(0.5033)	(0.7882)	(1.1466)
Log (Age)	0.0564	0.1236	0.1397	0.2422**	0.5623***	-0.1106
	(0.0618)	(0.2289)	(0.1373)	(0.1083)	(0.1726)	(0.2020)
Equity Net Issuance	-8.5545	-8.0881**	-1.6306	-0.1693	-10.0841	-4.7447
	(16.5671)	(3.8399)	(1.7780)	(1.2016)	(6.5200)	(5.0365)
ROA	23.3743	-9.9734	-20.1027	-22.0776	-25.7502	458.7112
	(45.8820)	(17.7114)	(12.3781)	(22.9903)	(17.7464)	(285.5782)
Sales / Assets	-5.9635	-34.3853**	-9.5958*	6.5481**	9.3171	-30.9828*
	(31.1042)	(15.2753)	(5.6564)	(3.3235)	(6.5670)	(17.9186)
Tobin's Q	-0.9130	-0.0890	0.2744***	0.0289	0.5436***	1.0259
	(1.2631)	(0.2491)	(0.0870)	(0.1230)	(0.1860)	(0.7011)
Book to Market Equity	-0.5981**	-0.3296	0.5503**	0.2214**	0.1445	0.7030**
	(0.2790)	(0.2345)	(0.2806)	(0.0965)	(0.2064)	(0.2767)
Operational Costs / Assets	-50.9772	45.8612**	12.5184**	-6.5073**	-12.0841*	33.5803*
	(40.3495)	(20.4040)	(5.8542)	(3.2787)	(7.0718)	(19.6969)
Quarter Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4543	184	594	1018	1235	409
pseudo R2	0.107	0.205	0.223	0.125	0.367	0.247

Notes: Probit regression of firm-level indicators of positive cyber risk total exposure on various balance sheet and income statement characteristics. Each column corresponds to a case where all but the relevant industry are dropped. Industries are defined by the NAICS 4-digit classification. All variables are defined in Table 3 of the Online Appendix. Standard errors are clustered at the firm level. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

G Cybersecurity ETFs

Figure 12: **Cybersecurity ETFs and Market Returns**



Notes: Monthly returns on the HACK and CIBR ETFs and the S&P500. The Pearson correlation coefficients between the HACK and CIBR ETFs with the market factor are 0.6299 and 0.6440 with the p-values of (0.000) and (0.000), respectively.

Table 11: **Determinants of Cybersecurity ETF Returns**

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	HACK ETF	100%	98%	61%	14%	49%	0%	6%	-6%	14%	8%	20%	6%	23%	1%
(2)	CIBR ETF	98%	100%	64%	13%	48%	-1%	9%	-6%	15%	7%	14%	2%	23%	5%
(3)	Market	61%	64%	100%	22%	34%	4%	6%	-5%	8%	4%	6%	1%	1%	6%
(4)	HML	14%	13%	22%	100%	18%	-5%	-10%	2%	-6%	-10%	-4%	-1%	-3%	-1%
(5)	SMB	49%	48%	34%	18%	100%	-2%	-5%	1%	-6%	-5%	0%	-3%	10%	3%
(6)	CyberExposure	0%	-1%	4%	-5%	-2%	100%	70%	41%	19%	30%	73%	82%	36%	24%
(7)	CyberUncertainty	6%	9%	6%	-10%	-5%	70%	100%	21%	17%	20%	54%	59%	26%	13%
(8)	CyberCountry	4%	2%	-7%	22%	13%	67%	28%	100%	-13%	0%	39%	62%	43%	-4%
(9)	CyberCrypto	14%	15%	8%	-6%	-6%	19%	17%	4%	100%	42%	22%	21%	23%	3%
(10)	CyberGlobal	8%	7%	4%	-10%	-5%	30%	20%	7%	42%	100%	40%	34%	4%	12%
(11)	CyberInsurance	20%	14%	6%	-4%	0%	73%	54%	32%	22%	40%	100%	75%	16%	9%
(12)	CyberLoss	6%	2%	1%	-1%	-3%	82%	59%	51%	21%	34%	75%	100%	39%	16%
(13)	CyberPolitics	23%	23%	1%	-3%	10%	36%	26%	14%	23%	4%	16%	39%	100%	44%
(14)	CyberSocial	1%	5%	6%	-1%	3%	24%	13%	9%	3%	12%	9%	16%	44%	100%

Notes: HACK and CIBR ETF pricing data was obtained from Bloomberg on March 09, 2021. The two daily time series of prices are aggregated to the monthly frequency and log-differenced. Market, HML, and SMB are the Fama and French (1993) three factors. All other variables are our cyber risk exposure indices, which we aggregate to the monthly frequency and log-difference.

References

Fama, E. and K. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1).

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