Dal Moro, Eric

Article

Towards an economic cyber loss index for parametric cover based on IT security indicator: A preliminary analysis

Risks

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Dal Moro, Eric (2020) : Towards an economic cyber loss index for parametric cover based on IT security indicator: A preliminary analysis, Risks, ISSN 2227-9091, MDPI, Basel, Vol. 8, Iss. 2, pp. 1-12, https://doi.org/10.3390/risks8020045

This Version is available at:
https://hdl.handle.net/10419/257999

risks

MDPI

Article

Towards an Economic Cyber Loss Index for Parametric Cover Based on IT Security Indicator: A Preliminary Analysis

Eric Dal Moro

SCOR Reinsurance Company, CH-8022 Zürich, Switzerland; Edalmoro@scor.com

Received: 27 January 2020; Accepted: 29 April 2020; Published: 8 May 2020

Abstract: As cyber events have virtually no geographical limitations and can result in economic losses on a global scale, the assessment of return periods for such economic losses is currently debated among experts. The potential accumulation of consequential insurance losses due to intrusions or viruses is one of the major reasons why the (re-)insurance industry has limited risk appetite for cyber related risks. In order to increase the risk appetite for cyber risk and based on a first batch of data provided by Symantec, the goal of this article is to: Check if IT activity, i.e., the number of virus or intrusions being blocked by Norton on end-user computers could be used as an index for parametric covers that reinsurance companies could propose to their cedants; Look into the correlations of this IT activity across different regions, thereby confirming the absence of geographical limitations for cyber risk, and hence confirming the systemic nature of this risk. This first study on the Symantec dataset shows that a cyber index based on IT activity could be a useful tool to design parametric reinsurance product.

Keywords: correlation; time-series; cyber risk; insurance linked securities; parametric insurance

1. Introduction

Cyber risk commonly refers to any risk of financial loss, disruption or damage to the reputation of an organization, resulting from the failure of its information technology systems. In order to protect organization or persons from cyber risk, a few companies propose protection tools such as Symantec, AVG or Microsoft. Among these companies, Symantec is one of the major providers of end-user (personal computer) protection with its software Norton. For this article, Symantec has provided a first batch of data that will help characterize the cyber risk, and could be a source for deriving an index of cybercrime activity.

In its 2018 Internet Crime Report (see Federal Bureau of Investigation 2018), the USA Federal Bureau of Investigation describes the mission of its IC3 department as "provide the public with a reliable and convenient reporting mechanism to submit information to the FBI concerning suspected Internet-facilitated criminal activity, and to develop effective alliances with industry partners. Information is analyzed and disseminated for investigative and intelligence purposes, for law enforcement, and for public awareness". The report shows a steady increase of Internet crime, totaling, in 2018, 1,509,679 complaints for a total loss estimated at USD 7.45 billion. As an illustration of the growing crime activity on the internet, Guy Carpenter and Cybercube mentioned (see Guy Carpenter 2019) that:

- The U.S. industry 1-in-100 years return period produces total annual cyber catastrophe insured losses of USD 14.6 billion (this can include one or more events within the same year);

- Both on-premise and cloud service providers face exogenous threats from malicious third parties. Focusing on cloud service providers, the calculated probability of ransomware is four times larger than the probability of other outages;
- The costliest cyber catastrophe scenario is widespread data loss from a leading operating systems provider with potential to generate up to USD 23.8 billion of insured loss;
- The most likely cyber catastrophe loss scenario is widespread data theft from a major email service provider.

In the context of the significant amount of losses already existing and of the lack of cyber risk characterization, insurance and reinsurance companies are still hesitating to propose covers which would be needed by the industry (see Abadie 2019). There have been many attempts to answer the lack of knowledge on cyber risk through scenarios (see Swiss Insurance Association (Switzerland) 2017), or through an agenda to work on this topic (see Falco et al. 2019). However, to this day, most of the stakeholders on this market still recognize the complexity of mastering this risk. This results in a lack of capacity provided to this market.

One way to increase the available cyber capacity of the risk transfer market, and to achieve an additional atomization of this accumulation exposure, are alternative risk transfer (ART) solutions, involving the capital market via insurance linked securities (ILS), i.e., cyber cat bonds and parametric insurance and industry loss warranties (ILW). ILS and ILW are defined as financial instruments whose values are driven by insurance loss events. Those such instruments that are linked to property losses due to natural catastrophes, for example, represent a unique asset class, the return from which is uncorrelated with that of the general financial market.

To achieve the participation of the capital market in index triggered ILS transactions, investors and insurers need to have confidence and full understanding of the composition of unbiased, reliable indices, based on transparent and robust IT security key performance indicators, parameters and metrics. When ART solutions will be made available by reinsurance companies to their cedants, insurers will have access to a significant capacity and will be in a position to propose more insurance policies with higher capacities to their clients.

In addition to the transparency and robustness of the envisaged IT activity-based trigger, the advantage of parametric covers is that they provide significant cost-saving opportunities in terms of loss indemnification. In the case of cyber risk, the identification of a loss, or the sources of it, can be very complex, especially due to the uncertain event duration: when it started, when was the peak, when it stopped. All these elements may be difficult to assess. As a consequence, ART may be a good solution to provide coverage provided a reliable cyber index can be built.

At this point, there have been some attempts to propose pricing models for such products (see Kasper 2019 or Eling and Jung 2018), but the lack of data is always limiting the applicability of the proposed model. As a result, with the data provided by Symantec, this article opens new areas of research for the creation of a reliable index based on IT activity, which could be used in the creation of a new set of ART products.

2. Review of Related Works

As mentioned above, the lack of data is limiting the ability of the insurance industry to propose coverage for cyber risk. In Marotta et al. (2017), an obvious explanation for the lack of available data is given: "Organizations are afraid of releasing too much information about their internal systems to prevent decrease of reputation as well as prevent leakage of knowledge about weaknesses of the system". As a result, in Eling and Schnell (2016), a thorough review of 209 papers conclude to "the immense difficulties to insure cyber risk, especially due to a lack of data and modelling approaches, the risk of change and incalculable accumulation risks". The same conclusion is reached by Florêncio and Herley (2013): as most of the information on cyber risk is gathered through surveys, they conclude that there should be a bias on the analysis of such data.

In this context, some proposals to overcome this issue are proposed including public-private partnerships. This is precisely the aim of the recently formed ASTIN working group (the non-life section of the International Actuarial Association—see www.actuaries.org) on cyber risk: it gathers experts from different fields, such as IT, (re)insurance actuaries, Insurance Linked Securities experts. As a non-profit working group, access to IT data is facilitated and Symantec has offered the database which is used in this article. This is one of the first time where a collaborative non-profit approach to cyber risk is proposed.

Finally, since 2018, the Geneva Association (see Geneva Association 2020) initiated a study to explore the opportunity of building an international cyber claims and cyber incident database, the Cyber Incident Data Exchange and Repository (CIDER). By giving insurers access to a pool of anonymized data, such a tool would aim to assist them in better understanding threat vectors and impacts, and in improving their ability to protect people and businesses from cyber incidents.

Following these different initiatives, it is likely that more and more IT and insurance data will be made available to researchers to explore cyber risk. This article is the first one providing some quantitative analysis on IT data, which could be used for providing parametric insurance cover in the future.

3. Methodology and Dataset Description

As mentioned above, based on a first batch of data provided by Symantec, this paper aims at confirming quantitatively some key features of cyber risk, in particular, the systemic nature of this risk and the possibility to use such IT data to create an index which would reflect the insurance losses. Many articles and books have been written on cyber risk (see Swiss Re 2017; KPMG 2016, or Edwards et al. 2016 to quote only a few of them), but the quantitative analysis of such risk is still limited, due to the limited data available, as the risk is relatively new. This article will therefore focus on the analysis of this first batch of data to derive some quantitative elements that could be used for pricing (re)insurance products covering cyber risk.

The data provided by Symantec includes the following elements on a daily basis between 1st July 2016 and 31st December 2018:

- The number of viruses blocked on each computer using Symantec software protection; virus protection software is designed to prevent viruses, worms and Trojan horses from getting onto a computer, as well as to remove any malicious software code that has already infected a computer;
- The sample covers all the countries where at least one computer is protected by Norton, the Symantec protection software;
- The number of intrusions blocked on each computer using Symantec software protection. The intrusion prevention system (IPS) is the Symantec Endpoint Protection client's second layer of defense after the firewall. The intrusion prevention system is a network-based system. If a known attack is detected, one or more intrusion prevention technologies can automatically block it. For example, this technology prevents malicious files from getting to a hard drive in the first place. Unlike an antivirus, which looks for known malicious files, IPS scans the network traffic stream in order to find threats using known exploits and attack vectors. IPS does not detect specific files, but rather specific methods that can be used to get malicious files onto a network. This allows IPS to protect against both known and unknown threats (see Holm 2014), even before antivirus signatures can be created for them;
- The daily numbers are split by country;
- For the antivirus extract (i.e., the number of viruses being blocked by Norton over all the computers worldwide), there are 127,231 records over the period 1st July 2016 and 31st December 2018. One record consists of the number of viruses being blocked for one day in one country;
- For the IPS extract, there are 128,054 records. One record consists of the number of IPS being blocked for one day in one country.

The Symantec data needs to be correlated to real loss data in order to assess the possibility to define a cyber loss index on the basis of such data. As real loss data is not available yet, the Symantec data is checked against the information included in the report of the Center for Strategic & International Studies (2006) and against the information available on the website PrivacyRights.org.

The following sections will focus on the information which we can be gathered from the analysis of the Symantec data:

- The first section will concentrate on Virus activities being blocked;
- The second section will concentrate on Intrusions being blocked;
- The third section will try to link the above two cyber activities with known events which happened on the Internet.

In the first two sections, the challenge is to visualize within Symantec data any surge in cyber activity which could be used as a trigger for a parametric insurance cover (see Figure 1 below). If such events are present in the Symantec data, there is a good chance that they are related to insurance losses. However, this next step consisting of correlating cyber activity extreme events with insurance loss is not part of this article.

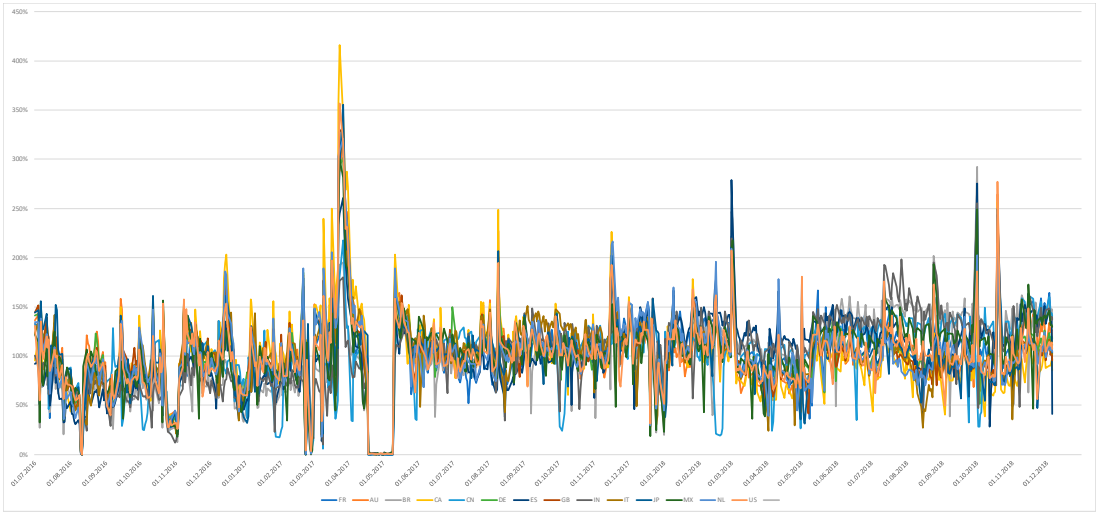

Figure 1. Number of virus being blocked on computers protected by Symantec protection software.

It must be noted that one strong assumption around this work lies with the idea that the virus/intrusions which affect computers not being protected by Symantec are proportional to virus/intrusions being blocked by Symantec. Even though such assumption is reasonable, there is no way in which such an assumption can be tested at this point. The need for this assumption would disappear in a business model where an insurer teams up with Symantec, so that all the insureds are actually Symantec-protected, and those not protected by Symantec are not insured. In such a case, the basis risk related to the proposed parametric trigger calculated on the basis of Symantec data would be significantly reduced.

4. Symantec Antivirus activities

The following is a visualization of the Symantec antivirus activity. For each day, the number of viruses being blocked by Symantec protection software on each end-used computer is shown per country.

The graph above has the following limitations:

- Data from weekends is filtered;
- Only the countries where number of blocks were more than 100,000 are shown.

Below is the result of a simple study of stationarity performed on the above time-series by country. The Augmented Dickey-Fuller (ADF) test with drift and trend is applied to each time-series: The above results (Table 1) confirm that the time-series are stationary.

Table 1. Results of the Augmented Dickey-Fuller test.

Lag	France		Australia		Brazil		Canada		China		Germany		Spain	
	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value
0	−10.23	0.01	−9.71	0.01	−12.73	0.01	−10.31	0.01	−10.55	0.01	−9.98	0.01	−10.28	0.01
1	−8.99	0.01	−8.71	0.01	−9.55	0.01	−8.65	0.01	−9.51	0.01	−8.93	0.01	−8.46	0.01
2	−7.39	0.01	−7.17	0.01	−7.28	0.01	−7.11	0.01	−8.31	0.01	−7.06	0.01	−6.82	0.01
3	−6.87	0.01	−6.94	0.01	−6.91	0.01	−6.44	0.01	−7.7	0.01	−6.66	0.01	−6.44	0.01
4	−6.24	0.01	−6.39	0.01	−5.94	0.01	−5.4	0.01	−7.29	0.01	−6.24	0.01	−6.16	0.01
5	−6.52	0.01	−6.48	0.01	−6.01	0.01	−6.34	0.01	−7.51	0.01	−6.36	0.01	−6.01	0.01
6	−6.36	0.01	−6.12	0.01	−5.72	0.01	−5.86	0.01	−7.29	0.01	−6.26	0.01	−5.78	0.01
Lag	**UK**		**India**		**Italy**		**Japan**		**Mexico**		**Holland**		**USA**	
	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value
0	−9.36	0.01	−12.37	0.01	−9.26	0.01	−10.35	0.01	−11.87	0.01	−9.75	0.01	−10.48	0.01
1	−8.19	0.01	−9.1	0.01	−8.08	0.01	−9.37	0.01	−9.69	0.01	−8.81	0.01	−9.15	0.01
2	−6.66	0.01	−7.63	0.01	−6.54	0.01	−8.05	0.01	−7.7	0.01	−6.91	0.01	−7.48	0.01
3	−6.59	0.01	−6.6	0.01	−6.13	0.01	−7.55	0.01	−6.59	0.01	−6.53	0.01	−6.99	0.01
4	−6.14	0.01	−5.19	0.01	−5.77	0.01	−7.35	0.01	−5.64	0.01	−6.22	0.01	−6.28	0.01
5	−6.47	0.01	−5.78	0.01	−6.1	0.01	−6.73	0.01	−6.4	0.01	−6.24	0.01	−6.85	0.01
6	−6.22	0.01	−6.06	0.01	−5.93	0.01	−6.58	0.01	−6.84	0.01	−6.2	0.01	−6.58	0.01

As expected, there are peak activities, in particular on the following days:

- 14–16 Dec. 2016;
- 24 Mar. 2017 (very high activity);
- 11 May 2017;
- 9 Aug. 2017;
- 16 Nov. 2017;
- 1 Mar. 2018;
- 1 Oct. 2018;
- 19 Oct. 2018.

From Figure 1 above, there seems to be a possibility to define an index if the above high activities would be related to insurance losses. In order to confirm this possibility to define an index, in further sections, we will try to provide some explanations for these high activities.

At this point, we can estimate correlations between the antivirus activities of each country. The correlation matrix is shown below (see Table 2):

Table 2. Correlation matrix between the antivirus activities in different countries.

	FR	AU	BR	CA	CN	DE	ES	GB	IN	IT	JP	MX	NL	US
FR	100%	80%	75%	72%	61%	88%	89%	84%	71%	86%	82%	83%	88%	83%
AU	80%	100%	56%	89%	52%	93%	68%	94%	49%	86%	90%	76%	91%	90%
BR	75%	56%	100%	45%	65%	64%	82%	57%	84%	58%	58%	80%	59%	66%
CA	72%	89%	45%	100%	37%	89%	60%	93%	38%	83%	86%	70%	90%	92%
CN	61%	52%	65%	37%	100%	55%	59%	48%	72%	49%	55%	67%	45%	54%
DE	88%	93%	64%	89%	55%	100%	77%	95%	58%	90%	90%	82%	95%	91%
ES	89%	68%	82%	60%	59%	77%	100%	73%	76%	77%	71%	80%	76%	74%
GB	84%	94%	57%	93%	48%	95%	73%	100%	49%	92%	89%	77%	96%	93%
IN	71%	49%	84%	38%	72%	58%	76%	49%	100%	53%	55%	77%	52%	60%
IT	86%	86%	58%	83%	49%	90%	77%	92%	53%	100%	85%	74%	90%	85%
JP	82%	90%	58%	86%	55%	90%	71%	89%	55%	85%	100%	75%	89%	88%
MX	83%	76%	80%	70%	67%	82%	80%	77%	77%	74%	75%	100%	76%	83%
NL	88%	91%	59%	90%	45%	95%	76%	96%	52%	90%	89%	76%	100%	90%
US	83%	90%	66%	92%	54%	91%	74%	93%	60%	85%	88%	83%	90%	100%

FR: France, AU: Australia, BR: Brazil, CA: Canada, CN: China, DE: Germany, ES: Spain, GB: UK, IN: India, IT: Italy, JP: Japan, MX: Mexico, NL: Holland, US: USA.

Countries with high correlation between each other (>80%) include France, Australia, Canada, Germany, UK, Italy, Holland, Japan and the US.

For Brazil and Mexico, it seems that local hacker groups focus on their country which results in smaller correlation to other countries.

Finally, it must be noted that the high correlations of the time-series of these countries are influenced by the fact that Symantec has high market shares in these countries.

In addition to the linear correlation analysis, the matrix of Kendall Tau (Table 3) is also shown below:

Table 3. Kendall Tau matrix between the antivirus activities in different countries.

	FR	AU	BR	CA	CN	DE	ES	GB	IN	IT	JP	MX	NL	US
FR	100%	48%	51%	38%	39%	60%	66%	51%	47%	56%	51%	54%	61%	52%
AU	48%	100%	29%	60%	25%	69%	36%	70%	24%	60%	64%	43%	63%	64%
BR	51%	29%	100%	19%	49%	36%	60%	26%	65%	29%	30%	56%	32%	40%
CA	38%	60%	19%	100%	11%	61%	27%	69%	16%	58%	59%	36%	62%	68%
CN	39%	25%	49%	11%	100%	28%	37%	19%	55%	22%	27%	50%	20%	27%
DE	60%	69%	36%	61%	28%	99%	45%	72%	33%	66%	66%	49%	73%	66%
ES	66%	36%	60%	27%	37%	45%	100%	38%	52%	46%	38%	50%	46%	42%
GB	51%	70%	26%	69%	19%	72%	38%	100%	22%	71%	64%	40%	74%	67%
IN	47%	24%	65%	16%	55%	33%	52%	22%	100%	26%	30%	57%	28%	36%
IT	56%	60%	29%	58%	22%	66%	46%	71%	26%	100%	61%	40%	67%	58%
JP	51%	64%	30%	59%	27%	66%	38%	64%	30%	61%	100%	42%	62%	61%
MX	54%	43%	56%	36%	50%	49%	50%	40%	57%	40%	42%	100%	43%	52%
NL	61%	63%	32%	62%	20%	73%	46%	74%	28%	67%	62%	43%	100%	65%
US	52%	64%	40%	68%	27%	66%	42%	67%	36%	58%	61%	52%	65%	100%

Based on the above, if we define systemic risk as the risk of the occurrence of an event that threatens the well-functioning of the system of interest (financial, payments, banking, etc.), sometimes to the point of making its operation impossible (see Billio et al. 2012), it shows clearly that cyber risk has a systemic nature: many countries are hit by Virus attacks at the same time. Finally, we must bear in mind that this conclusion is limited to the available dataset.

5. Symantec Intrusion Activities

The following is a visualization of the Symantec intrusion activity. For each day, the number of intrusion attempts being blocked by Symantec protection software on each end-used computer is shown per country.

As for the Antivirus activities, the graph above (Figure 2) has the following limitations:

- Data from weekends is filtered;
- Only the countries where the number of blocks were more than 100,000 are shown.

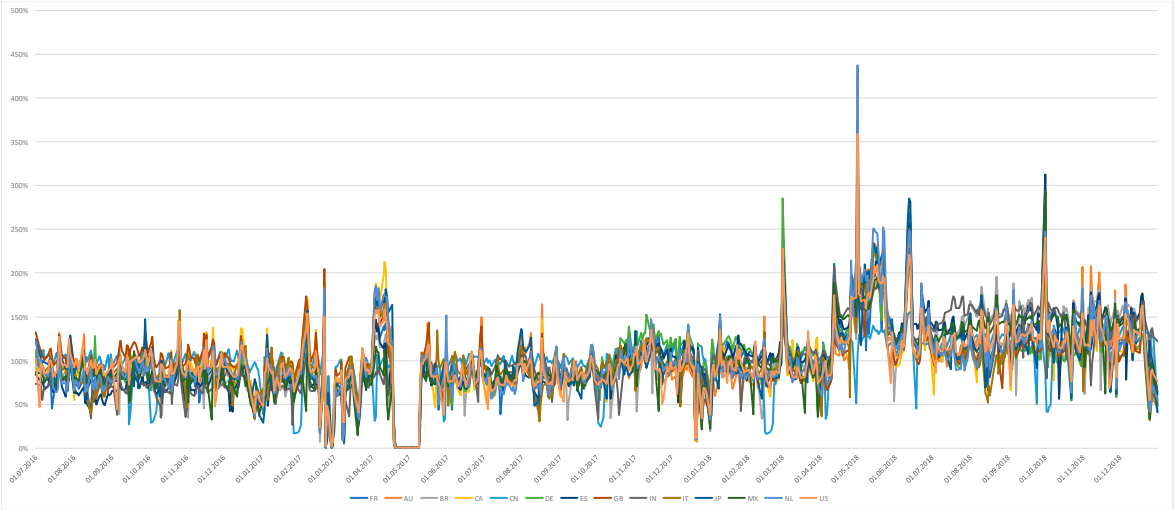

Figure 2. Number of intrusions being blocked on computers protected by Symantec protection software.

As with the antivirus, a simple study of stationarity performed on the above time-series by country was performed. The Augmented Dickey-Fuller (ADF) test with drift and trend is applied to each time-series:

As with the antivirus activities, the above results (Table 4) confirm that the time-series are stationary.

Also, as expected, there are peak activities, in particular, on the following days:

- 21 Feb 2017;
- 6 and 12 Apr 2017;
- 2 Mar 2018;
- 1 May 2018 (highest activity recorded);
- 12 Jun 2018;
- 1 Oct 2018.

Table 4. Results of the Augmented Dickey-Fuller test.

Lag	France		Australia		Brazil		Canada		China		Germany		Spain	
	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value
0	−9.71	0.01	−9.62	0.01	−10.56	0.01	−10.37	0.01	−9.66	0.01	−9.36	0.01	−10.15	0.01
1	−8.44	0.01	−8.22	0.01	−8.56	0.01	−8.5	0.01	−8.77	0.01	−8.05	0.01	−8.31	0.01
2	−7.06	0.01	−6.45	0.01	−6.68	0.01	−7.08	0.01	−7.34	0.01	−6.21	0.01	−6.81	0.01
3	−7.06	0.01	−5.92	0.01	−6.19	0.01	−6.92	0.01	−7.52	0.01	−6.02	0.01	−6.48	0.01
4	−6.23	0.01	−5.18	0.01	−5.32	0.01	−6.15	0.01	−7.07	0.01	−5.43	0.01	−5.63	0.01
5	−5.69	0.01	−5.39	0.01	−5.14	0.01	−6.1	0.01	−6.55	0.01	−5.17	0.01	−5.31	0.01
6	−5.25	0.01	−5.25	0.01	−4.96	0.01	−5.75	0.01	−5.92	0.01	−4.79	0.01	−4.96	0.01
Lag	**UK**		**India**		**Italy**		**Japan**		**Mexico**		**Holland**		**USA**	
	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value	ADF	*p*.Value
0	−9.46	0.01	−10.84	0.01	−9.68	0.01	−10.14	0.01	−10.05	0.01	−9.98	0.01	−10.27	0.01
1	−8.07	0.01	−7.86	0.01	−8.16	0.01	−8.57	0.01	−7.83	0.01	−8.5	0.01	−8.75	0.01
2	−6.75	0.01	−6.78	0.01	−6.76	0.01	−6.96	0.01	−6.16	0.01	−6.8	0.01	−7.1	0.01
3	−6.84	0.01	−6.13	0.01	−6.84	0.01	−6.55	0.01	−5.53	0.01	−6.59	0.01	−6.73	0.01
4	−5.92	0.01	−5.38	0.01	−5.89	0.01	−6.03	0.01	−4.64	0.01	−5.91	0.01	−6.11	0.01
5	−5.79	0.01	−4.98	0.01	−5.84	0.01	−5.65	0.01	−4.56	0.01	−5.66	0.01	−5.92	0.01
6	−5.41	0.01	−4.45	0.01	−5.24	0.01	−5.18	0.01	−4.08	0.01	−5.39	0.01	−5.59	0.01

From Figure 2 above, there seems to be a possibility to define an index if the above high activities would be related to insurance losses (see Figure 3 for a comparison of blocked intrusions vs number of records stolen). Interestingly, such an intrusions-based index would trigger the insurance cover at different times as compared to an antivirus-based index. Therefore, it would be interesting to see if the index based on antivirus or on intrusions would be performing better in terms of insurance loss prediction. Alternatively, a blend of the two indexes could be the best performer.

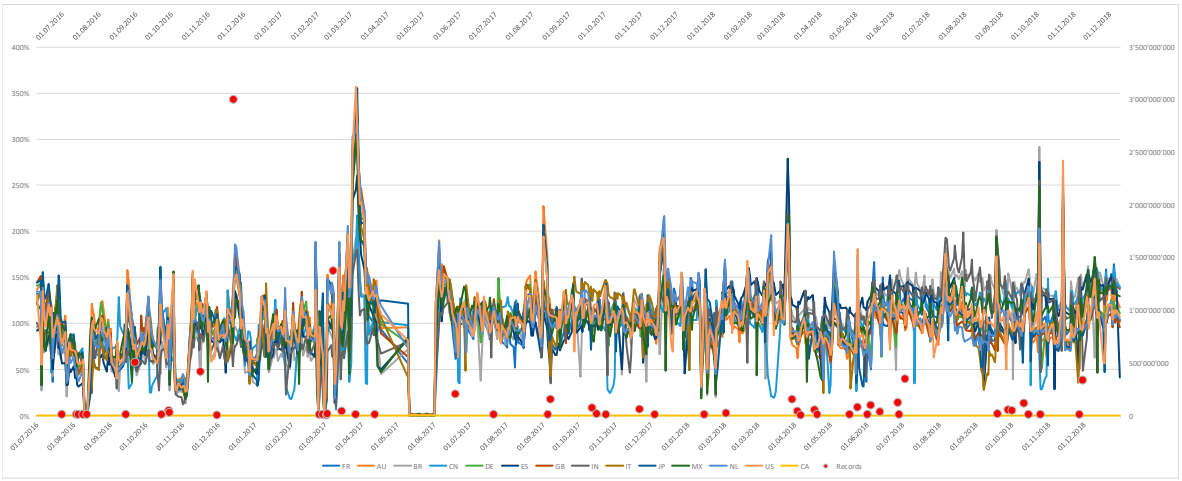

Figure 3. Number of blocked intrusions vs. the number of records stolen.

As for antivirus activities, we can estimate correlations between the activities of each country. The correlation matrix (Table 5) is shown below:

Table 5. Correlation matrix between the intrusion activities in different countries.

	FR	AU	BR	CA	CN	DE	ES	GB	IN	IT	JP	MX	NL	US
FR	100%	84%	84%	85%	60%	93%	93%	84%	82%	93%	87%	87%	92%	88%
AU	84%	100%	80%	94%	59%	85%	81%	92%	72%	85%	91%	82%	93%	92%
BR	84%	80%	100%	77%	66%	77%	89%	76%	87%	79%	81%	93%	81%	85%
CA	85%	94%	77%	100%	55%	88%	79%	95%	68%	87%	91%	79%	94%	95%
CN	60%	59%	66%	55%	100%	60%	62%	57%	68%	58%	61%	67%	56%	62%
DE	93%	85%	77%	88%	60%	100%	85%	87%	72%	88%	87%	79%	92%	88%
ES	93%	81%	89%	79%	62%	85%	100%	77%	89%	89%	83%	91%	86%	85%
GB	84%	92%	76%	95%	57%	87%	77%	100%	65%	88%	89%	77%	92%	93%
IN	82%	72%	87%	68%	68%	72%	89%	65%	100%	74%	76%	90%	74%	78%
IT	93%	85%	79%	87%	58%	88%	89%	88%	74%	100%	87%	81%	91%	87%
JP	87%	91%	81%	91%	61%	87%	83%	89%	76%	87%	100%	81%	92%	92%
MX	87%	82%	93%	79%	67%	79%	91%	77%	90%	81%	81%	100%	83%	87%
NL	92%	93%	81%	94%	56%	92%	86%	92%	74%	91%	92%	83%	100%	93%
US	88%	92%	85%	95%	62%	88%	85%	93%	78%	87%	92%	87%	93%	100%

All countries except China and India show high correlation between each other (>80%).

6. Cyber Activity vs. Known Data Breaches

In order to explain peak activities, we now turn to the use of the data shown on privacyrights.org, where the main data breach events are provided with the number of records being stolen. There are a few limitations to this data:

- The date of the event shown on the database is the date when the event was disclosed publicly. However, it can relate to a time span which was much earlier and could have spread over many months;
- The events are only related to data breaches and do not include other types of cyber events.

Despite the above limitations, we will try to understand the peak activities described in the above two sections.

Privacy Rights Clearinghouse (hereinafter "PRC" from the website PrivacyRights.org) was founded in 1992 as a program of the University of San Diego School of Law's Center for Public Interest Law. In 1996, the first iteration of their website was launched, and it started tracking data breaches. The PRC data does not only contain data theft, but also the non-theft category unintended disclosure (DISC), and some partial theft, partial non-theft categories, such as physical loss (PHYS), portable (PORT) and stationary (STAT) losses, and unknown (UNKN). For this section, all categories of data breaches are aggregated together. It has to be noted that the PRC data are collected only for USA events. However, due to the high correlations demonstrated in the previous section, it is possible that the PRC data reflects also the worldwide cybercrime activity.

In this section, we are also going to use data from the Centre for Strategic and International Studies (CSIS). CSIS is a nonprofit USA policy research organization dedicated to advancing practical ideas to address the world's greatest challenges. CSIS records significant cyber incidents since 2006 and focuses on cyber-attacks on government agencies, defense and high-tech companies, or economic crimes with losses of more than a million dollars.

We will first plot the main data breaches (number of records lost as shown in the website PrivacyRights.org) against the antivirus activity of Symantec, to see if some correlations between these events can be drawn:

The red dots show the number of data stolen at each data breach events (PRC data).

The estimation of the correlation between data breach events and Symantec antivirus activity is small. It must be noted that a simple correlation assessment is unlikely to work as there are a number of elements to take into account:

- There is likely a delay (positive or negative) between the data breach events (positive or negative) and antivirus activity. In addition, the date of the event recorded on PrivacyRights.org is the date of the event being disclosed; it is not the date of the occurrence of the event. Such occurrence date is anyway difficult to assess as it may not be known at all (e.g., the Yahoo data breach occurred over many years/months, and the start date is not known precisely);
- The antivirus activity of Symantec may simply not be reflective of data breaches and the intrusion activity may be a better indicator;
- The data breaches recorded on PrivacyRights.org relate to large corporates, while both intrusion and antivirus activities of Symantec relate to end-user computers and not big servers of large corporates.

Taking all these elements into account, there seems to be little hope to find a dependency model that would take into account at least the delay between the Symantec activity and the data breach.

When we try to understand the high data breach activity with the reports from the CSIS, we find that the events below may be an explanation for the peaks (see Figure 4 below):

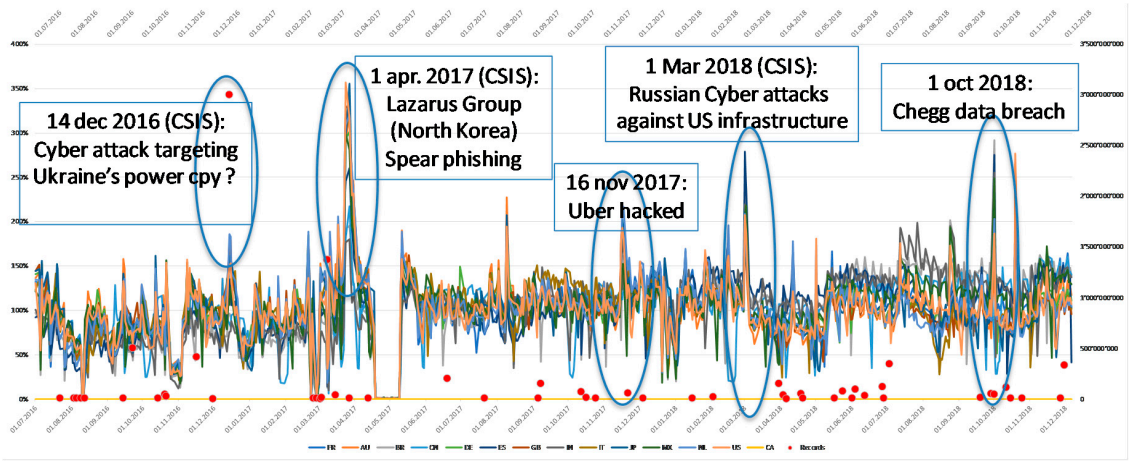

Figure 4. Number of blocked viruses vs. the number of data stolen vs. Centre for Strategic and International Studies (CSIS) known events.

However, such analysis has a lot of limitations as it is just one source of information. Still, it shows that there are elements of "Cyber war" that may influence the Symantec activity. When a legal definition for "Cyber war" will be available, such risk elements should definitely be excluded from the parametric insurance cover proposed by the (re)insurance industry.

In the same way as above, we plot the number of data stolen from known data breach events vs. the Intrusion activity of Symantec, and try to link it with known events given by the CSIS (see Figure 5):

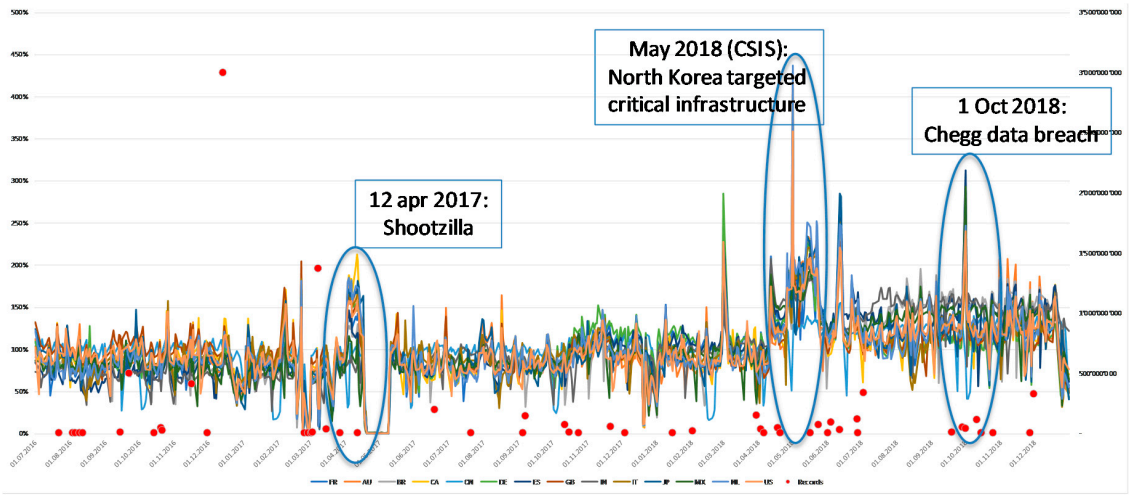

Figure 5. Number of blocked intrusions vs. number of data stolen vs. CSIS known events.

Overall, there is a negative correlation between the number of data stolen from known data breaches and the intrusion activity of Symantec. It could be a sign that the Symantec intrusion software is efficient as it blocks effectively the data breach crimes.

However, as mentioned above, there is still a lot of uncertainty on any conclusion that could be drawn from such analysis.

7. Conclusions

From an academic point of view, bearing in mind the lack of available data at this point, this paper provides some first quantitative elements:

- The correlation matrix between IT activity in different countries confirms the systemic nature of this risk;
- The Symantec data shows that there are peak activities that could be the foundation for a cyber loss index.

Even though these elements are encouraging, the following studies should be performed to explain the characteristics of cyber risk:

- Obtain a good understanding of the IT event succession when cyber activity is seen. It relates, in particular, to the definition of the cyber event (e.g., when it started, when was at the peak, when it stopped) but also to the way in which the IT companies update their antivirus/intrusion software and how long such update would take;
- Get insurance loss data or at least loss data (e.g., from a police claims database) and try to model a dependence between IT activity and such loss data.

On the basis of these findings, (re)insurance companies could consider parametric insurance to provide the needed coverage to the industry. For such coverage, a cyber loss index based on IT activity could be considered when partnering with Symantec. As an example, the coverage would be limited to the computers which are protected by Symantec as the proposed cyber loss index would reflect such activity and reduce the basis risk related to such parametric cover.

Finally, the development of non-profit initiatives (e.g., ASTIN, Geneva Association) should make more and more data available for the study of cyber risk. When such data will have been analyzed by academics, this article shows that, like for any other types of systemic risk (e.g., credit & surety), (re)insurance companies should be in a position to provide an adequate price for a given parametric cover.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

Abadie, Aurélie. 2019. *Cyber, la Reassurance Veut Lever Les Freins*. Antony: L'argus de L'assurance.

Billio, Monica, Mila Getmansky, Andrew W. Loc, and Loriana Pelizzon. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104: 535–59. [CrossRef]

Center for Strategic & International Studies. 2006. Significant Cyber Incidents since 2006. Available online: www.csis.org (accessed on 5 January 2020).

Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. 2016. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity* 2: 3–14. [CrossRef]

Eling, Martin, and Kwangmin Jung. 2018. Copula approaches for modeling cross-sectional dependence of data breach losses. *Insurance: Mathematics and Economics* 82: 167–80. [CrossRef]

Eling, Martin, and Werner Schnell. 2016. What do we know about cyber risk and cyber risk insurance? *The Journal of Risk Finance* 17: 474–91. [CrossRef]

Falco, Gregory, Martin Eling, Danielle Jablanski, Virginia Miller, Lawrence A. Gordon, Shaun Shuxun Wang, Joan Schmit, Russell Thomas, Mauro Elvedi, Thomas Maillart, and et al. 2019. *A Research Agenda for Cyber Risk and Cyber Insurance*. Stanford: Stanford University.

Federal Bureau of Investigation. 2018. *Internet Crime Report*. Washington, DC: Federal Bureau of Investigation.

Florêncio, Dinei, and Cormac Herley. 2013. Sex, Lies and Cybercrime Surveys. In *Economics of Information Security and Privacy III*. New York: Springer, pp. 35–53.

Geneva Association. 2020. *Exploring the opportunity for a Cyber Incident Data Exchange and Repository (CIDER)*. Geneva: Geneva Association.

Guy Carpenter, Cybercube. 2019. *Looking beyond the Clouds*. New York: Guy Carpenter&Company LLC.

Holm, H. 2014. Signature Based Intrusion Detection for Zero-Day Attacks: (Not) A Closed Chapter? Paper presented at the 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, January 6–9; pp. 4895–4904.

Kasper, Daniel. 2019. *Analyzing the Feasibility of Cyber Bonds by Stochastically Solving a Copula-based Model with Differential Evolution*. Köln: University of Cologne.

KPMG. 2016. *Small Business Reputation & the Cyber Risk*. Amstelveen: KPMG.

Marotta, Angelica, Fabio Martinelli, Stefano Nanni, Albina Orlando, and Artsiom Yautsiukhin. 2017. Cyber-insurance survey. *Computer Science Review* 24: 35–61. [CrossRef]

Swiss Insurance Association (Switzerland). 2017. *Economic Impact of Cyber Accumulation Scenarios*. Zürich: Cyber Working Group.

Swiss Re. 2017. *Cyber: Getting to Grips with a Complex Risk*. Sigma Report 2017. Zürich: Swiss Re.