Cyber risk modeling: A discrete multivariate count process approach

Yang Lu[*], Jinggong Zhang[†], Wenjun Zhu[‡]

Abstract

In the past decade, cyber risk has raised much interest in the economy, and cyber risk has evolved from a type of pure operational risk to both operational and liability risk. However, the modeling of cyber risk is still in its infancy. Compared with other financial risks, cyber risk has some unique features. In particular, discrete variables regularly arise both in the frequency component (e.g. number of events per unit time), and the severity component (e.g. the number of data breaches for each cyber event). In addition, the modeling of these count variables are further complicated by nonstandard properties such as zero inflation, serial and cross-sectional correlations, as well as heavy tails. Previous cyber risk models have largely focused on continuous models that are incompatible with many of these characteristics. This paper introduces a new count-based frequency-severity framework to model cyber risk, with a dynamic multivariate negative binomial autoregressive process for the frequency component, and the generalized Poisson inverse-Gaussian distribution for the severity component. We unify these new modeling tools by proposing a tractable Generalized Method of Moments for their estimation and apply them to the Privacy Rights Clearinghouse (PRC) dataset.

Keywords: Operational risk, Cyber risk, Systemic and pairwise contagion, Self-excitation, Negative binomial autoregressive process, Generalized Poisson inverse-Gaussian distribution, Discrete stable distribution.

JEL: C22, G17

[*]Department of Mathematics and Statistics, Concordia University, Montreal, QC, Canada. E-mail: yang.lu@concordia.ca.

[†]Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore 639798. E-mail: jgzhang@ntu.edu.sg.

[‡]Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore 639798. E-mail: wjzhu@ntu.edu.sg.

1 Introduction

With the increasingly pervasive dependence on new disruptive technologies, such as mobiles and Internet of Thing (IoT) devices, more corporations are exposed to cyber risk, from unintended data disclosure and theft to data breach and ransomware.[1] Recent global pandemic has also boosted remote workforces, leading to new and more frequent cyber risk exposures. As a result, over the past decade, cyber risk has become a high-profile risk that threatens business stability as well as corporation's reputation. Different stakeholders including commercial corporations and governments are increasing their investments in safeguarding their information systems. The worldwide end user spending in information security & risk management is forecast to be $150.4 billion in 2021, according to Gartner, Inc. (Gartner, 2021). It is also forecast that the global information security market is to reach $170.4 billion in 2022 (Kim et al., 2018). Furthermore, around the globe, $9-12 trillion of economic value could depend on the strength of the cyber security environment (Bailey et al., 2014). Cyber risk is also costly. As of 2020, the average cost of a data breach is estimated to be $3.86 million (IBM, 2020).

The insurance market can make an important contribution to cyber risk management with their expertise in risk quantification and exposure measurement, encouraging risk prevention and risk mitigation, facilitating crisis management and risk reduction, etc. Indeed, in recent years, we observe extremely fast growth of the cyber insurance market with a compound annual growth rate of 25% from 2012 to 2015 (OECD, 2017). The momentum continues and it is estimated by Fitch that the direct written premium for cyber insurance increased by over 22% in 2020 (FitchRatings, 2021). This unprecedented opportunity, however, also spells numerous challenges for insurers. Indeed, cyber risk has evolved from a type of pure operational risk to both operational and liability risk, creating potential correlated shocks on different parts of the balance sheet of an insurer.[2] In addition, while the cyber insurance market is growing rapidly, it is generally perceived as being still in its infancy and the coverage remains low (Swiss Re, 2017). The market size of cyber insurance market is relatively small compared with the

[1] Currently, there is no consensual definition of cyber risk in the insurance sector. But a few associations of insurance companies have put forward some definitions. For example, the Geneva Association (2016) defined cyber risk as "any risk emerging from the use of information and communication technology that compromises the confidentiality, availability, or integrity of data or services." Similarly, Cebula and Young (2010) define cyber risk as "operational risks to information and technology assets that have consequences affecting the confidentiality, availability or integrity of information or information systems".

[2] Traditionally, operational risk is assumed to be either uncorrelated, or negative correlated with other risks (Gatzert and Kolb, 2014).

property and liability insurance market in general. In the U.S., for example, the direct written premia of cyber insurance in 2019 is $2.25 billion (AM Best, 2019), compared to the $1,495 billion non-life insurance market (OECD, 2019). In particular, an important challenge that impedes further development of cyber insurance market is the lack of appropriate quantitative models for cyber risk (Maillart and Sornette, 2010; Biener et al., 2015; Falco et al., 2019). Improved cyber risk modeling could reduce uncertainty, encourage underwriting, and hence increase the willingness of insurers to extend the cyber insurance coverage.

Besides the scarcity of data, the modeling challenge can be explained by the numerous unique characteristics that distinguish cyber risk from traditional insurance and financial risks. In particular, discretely valued variables regularly arise both for the frequency components (e.g. number of events per unit time), and the severity components (e.g. the number of data breaches for each cyber event). In addition, the modeling of these count variables are further complicated by nonstandard properties such as zero inflation, serial and cross-sectional correlations, as well as heavy tails. On the other hand, most existing cyber risk models are borrowed from traditional insurance risks,[3] and cannot well accommodate for many of these features.

This paper aims to introduce a new count-based multivariate frequency-severity framework that is tailored to the aforementioned stylized facts of cyber risk data. The model includes a (multivariate, dynamic) frequency component and a (cross-sectional) severity component, and both of them are count-valued to ensure compatibility with observed data. Our frequency component is the multivariate negative binomial autoregressive (NBAR) process, which is an extension of the negative binomial distribution to the dynamic *and* multivariate context. In addition, the proposed frequency model allows for two separate mechanisms of contagion, including a ***pairwise*** contagion channel and a ***systemic*** contagion channel. Our severity model is based on the Generalized Poisson inverse-Gaussian (GPIG) distribution, which is known for its capability of capturing heavy tails. Both models are quite flexible, but also highly tractable thanks to their closed form probability generating function (or equivalently Laplace transform).

[3] A recent literature introduces network-based (Fahrenwaldt et al., 2018; Jevtić and Lanchier, 2020) or epidemiology-based (Xu and Hua, 2019; Hillairet and Lopez, 2021) frequency models. Even though these models have appealing theoretical properties, they are not suitable for currently available cyber risk databases. As Hillairet and Lopez (2021) put it, "Using the model to replicate the attack is a difficult task, since few available data support the calibration of the parameters." As a consequence, the illustration of these models have been largely confined to simulation studies.

This allows us to propose a unified, Generalized Method of Moment (GMM) method to estimate these models. We apply our methodology to the Privacy Rights Clearinghouse (PRC) database on data breach events. Our empirical results confirm the statistical significance of both contagion channels in the cyber risk dataset, as well as the unbounded expectation (i.e., extremely heavy tail property) of the severity distribution.

Our paper makes several contributions to the literature. From the frequency side, previous cyber risk research is largely focused on the (static) distribution of number of events per unit of time, e.g., hour, day, week, month, etc. (Edwards et al., 2016; Wheatley et al., 2016; Eling and Loperfido, 2017; Peng et al., 2017; Xu et al., 2017; Peng et al., 2018; Eling and Jung, 2018; Eling and Wirfs, 2019; Bessy-Roland et al., 2021; Wheatley et al., 2021). For example, using the Privacy Rights Clearinghouse (PRC) dataset, Edwards et al. (2016) show that negative binomial distribution fits well to the daily number of data breaches. This result is confirmed by Eling and Loperfido (2017), who conduct multidimensional scaling and goodness-of-fit tests, and find that negative binomial distribution provides better fits than Poisson. However, most existing models either do not account for the cross-sectional dependence of the different frequency variables, or omit the serial correlation across different time periods, and ignore the dynamic nature of the data. Only a few attempts have been made to capture either the serial correlation, through time series models (such as ARMA/GARCH models), or the cross-sectional dependence, via copulas. Moreover, they do not account for the fact that the frequency is discrete (e.g., Xu et al., 2017; Peng et al., 2018). We contribute to the literature by introducing the multivariate NBAR process. This model is the first to account for both cross-sectional and serial dependence between discrete loss frequencies of different types of cyber events. Moreover, our proposed frequency model has insightful economic meaning. Indeed, because different types of cyber risk events feature complicated dependence pattern, we also pioneer the introduction of two separate mechanisms of contagion, with one latent multivariate factor accounting for the ***pairwise*** contagion channel, and another capturing the ***systemic*** contagion channel. These two transmission channels are disentangled by exploring the extent to which events of different types occur *simultaneously*.

This paper also contributes to the cyber risk severity modeling. It is widely established that the severity distribution of cyber risk is heavy tailed (Maillart and Sornette, 2010; Edwards et al., 2016; Wheatley et al., 2016; Eling and Loperfido, 2017; Peng et al., 2017; Eling and Jung,

2018; Farkas et al., 2021; Jung, 2021; Wheatley et al., 2021). For example, using a combined dataset from the Open Security Foundation and PRC, Maillart and Sornette (2010) identify an extremely heavy-tailed power-law distribution for loss severity. Recently, extreme value theory have gained popularity in this literature (Eling and Wirfs, 2019; Sun et al., 2020; Jung, 2021). However, for data breach distribution modeling, existing models do not take into account the two important stylized facts of data: *i*) the severity is often integer-valued (such as number of records of data breaches); *ii*) loss severity has a large probability of taking a value of zero. Hence we introduce the GPIG distribution to the insurance literature, that can simultaneously account for the above mentioned two stylized facts, and account for heavy tail at the same time.

The rest of the paper is organized as follows. Section 2 describes the database and motivates the need of a new modeling approach. Section 3 introduces the new frequency-severity framework. Section 4 deals with the estimation methodology of the model, which is implemented in Section 5 with the PRC dataset. Section 6 concludes. Proofs, technical details and additional estimation results are gathered in the appendices.

2 Data description and motivation

In this section, we describe some important characteristics of the data used in this paper. These features will motivate the model(s) that we propose in Section 3. We explore data breaches derived from the Privacy Rights Clearinghouse (PRC).[4] The PRC lists data breach events involving US citizens, and is the largest public database on breached data information (Edwards et al., 2016). For each event, it provides the organization, the type of data breach, the breach severity (i.e., total number of records breaches), and the date of the event. The PRC dataset is reliable in the sense that each breach event is confirmed by at least one major media source. It is continuously updated by the PRC, and has been extensively used both in academia (see, e.g., Eling and Loperfido, 2017; Eling and Jung, 2018; Eling and Wirfs, 2019; Sun et al., 2020; Bessy-Roland et al., 2021; Farkas et al., 2021; Wheatley et al., 2021) and in industry practice (see Edwards et al., 2016, for a review). Moreover, due to the lack of sufficient proprietary cyber risk data, the public databases serve as important supplementary

[4]The Privacy Rights Clearinghouse database is made publicly available at `https://privacyrights.org/data-breaches`. This article is based on a download on 2/27/2021, covering events between 2005 and 2019.

source of information for insurance companies (Farkas et al., 2021).

Previous researches show that different types of data breaches need to be distinguished, in order to account for risk heterogeneity (Böhme and Kataria, 2006; Eling and Loperfido, 2017; Eling and Jung, 2018), as well as the inter-dependence between different types of events (Eling and Jung, 2018; Xu et al., 2017). In this paper, we follow the definition of Privacy Rights Clearinghouse (PRC) and Eling and Jung (2018) and segment the events into four types of data breach.[5] Panel A of Table 1 lists the definitions of the four types of data breaches. Then we discretize the date of the event into weeks, and obtain four nonnegative frequency counts $Y_{1t}, .., Y_{4t}$, corresponding to the number of events of type 1,...,4 within each week t, respectively, for week $t = 1, ..., T = 772$.[6] In other words, for type j, $j = 1, ..4$, we observe a total of $N_j = \sum_{t=1}^{T} Y_{jt}$ events. The severity of these events will be analyzed separately for each type. That is, for loss severity, we work with the granular, event-level data as in Farkas et al. (2021), instead of aggregated weekly total loss to avoid the double-counting issue.

2.1 Data on frequency

Panel B of Table 1 provides summary statistics of the time series of loss frequency counts. We observe several characteristics of the loss frequency data. First, the frequency counts are typically small. For example, the mean (maximum) loss frequency of the four breach types are 2.41 (13), 3.28 (24), 0.78 (8), and 4.08 (18), respectively. Second, a significant proportion of the frequency count takes a value of zero. For example, the breach type INSD has the highest proportion of zero counts of 56.99%, while the breach type ELET has 13.99% zero counts which is the lowest among the four types. Third, all four types of data breach counts are highly auto-correlated. For instance, the auto-correlation of HACK is as high as 0.64 and the auto-correlation coefficient of ELET is 0.51.

To further investigate the serial and cross-sectional dependence structure of the four fre-

[5] Eling and Jung (2018) also suggest a categorization according to organization types. Our empirical analysis in Section 5 will focus on classification according to breach types due to length restriction. We find consistent results across different organization types, and we collect the estimation results for organization types in Appendix C.

[6] The weekly frequency has also been adopted by Jung (2021). By using weekly counts, we obtain a much longer time series, which hugely facilitates the parameter estimation of our model. Similarly, higher-frequency count data have also been analyzed in the cyber risk literature by, e.g. Xu et al. (2017); Peng et al. (2017). Other choices of frequencies include bi-weekly (Jung, 2021), monthly (Wheatley et al., 2016; Eling and Jung, 2018; Eling and Wirfs, 2019; Xu and Zhang, 2021), bi-annually (Wheatley et al., 2021), and yearly (Wheatley et al., 2016).

Table 1: **Summary statistics of cyber loss data.** This table presents summary statistics of cyber loss data including frequency and severity. Loss data are classified by breach types according to the definition of Privacy Rights Clearinghouse (PRC) and Eling and Jung (2018).

Panel A: List of types of data breaches	
Label	Event
DISC	Unintended disclosure (e.g., sensitive information posted publicly, mishandled or sent to the wrong party via publishing online, sending in an email, sending in a mailing or sending via fax).
HACK	Hacked by an outside party or infected by malware.
INSD	Insider breach (e.g., employee, contractor or customer).
ELET	Lost, discarded or stolen physical, portable device or stationary computer loss.

Panel B: Loss Frequency (weekly number of events for each event type)											
Label	N obs	N zeros	% of zeros	Mean	Std	Skewness	Kurtosis	Min	$Q_{25\%}$	$Q_{75\%}$	Max
DISC	$T = 772$	146	18.91%	2.41	2.14	1.27	2.17	0	1	4	13
HACK	$T = 772$	138	17.88%	3.28	3.34	1.60	3.31	0	1	5	24
INSD	$T = 772$	440	56.99%	0.78	1.21	2.10	5.42	0	0	1	8
ELET	$T = 772$	108	13.99%	4.08	3.50	1.17	1.41	0	1	6	18
Auto-correlation:		DISC:	0.33	HACK:	0.64	INSD:	0.36	ELET:	0.51		

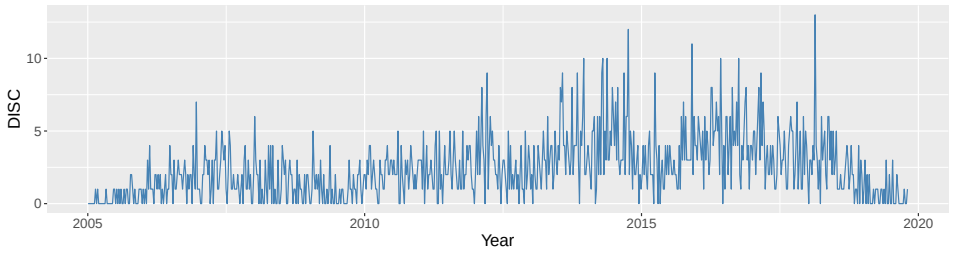
Panel C: Loss Severity (number of total breach records for each event type)											
Label	N obs	N zeros	% of zeros	Mean	Std	Skewness	Kurtosis	Min	$Q_{25\%}$	$Q_{75\%}$	Max
DISC	$N_1 = 1861$	308	16.55%	1410615	33450096	37.54	1509.26	0	180	4586	1.37E+09
HACK	$N_2 = 2533$	930	36.72%	2847809	62519680	43.99	2085.31	0	0	10000	3.00E+09
INSD	$N_3 = 606$	230	37.95%	137921.5	1468429	14.94	251.01	0	0	1329.75	2.82E+07
ELET	$N_4 = 3153$	621	19.70%	76636.61	1475875	45.27	2252.57	0	145	5806	7.60E+07
Overall	$N_{\text{total}} = 8153$	2089	25.62%	1246642	38363545	64.94	4782.69	0	0	5875	3.00E+09
M-K test (Monotonic trend)				DISC:	trend	HACK:	no trend	INSD:	no trend	ELET:	no trend
WAVK test (Any, possibly nonmonotonic trend)				DISC:	no trend	HACK:	no trend	INSD:	no trend	ELET:	no trend

quency processes, we visualize the four weekly time series in Figure 1, and report their empirical pair-wise correlation matrix in Table 2. We see that DISC and HACK has the highest correlation of 0.48, followed by the correlation between INSD and ELET (0.43). Other pairwise correlations are less than 0.2 but are still statistically significant at a level of 10% or lower.

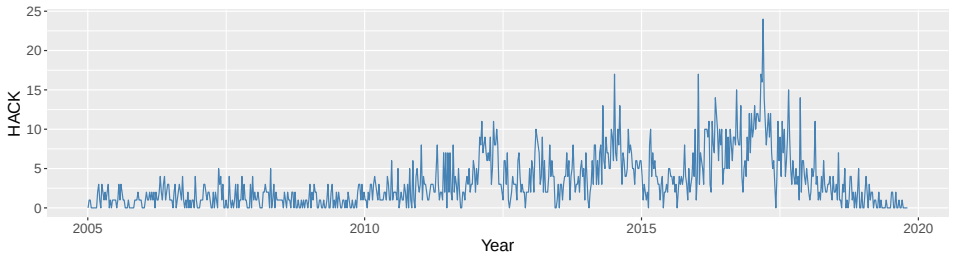
Table 2: **Cross-sectional correlation matrix of loss frequency.** This table presents empirical (cross-sectional) correlation matrix of the loss frequency data classified by breach types. $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. Since the correlation matrix is symmetric, only its upper triangular part is reported.

	DISC	HACK	INSD	ELET
DISC	1	0.48^{***}	0.10^{***}	0.14^{***}
HACK		1	0.06^{*}	0.08^{***}
INSD			1	0.43^{***}
ELET				1

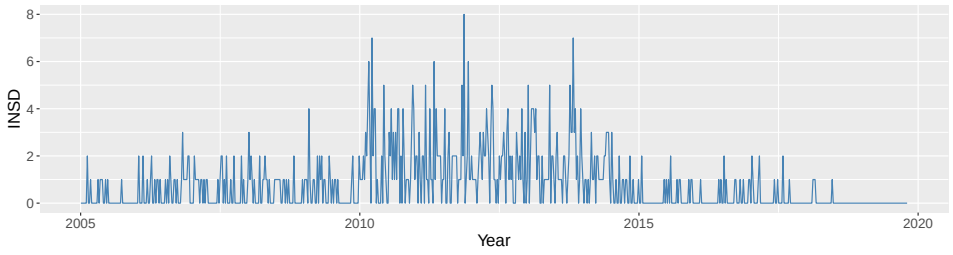
To summarize, the time series of frequencies feature both inter-temporal dependence, and cross-sectional dependence. Yet the current approaches in cyber risk literature typically do not capture both types of dependence. For instance, Eling and Jung (2018) model the cross-sectional dependence only; Peng et al. (2017) model the serial dependence of a univariate time



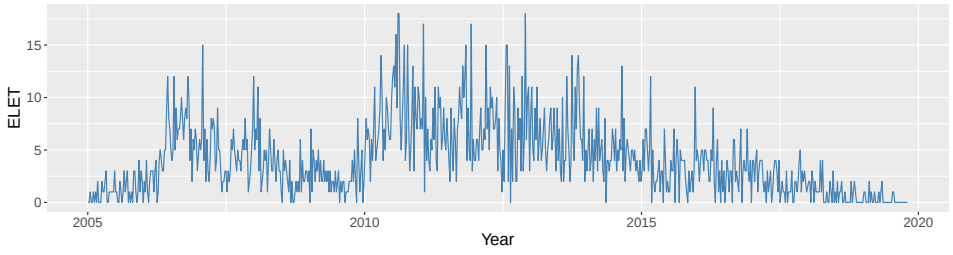
(a) DISC



(b) HACK



(c) INSD



(d) ELET

Figure 1: **Cyber loss frequency time series.** This figure visualizes weekly time series of cyber loss frequencies for four breach types DISC, HACK, INSD, and ELET.

series of counts. To the best of our knowledge, there are only two approaches that account for both the serial and cross-sectional dependences of the frequency counts. The first one is the ARMA/GARCH-copula approach in (Xu et al., 2017; Peng et al., 2018), in which univariate ARMA/GARCH (resp. copula) models are used to capture serial (resp. cross-sectional) dependence within (resp. between) individual frequency processes. However, the ARMA/GARCH model does not ensure the positivity of the frequencies, and hence has the potential of introducing regulatory arbitrage opportunities. Moreover, both the ARMA/GARCH and copula models are initially designed for continuous variables, and are therefore unable to account for the discreteness of the counts, and in particular the non-zero probability that the frequencies equal zeros. See Genest and Nešlehová (2007) for a cautionary note on the dangers and limitations of copula-based count models. This difficulty has recently also been identified in the cyber risk literature. As Eling and Jung (2018) put it, "the zero-inflation could generate a mis-specification of a dependence model due to discontinuous probability function".

The second approach accounting for serial and cross-sectional correlations of the cyber risk is through continuous-time, multivariate Hawkes process (Bessy-Roland et al., 2021). The Hawkes process is popular thanks to its tractable likelihood function, but has also several limitations. First, in most cyber risk databases known to the academic literature, time is discrete. For instance, in the PRC data, the occurrence of the events are reported on a daily basis. Even though the finance literature sometimes *approximates* daily (multivariate) counting process data using continuous-time Hawkes process (see Embrechts et al., 2011), such an approach is only appropriate if the probability of observing simultaneous events of different types is effectively zero, which is implicitly assumed in the multivariate Hawkes process model. This assumption, however, may sound restrictive for cyber risk. For instance, the 2530 events of HACK type in the PRC data occurred on 1580 different days, hence the average number of HACK events that occur within a single day is well above 1. Second, the date at which a data breach event is reported is likely subject to reporting delay, i.e., they are "Incurred But Not Reported (IBNR)" before confirmation. One can also imagine that several events reported by different companies could be triggered by the same malware at roughly the same time. As a consequence, treating time as effectively discrete allows to minimize the impact of such measurement error on our statistical analysis. Third, in many real-time forecasting problems,

the horizon is an integer unit of time.[7] Even though the *instantaneous* intensity is tractable for a (multivariate) Hawkes process, the derivation of the forecasting formula for a positive horizon is usually rather complicated.[8] Finally, it has been well documented that the number of cyber events per unit time usually features over-dispersion (see, e.g., Eling and Loperfido, 2017; Sun et al., 2020), and hence, negative binomial type models typically outperform Poisson models. The Hawkes process, however, is a dynamic extension of the Poisson process and is hence less convenient to capture this over-dispersion feature.

As a partial summary, existing frequency models in the cyber risk literature either do not account for serial and cross-sectional dependence, or are based on continuous models that are unsuitable for count data. Hence we need a (multivariate and dynamic) discrete model that is capable of replicating the aforementioned stylized facts of the frequency data.

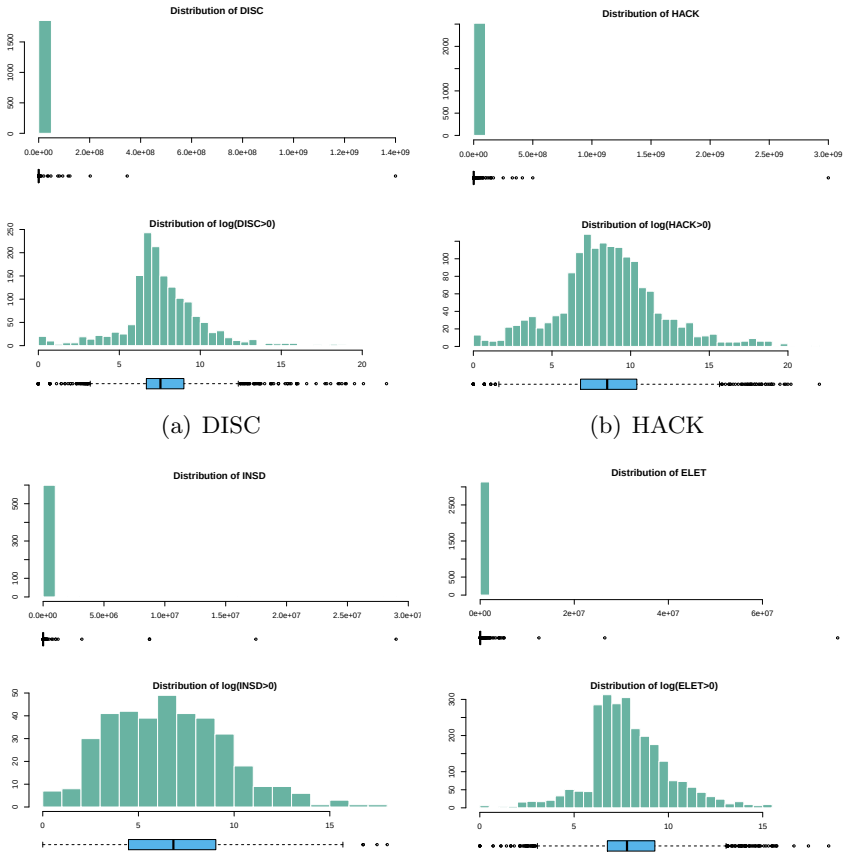
2.2 Data on severity

We now move onto the severity component. Panel C of Table 1 provides summary statistics of loss severity data for each of the 4 types of events. We see that the severity component is associated with two stylized facts. First, the data breach records are integer-valued,[9] and has a significant probability of taking a value of zero. For example, among the 8153 observations, nearly 25.62% have zero records. Second, the distribution of the records is strongly heavy-tailed. From Table 1, we see that the sample average is much larger than the 75% percentile, and all sections are extremely skewed and leptokurtic. Histograms of the four types of breach severity in Figure 2 clearly visualize these two properties.

[7] Depending on the applications, this unit might be one year, in the case of computation of Value-at-Risk for a bank (Chavez-Demoulin et al., 2005, 2016), or one hour (Peng et al., 2017).

[8] Bessy-Roland et al. (2021) show that the computation of the *unconditional* expectation of the process involves matrix equations of dimension $K(K+1) \times K(K+1)$, where K is the dimension of the Hawkes process. The *conditional* expectation given observation up to a certain time t has not been derived.

[9] Some of the commercial databases, such as SAS OpRisk Global data (e.g., Eling and Wirfs, 2019; Eling and Schnell, 2020), or IBM's annual Cost of Data Breach Report (Eling and Jung, 2018), also provide the monetary cost of these cyber events. However, these datasets are not publicly available and most existing cyber risk studies rely on data which provide integer-valued severity variables. See also Section 6 and Appendix B for further discussion.



(a) DISC (b) HACK

(c) INSD (d) ELET

Figure 2: **Distributions of the severity of cyber events.** This figure plots distributions of loss severity for four breach types, including DISC, HACK, INSD, and ELET. For each figure, top panel plots the histogram and box plot of the whole distribution of each breach type; bottom panel plots the histogram and box plot of the logarithmic transformation of positive losses for each breach type.

To the best of our knowledge, none of the existing models accommodate for both properties. Indeed, even though some studies take into account the heavy tail (see, e.g,. Chavez-Demoulin et al., 2005; Maillart and Sornette, 2010; Chavez-Demoulin et al., 2016; Edwards et al., 2016; Eling and Loperfido, 2017; Farkas et al., 2021; Jung, 2021), all these models rely on continuous distributions, by either removing events associated with zero-loss records, or only considering events whose severity exceeds a certain threshold.

Such a continuous approach has several limitations. First, a "zero" (small) record does not necessarily imply zero (small) economic loss. Indeed, once an IT system is breached, regardless of the final number of loss records (which might take a long while to determine), the information

system should be fixed, daily operations of that organization might be halted, the reputation could be damaged, leading to client and investor defection (Cummins et al., 2006). Such indirect costs are currently unincorporated into many major cyber risk databases (Eling and Wirfs, 2019), but have already become a huge concern for many organizations seeking cyber risk coverage. This explains why many of the current stand-alone cyber insurance policies provide coverage for costs aimed at minimizing (potentially long-term) reputation damages.[10] Similarly, the dataset also includes a few events with a small number of records. For example, 128 events (or 1.6%) in our sample involve small records that are between 1 and 10. This probability of observing a small, positive severity should not be omitted, either, since for some cyber events, only a small number of selected, highly sensitive records are targeted. In this case, even a small number of data loss can lead to a significant dollar loss.[11]

Second, by removing these zero-loss observations, the frequency of the cyber events is under-estimated. Let us for instance imagine a malware which breaches the server of several organizations, and causes data losses for a subset of them. If we do not account for the attacks suffered by companies that end up with zero loss of records, then we under-estimate the transmissibility of the malware. We can also draw an analogy with credit risk analysis, where loan default can be associated with a zero Loss-Given-Default (LGD), but should still be counted as a default since even a default with $LGD = 0$ may incur indirect costs such as legal fees, collection fees, etc. (see e.g. Tanoue et al., 2017). As a consequence, we follow Sun et al. (2020), Bessy-Roland et al. (2021), and Farkas et al. (2021) to include all the data breach events, including those with zero data breach records. Moreover, in order to correctly characterize the probability that the observed severity takes small, integer values, in the next section we will introduce a fully discrete model, instead of approximating this severity distribution with continuous models.

We also test for possible time trend in the severity data with test results listed in the last two rows in Panel C of Table 1. We consider two trend tests, one being Mann-Kendall test (Mann, 1945; Kendall, 1975), which tests for monotonic trends, and the other being WAVK

[10] According to the RMS-CCRS review, 46% of the reviewed policies included coverage for reputation damage (RMS-CCRS, 2016).

[11] For instance, on 9/13/2016, the World Anti-Doping Agency (WADA) confirmed the hack of its database on medical files of Olympic athletes by the Russian group Fancy Bear. This event led to only 4 lost records, but its economic consequence could be difficult to evaluate, yet potentially large, especially if the stolen information was used by rival athletes.

test (Wang et al., 2008), which tests for any, potentially non-monotonic trend. We see that the WAVK test indicates no trend for all breach types. For Mann-Kendall test, except for DISC, all the other three breach types are tested no monotonic trend.[12]

3 The model

3.1 Frequency component

In Section 2.1 we have mentioned the importance of accounting for both serial and cross-sectional dependence between the multiple time series of counts. In order to identify the most suitable candidate process, let us now briefly review the relevant time series literature. During the last two decades, a large number of univariate count time series models have been proposed (see Davis et al., 2021, for a review). However, research on their multivariate counterparts has been much more sparse, and most of which are not suitable for the application we have in mind.

For instance, Jørgensen et al. (1999), Jung et al. (2011), Livsey et al. (2018), and Catania and Di Mari (2021) propose models with latent factors that are suitable for potentially high-dimensional count time series, but the estimation and forecasting in these models are computationally intensive because of the need to integrate out the latent factor. In addition, Heinen and Rengifo (2007) and Fokianos et al. (2020) propose models without latent factors. These alternative models are easy to estimate and forecast, but the cross-sectional dependence in these models is captured through copulas. The downside of the copula approach is that *i*) the stationarity of such dynamic models is usually not clear; *ii*) when applied to counts, it remains an open question as to whether the copula is well identified (see Genest and Nešlehová, 2007; Trivedi and Zimmer, 2017, for more discussion). Finally, only a few recent models allow for both flexible conditional distributions and simple estimate/forecasting procedures (Darolles et al., 2019; Gouriéroux and Lu, 2019), but this literature is up to now focused on the bivariate case due to the curse of dimensionality. In this paper, we propose a multivariate extension of the (bivariate) negative binomial autoregressive (NBAR) process of Gouriéroux and Lu (2019)

[12]Trend in cyber data has been discussed in the literature. For example, Edwards et al. (2016) report no time trend for the severity over the past decade. Eling and Loperfido (2017) report no clear upward or downward trend in the PRC data from January 10, 2005, to December 15, 2015. According to Eling and Jung (2018), they empirically do not observe any non-stationarity or serial dependence for different risk types. Jung (2021) reports a break in loss severity trend using 50 days rolling windows.

for the time series of counts $Y_t = (Y_{1,t}, .., Y_{K,t})'$, where K is the dimension of frequency counts at each period. For instance, in the empirical application in Section 5, where we classify events by breach types, we take $K = 4$. The methodology remains valid and computationally realistic for other mildly large values of K.[13]

The extension we propose involves the Wishart-gamma distribution, recently introduced by Denuit and Lu (2021). This latter is the diagonal marginal distribution of the Wishart distribution, which in turn is the (symmetric positive definite) matrix extension of the gamma distribution.

Definition 1 (Wishart-gamma distribution)**.** A K-dimensional positive random vector $X = (X_1, .., X_K)'$ follows the Wishart-gamma distribution if its Laplace transform is given by:

$$\mathbb{E}\left[e^{-s_1X_1-\cdots-s_KX_K}\right] = \frac{1}{\left|\mathrm{Id}_K + 2\mathrm{Diag}(s_1, s_2, \cdots, s_K)\Sigma\right|^{\frac{\delta_0}{2}}}, \qquad \forall\ s_1, .., s_K \geq 0, \tag{1}$$

where Id_K is the K-dimensional identity matrix, symbol $|\cdot|$ denotes the determinant of a matrix, the degree of freedom parameter δ_0 is either a positive integer, or a real number not smaller than $K-1$,[14] and (matrix) scale parameter Σ is $(K \times K)$ symmetric positive definite.

In Equation (1), by letting $s_1, ..., s_K$ equal to zero except $s_i, s_j \geq 0$, where $i \neq j$, we get the joint Laplace transform of (X_i, X_j):

$$\mathbb{E}\left[e^{-s_iX_i-s_jX_j}\right] = \frac{1}{\left[1 + 2s_i\sigma_{ii} + 2s_j\sigma_{jj} + 4s_is_j(\sigma_{ii}\sigma_{jj} - \sigma_{ij}^2)\right]^{\frac{\delta_0}{2}}}, \qquad \forall\ s_i, s_j \geq 0.$$

In particular, the joint distribution of (X_i, X_j) is bivariate gamma, and each component X_j is marginally gamma $\gamma\left(\frac{\delta_0}{2}, 2\sigma_{jj}\right)$ distributed, and simple algebra shows that their correlation coefficient is:

$$\mathrm{Corr}[X_i, X_j] = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}} \geq 0, \qquad \forall\ i, j = 1, .., K. \tag{2}$$

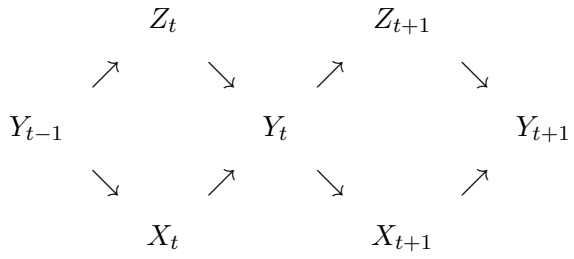
Thus, even though the off-diagonal elements of Σ can be negative, the components of X will always feature nonnegative correlation.[15]

[13] For instance, if we were to count the cyber events in PRC by organization types as in Eling and Jung (2018), then we get $K = 5$. These results are collected in Appendix C.

[14] Under some conditions, the domain of δ_0 can be further relaxed, see Denuit and Lu (2021) for a discussion.

[15] Even though the covariance formula (2) does not depend on the sign of the entries Σ, the off-diagonal

Because of its closed-form Laplace transform, the Wishart-gamma distribution is very convenient when it comes to correlating different count variables in a cross-sectional setting (Denuit and Lu, 2021). However, recall that the multivariate frequency process we are interested in features *both* cross-sectional *and* serial correlations. Thus in order to account for the serial correlation between the different counts observed at different periods, we now use this Wishart-gamma distribution as a building block to define a time series of multivariate counts $Y_t = (Y_{1,t}, .., Y_{K,t})'$. This process is Markov, and is defined alternately with two intermediate, latent factors, Z_t and X_t, using the following causal scheme:



(3)

Here we decompose the intermediate factors into two parts, Z_t and X_t, and it will become clear shortly afterwards that these two factors represent different contagion channels in the transition rules from Y_{t-1} to (Z_t, X_t) and then to Y_t.

Definition 2 (K-dimensional negative-binomial autoregressive (NBAR) process)**.** The K-dimensional NBAR process is such that:

- Given $\underline{Y_t} := (Y_t, Y_{t-1}, ..., Y_1, ...)$, variables $Z_{j,t+1}, j = 1, .., K$, are independent, with conditional gamma distribution $\gamma(\delta_j + \beta_j' Y_t, c_j)$, respectively, where $\beta_j' Y_t = \beta_{j,1} Y_{1t} + \cdots + \beta_{j,K} Y_{Kt}$, and all the components of $\beta_j = (\beta_{j,1}, \ldots, \beta_{j,K})'$ as well as δ_j and c_j are nonnegative for each $j = 1, ..., K$.

- The conditional distribution of the common intensity factors $X_{t+1} = (X_{1,t+1}, .., X_{K,t+1})$ is Wishart-gamma with conditional Laplace transform:

$$\mathbb{E}\left[e^{-s_1 X_{1,t+1} - \cdots - s_K X_{K,t+1}} | Y_t\right] = \frac{1}{\left|\mathrm{Id}_K + 2\mathrm{Diag}(s_1, s_2, \cdots, s_K)\Sigma\right|^{\frac{\delta_0}{2} + \alpha' Y_t}}, \quad \forall\ s_1, ..., s_K \geq 0. \tag{4}$$

elements of Σ should not be constrained to be nonnegative. Indeed, the distribution of X is not multivariate normal, hence it is not fully-characterized by its first two moments.

where $\alpha' Y_t = \alpha_1 Y_{1t} + \cdots + \alpha_K Y_{Kt}$, with α_j nonnegative for each $j = 1, ..., K$.

- The counts $Y_{j,t+1}, j = 1, ..., K$ are conditionally independent given $\underline{X_{t+1}}$, $\underline{Y_t}$ and $Z_{j,t+1}$, following Poisson distributions with parameters $X_{j,t+1} + Z_{j,t+1}$, respectively.

The process introduced above is slightly different from the original, bivariate version of Gouriéroux and Lu (2019) for parsimony reasons.[16] Nevertheless, we continue to call it NBAR, since in the limiting case where Σ reduces to the null matrix, each $Y_{i,t+1}$ is conditionally mixed Poisson distributed with gamma mixing distribution (and hence negative binomial) given the past observation Y_t.

In the causal chain (3), the stochastic properties of the two latent (multivariate) processes (Z_t) and (X_t) are different: given Y_{t-1}, the components of Z_t are conditionally independent, whereas those of X_t feature positive dependence [see equation (2)]. Thus, even though both unobserved factors capture the causality of Y_{t-1} and Y_t, they correspond to two different channels of *contagion*: the components of Z_t are conditionally independent given Y_{t-1}, with each component $Z_{j,t}$ measuring the effect of previous event counts $Y_{1,t-1}, ..., Y_{K,t-1}$ on $Y_{j,t}$. In other words, Z_t captures the analog of the self-[17] and mutual-excitation[18] in continuous-time Hawkes process (Aït-Sahalia et al., 2014; Bessy-Roland et al., 2021), but by further allowing for stochastic intensity. By contrast, the components of X_t are conditionally dependent, and positively correlated given the past. Hence X_t captures the possibility that some events simultaneously cause several types of future events. To summarize, our model is one of the first in finance and insurance literature to allow for two separate mechanisms of contagion, with the latent multivariate factor Z_t accounting for the ***pairwise*** contagion channel, and X_t capturing the ***systemic*** contagion channel. These two channels differ by the fact that only the systemic channel is capable of generating simultaneous events. [19]

The introduction of stochastic intensities is a natural way to account for over-dispersion,

[16] This latter involves quadratic forms of bivariate Wishart stochastic matrices. The problem of this specification is that at dimension K, each such quadratic form involves $K(K+1)/2$ parameters, and there are as many quadratic forms as there are components of Y_t. In contrast, our new model only involves the Wishart-gamma distribution, which is the diagonal of the Wishart distribution and is hence much more parsimonious.

[17] The effect of self-excitation is captured by parameters $\beta_{j,j}, j = 1, ..., K$.

[18] The effect of mutual-excitation is captured by parameters $\beta_{i,j}, i \neq j$.

[19] Note that the term "systemic contagion" does not necessarily mean that one systemic event generates simultaneously cyber events of different types. This terminology is mainly introduced to emphasize the fact that in contrast to self-excitation which generates counts that are conditionally independent given the past, the systemic channel, through the common factor X_t, captures the positive correlation of $Y_{1,t+1}, ..., Y_{K,t+1}$ given past information.

which, in a dynamic context, is also the synonym of "higher" risk than implied by Poisson models. However, our approach differs from other latent factor-based count process models (Jørgensen et al., 1999; Jung et al., 2011), or the analogous continuous-time Cox doubly stochastic intensity model, which are generally computationally intractable. In the latter models, the latent factor has its own exogenous dynamics, whereas in our model, the dynamics of the latent intensity factors Z_t, X_t given all the past depends on Y_t only. This will make the model much more tractable.

Finally, in our modeling, we have assumed that the count process (Y_t) is (first-order) Markov. Even though it is straightforward to extend a first-order Markov model to a higher-order one while at the same time staying within the affine process framework (Darolles et al., 2006), for the cyber risk application, we argue that this extension is not strictly necessary. Indeed, cyber risk likely evolves very fast (Wheatley et al., 2021; Eling and Wirfs, 2019), and because our counts are aggregated at a weekly frequency, even a first-order model is capable of capturing the contagion effect in the next week. Therefore, the benefit of introducing a more parameter-heavy higher-order model is likely smaller than the extra burden in terms of model estimation.

We are now ready to give the economic interpretations of model parameters. We will further illustrate the magnitudes and significance of these parameters in the empirical application in Section 5.

- $\alpha_1, .., \alpha_K$ measure the contribution of each of the individual processes to the systemic risk.
- $\sigma_{11}, .., \sigma_{KK}$ measure the impact of the systemic risk on individual processes.
- Off-diagonal terms $\sigma_{ij}, i \neq j$, of matrix Σ capture the instantaneous dependence between individual processes due to their exposure to the systemic risk.
- Vectors $\beta_1, .., \beta_K$ measure the spillover effects between different individual processes, through the direct channel.
- $\delta_1, .., \delta_K$ measure the propensity of idiosyncratic events that cannot be interpreted as systemic or pairwise contagion.

- Scalar parameters $c_1, .., c_K$ account for the heterogeneity in level of the different components of Y_t.

We now discuss some properties of the multivariate count process $Y_t = (Y_{1,t}, .., Y_{K,t})'$.

Mean-variance predictions For the conditional first moment we have:

$$\mathbb{E}[Y_{j,t}|Y_{t-1}] = c_j \left(\delta_j + \beta_j' Y_{t-1}\right) + 2\sigma_{jj} \left(\frac{\delta_0}{2} + \alpha' Y_{t-1}\right), \quad j = 1, .., K. \tag{5}$$

These equations can be conveniently written into the matrix form:

$$\begin{aligned}\mathbb{E}[Y_t|Y_{t-1}] &= \text{Diag}(c)(\Delta + BY_{t-1}) + 2\left(\frac{\delta_0}{2} + \alpha' Y_{t-1}\right) S \\ &= \text{Diag}(c)\Delta + \delta_0 S + \left[\text{Diag}(c)B + 2S\alpha'\right] Y_{t-1},\end{aligned} \tag{6}$$

where vector $S = (\sigma_{11}, \sigma_{22}, \cdots, \sigma_{KK})'$ is the diagonal of the symmetric positive definite matrix Σ; $\text{Diag}(c)$ is the diagonal matrix with diagonal $(c_1, ..., c_K)'$; column vector $\Delta = (\delta_1, \cdots, \delta_K)'$; and matrix $B = (\beta_{ij})_{i,j=1,..,K}$.

Similarly, by using the total (co-)variance formula, we can get the conditional (co-)variance of Y_t. For $j, k = 1, .., K$, we have:

$$\begin{aligned}\mathbb{C}ov[Y_{j,t}, Y_{k,t}|Y_{t-1}] &= \mathbb{C}ov[\mathbb{E}[Y_{j,t}|X_{j,t}, Z_{j,t}], \mathbb{E}[Y_{k,t}|X_{j,t}, Z_{j,t}]|Y_{t-1}] \\ &\quad + \mathbb{E}[\mathbb{C}ov[Y_{j,t}, Y_{k,t}|X_{j,t}, Z_{j,t}]|Y_{t-1}] \\ &= \mathbb{C}ov[X_{j,t} + Z_{j,t}, X_{k,t} + Z_{k,t}|Y_{t-1}] + \mathbb{1}_{j=k}\mathbb{E}[X_{j,t} + Z_{j,t}|Y_{t-1}] \\ &= \mathbb{1}_{j=k}\Big[(c_j + c_j^2)(\delta_j + \beta_j' Y_{t-1}) + 2\sigma_{jj}\left(\frac{\delta_0}{2} + \alpha' Y_{t-1}\right)\Big] \\ &\quad + 4\sigma_{ij}^2\left(\frac{\delta_0}{2} + \alpha' Y_{t-1}\right),\end{aligned} \tag{7}$$

where we have used the fact that the conditional distribution of $X_{j,t}$ given Y_{t-1} is gamma, and the indicator $\mathbb{1}_{j=k}$ if equal to 1 if and only if $j = k$, and 0, otherwise.

Nonlinear prediction through probability generating function (pgf) Because (Y_t) are count valued, the linear forecasting formula does not characterize the predictive distribution of the future counts. Thus let us now work out the nonlinear prediction formula, through the con-

ditional joint probability generating function (pgf). By the iterative expectation formula, for any $u_j \in [0,1], j = 1,..,K$, we have:

$$\begin{aligned}
\mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right] &= \mathbb{E}\Big[\mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | X_{t+1}, Z_{t+1}\right] | Y_t\Big] \\
&= \mathbb{E}\Big[\exp\left[(X_{1,t+1} + Z_{1,t+1})(u_1 - 1) + \cdots + (X_{K,t+1} + Z_{K,t+1})(u_K - 1)\right] | Y_t\Big] \\
&= \frac{1}{[1 + c_1(1-u_1)]^{\delta_1 + \beta_1' Y_t}} \cdots \frac{1}{[1 + c_K(1-u_K)]^{\delta_K + \beta_K' Y_t}} \\
&\quad \times \frac{1}{\left|\mathrm{Id} + 2\mathrm{Diag}(1-u_1, 1-u_2, \cdots, 1-u_K)\Sigma\right|^{\frac{\delta_0}{2} + \alpha' Y_t}}. \qquad (8)
\end{aligned}$$

Because the conditional joint pgf (or equivalently the conditional Laplace transform) is an exponential affine function of the conditioning variable Y_t, this is an example of affine process (Darolles et al., 2006). This type of processes has been shown to have appealing analytic properties for insurance and finance applications (Gourieroux et al., 2006; Lu, 2018).

The stationary condition Roughly speaking, process (Y_t) is stationary, if there is neither deterministic, nor stochastic trend leading to an explosion of process (Y_t), when time t goes to infinity. The problem of testing for trend of cyber risk has previously been addressed by Edwards et al. (2016), but their analysis only allows for deterministic trend.[20]

By iterating equation (5), we get:

$$\mathbb{E}[Y_{t+h} | Y_t] = M^h Y_t + \left(M^{h-1} + \cdots + \mathrm{Id}\right)[\mathrm{Diag}(c)\Delta + \delta_0 S], \qquad (9)$$

where

$$M = (m_{i,j}) := \mathrm{Diag}(c)B + 2S\alpha'. \qquad (10)$$

The matrix M has insightful economic meaning: each element of M, $m_{i,j}$, can be interpreted as the average number of contagious event of type j generated by a previous event of type i. Based on M, the stationarity condition for process (Y_t) can be deduced easily as follows:

Proposition 1. Process (Y_t) is strictly stationary[21] if and only if all the eigenvalues of matrix M is smaller than 1 in modulus.

[20] See Chavez-Demoulin et al. (2016) for a similar, deterministic approach for the modeling of operational risk.

[21] We say that process (Y_t) is strictly stationary, if any finite dimensional marginal distribution of $(Y_{t+k_1}, Y_{t+k_2}, ..., Y_{t+k_n})$, where $n, k_1, ..., k_n$ are integers, are time-invariant.

Thus the spectral radius $\rho(M)$ of matrix M, which is its biggest eigenvalue, is a measure of the persistence of the vector process (Y_t). In the limiting case where $\rho(M)$ is equal to 1, the components of (Y_t) become random walks, corresponding to an increasing trend of operational risk occurrence. If, on the other hand, $\rho(M)$ is smaller than 1, then the cyber risk is not exploding.

3.2 Severity component

Let us now introduce the Generalized Poisson inverse-Gaussian (GPIG) family of distributions for the severity component. This count model, first introduced by Zhu and Joe (2009),[22] is the only one we are aware of that is compatible with the unique data features, including zero-inflation and heavy tail.

Definition 3 (GPIG distribution)**.** A random variable W is said to follow the GPIG distribution, denoted as $GPIG(a; b; c)$, if its pgf is:

$$\mathbb{E}\left[u^W\right] = \exp\left\{b[(1-c)^a - (1-cu)^a]\right\}, \qquad \forall\ u \in [0,1], \tag{11}$$

where shape parameter $a \in (0, 1]$, scale parameter $b > 0$, and probability parameter $c \in [0, 1]$.[23]

Assumption 1 (Distribution of the severity variable)**.** For each event type j, $j = 1, \ldots, K$, the total breach records, N_j, $j = 1, \ldots, K$, are i.i.d. following GPIG(a_j, b_j, c_j) distribution, where a_j, b_j, c_j are event-specific model parameters.

This assumption implies that for each given event type, the frequency and severity variables are independent. This assumption is rather standard in the cyber-risk literature and beyond (see, e.g., Chavez-Demoulin et al., 2016; Eling and Jung, 2018; Farkas et al., 2021; Bessy-Roland et al., 2021).[24] Moreover, the "i.i.d." assumption also means that we do not consider

[22] See also El-Shaarawi et al. (2011) for an application in environmental sciences.

[23] See Zhu and Joe (2009) for the interpretation of these three parameters in terms of shape, scale, and probability, respectively.

[24] Recently, Sun et al. (2020) investigate the dependence between the severity and frequency at the *firm*-level and found a nonlinear, positive dependence. However, their analysis is based on firms that have reported claims only, and is hence subject to *selection bias*. Moreover, their approach does not allow for serial correlation, and cannot be applied at the *portfolio* level. Because events are disaggregated into different breach categories, some of the heterogeneity has already been taken into account. As a consequence, the independence assumption between frequency and severity for each of the event types seems reasonable.

the possibility that claim size features a trend across time, as we have observed no trend in our data from the Panel C of Table 1.

Let us now study some properties of the GPIG distribution and explain why it is particularly appropriate for the loss severity we observe. We first look at a few special cases.

- When $a = 1$ and $c \neq 0$, we get the pgf of Poisson with mean bc.
- When $a = \frac{1}{2}$, we get the Poisson-Inverse Gaussian distribution (Willmot, 1987), that is, a mixed Poisson distribution with inverse-Gaussian mixing distribution, hence the terminology GPIG distribution.
- When $c = 1$ and $a < 1$, we get $\mathbb{E}\left[u^W\right] = \exp\left[-b(1-u)^a\right]$, which is the pgf of a discrete stable distribution (Steutel and van Harn, 1979; Christoph and Schreiber, 1998). This distribution is again a mixed Poisson distribution, in which the mixing distribution is (one-sided) α-stable with support $\mathbb{R}_{>0}$, and Laplace transform:

$$\mathbb{E}\left[e^{-uW}\right] = e^{-bu^a}, \qquad \forall\ u \geq 0.$$

  In this case, parameter a (resp. b) is also the stability (resp. scale) parameter of the α-stable distribution. Because this latter distribution has an infinite mean when $a \in (0,1)$, the discrete stable distribution also has infinite mean. This special case will be of particular interest in our empirical investigation, given the heavy-tail phenomenon highlighted in Section 2.

In the general case, when both a and c are strictly small than 1, Zhu and Joe (2009) show that this distribution is mixed Poisson, with the mixing distribution being the tempered stable distribution (Hougaard, 1986; Carr et al., 2002), also called Tweedie distribution or power variance function distribution (Jørgensen, 1987). The support of this mixing distribution is $(0,\infty)$, and its density is obtained by exponentially tilting an α-stable distribution with shape parameter a (Hougaard, 1986).

Expectation Replacing u in Equation (11) by e^{-s}, where $s > 0$, we obtain the Laplace transform of W:

$$\mathbb{E}\left[e^{-sW}\right] = \exp\left\{b\left[(1-c)^a - (1-ce^{-s})^a\right]\right\}. \tag{12}$$

Taking the first-order Taylor's expansion with respect to s at $s = 0$, we get:

$$\mathbb{E}[W] = abc(1-c)^{a-1}. \tag{13}$$

In particular, when c goes to 1, the expectation $\mathbb{E}[W]$ goes to infinity and in the limiting case, we recover the fact that the discrete stable distribution has infinite mean.

Probability mass function (pmf) By taking $u = 0$ in Equation (11), we get:

$$\mathbb{P}[W = 0] = \exp\Big\{b[(1-c)^a - 1]\Big\}. \tag{14}$$

The expression of $\mathbb{P}[X = n]$ for $n \geq 1$ can be obtained by computing the Taylor's expansion of Equation (11) with respect to u to the order n. For instance, for $n = 1$, we get:

$$\mathbb{P}[W = 1] = \exp\Big\{b[(1-c)^a - 1]\Big\}abc. \tag{15}$$

Such a Taylor's expansion, however, becomes more and more cumbersome to compute for large n.[25]

Let us now discuss the implications of Equations (13), (14) and (15) and study the zero-inflation property of the GPIG distribution.

Zero-inflation property For $c < 1$ such that the mean $\mathbb{E}[W]$ is finite, we compare this distribution with a Poisson distribution with the same mean. This latter distribution's pmf (p_n^*) is such that:

$$p_0^* = \exp\Big\{-abc(1-c)^{a-1}\Big\}, \tag{16}$$

$$p_1^* = abc(1-c)^{a-1}\exp\Big\{-abc(1-c)^{a-1}\Big\}. \tag{17}$$

Proposition 2. We have,

[25] Zhu and Joe (2009) propose an alternative recursive formula to compute the pmf, but the computational cost of this recursion is still very high for large n and it is known that such recursion may have large numerical error (Wang and Panjer, 1993).

- The mass at zero $\mathbb{P}[W = 0]$ is such that:

$$\mathbb{P}[W = 0] > p_0^*. \tag{18}$$

- The odds ratio $\dfrac{\mathbb{P}[W = 1]}{\mathbb{P}[W = 0]}$ is such that:

$$\frac{\mathbb{P}[W = 1]}{\mathbb{P}[W = 0]} < \frac{p_1^*}{p_0^*}. \tag{19}$$

Inequality (18) means that the distribution of W is zero-inflated compared to the Poisson distribution. In the limiting case, when c goes to 1 while a and b remain constant, we have $\mathbb{P}[W = 0] \to e^{-b}$ whereas $p_0^* \to 0$. Thus, the more its distribution is heavy-tailed, the more it is zero-inflated. Similarly, Inequality (19) implies that the point mass at 1 is less important than the point mass at 0 for the GPIG distribution than a Poisson distribution with the same mean.

Proof. See Appendix A. □

Finally, an open question in the literature on the PRC database is how the number of breach records is related to the actual economic losses (Eling and Loperfido, 2017), which is not reported in the data. This issue is clearly out of the scope of the present paper, given the popularity of this database. Nevertheless, in Appendix B, we discuss how a GPIG-based count model can be used to evaluate the expected economic cost of the events under a simple relational model proposed in the literature.

4 Estimation methodology

In this section, we discuss the estimation of our framework. We will start with the frequency component, before proceeding to the severity model. In both cases, we propose to use Generalized Method of Moments (GMM), based on (conditional) pgf. As a standard workhorse of modern econometrics, the GMM estimator is discussed in various leading textbooks. We refer to Mátyás (1999) for an introduction of GMM and Hall (2005) for applications of GMM to time series data.

4.1 Estimating the frequency model

Because the dynamics of (Y_t) is most conveniently characterized by the conditional pgf (or equivalently the conditional Laplace transform, or the characteristic function), the most natural estimation approach is the GMM.[26] More precisely, for any combinations of $u = (u_1, u_2, \cdots, u_K)$ belonging to a set of arguments $\mathcal{U} \subset [0,1]^K$, we consider the moment conditions:

$$\mathbb{E}\Big[\psi_{u_1,u_2,\cdots,u_K}(Y_t)\Big(u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} - \mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right]\Big)\Big] = 0, \qquad \forall\ u \in \mathcal{U}, \tag{20}$$

where $\psi_{u_1,u_2,\cdots,u_K}(Y_t)$ can be any measurable function of Y_t, called the instrumental function. For convenience, in this paper, we choose the instrumental function to be:

$$\psi_{u_1,u_2,\cdots,u_K}(Y_t) = \frac{1}{\sqrt{\mathbb{V}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right]}}, \tag{21}$$

where

$$\mathbb{V}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right] = \mathbb{E}\left[u_1^{2Y_{1,t+1}} \cdots u_K^{2Y_{K,t+1}} | Y_t\right] - \left\{\mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right]\right\}^2,$$

which has a closed form by the formula of the joint pgf [see Equation (8)].

Indeed, this moment condition has the advantage that the integrand in Equation (20) has unitary conditional variance given Y_t:

$$\mathbb{E}\left[\psi^2_{u_1,u_2,\cdots,u_K}(Y_t)\left(u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} - \mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t\right]\right)^2 | Y_t\right] = 1.$$

In other words, the temporal heteroscedasticity of the process (Y_t) is accounted for.

The GMM estimator of the vector of parameters $\theta = (\Delta', c', vec(B), vec(\Sigma), \alpha', \delta_0)'$ is obtained by minimizing the quadratic function:

$$\hat{\theta}_T = \arg\min_{\theta} h(\theta)' V h(\theta),$$

[26] Note, however, that it is possible to compute the conditional pmf of a multivariate count process by series expansion, if the conditional pgf is available in closed form. See Lu (2021) for a treatment of the univariate case, and Darolles et al. (2019) for an example in dimension $K = 2$. The disadvantage of this approach is that the computation becomes much more costly in high dimensions, say $K = 4$, even though one could rely on pairwise conditional pgf $p(Y_{t,i}, Y_{t,j}|Y_{t-1})$, $1 \leq i \neq j \leq 4$, and estimate the model using (pairwise) composite likelihood approach (see Pedeli and Karlis, 2013, for an example in multivariate count time series analysis). As a consequence, in this paper, we focus on the GMM approach, which is much more computationally tractable.

where V is a symmetric positive definite matrix known as the weighting matrix; vector function $h = (h_1, ..., h_L)'$ has the same dimension L as the number of moment conditions, with its l-th component computing the empirical counterpart of the left hand side of Equation (20). More specifically,

$$h_k(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} \psi_{u_1,u_2,\cdots,u_K}(Y_t) \left(u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} - \mathbb{E}\left[u_1^{Y_{1,t+1}} \cdots u_K^{Y_{K,t+1}} | Y_t \right] \right).$$

The number of moment conditions L is chosen to be larger than the dimension of θ to ensure asymptotic consistency of the GMM estimator. We follow the standard two-step GMM estimation procedure (Hansen, 1982). That is, we first use an identity weighting matrix and obtain the first stage estimator $\hat{\theta}_{T,1}$. Then we use the optimal weighting matrix, given by the inverse of the covariance matrix of the empirical moments, to obtain the second stage estimator $\hat{\theta}_{T,2}$.

4.2 Estimating the severity component

Let us assume that we have i.i.d. loss severity count observations $W_1, ..., W_n$, following the GPIG distribution. Zhu and Joe (2009) suggest to estimate its parameters by maximum likelihood (ML) method, which requires to compute the pmf. Unfortunately, as we have mentioned in Section 3.2, its computational cost could be forbidding for extremely large arguments. Indeed, from the summary statistics in Table 1, we see that in our data, some of the loss severity exceed 1,000 million. Hence we resort to the GMM, which is computationally tractable, thanks to the simple pgf of the GPIG distribution.

Similar to the frequency model, let us denote by $\mathcal{U}^* = (u_1, ..., u_J)$ another set of arguments. Then we compute the difference between the empirical pgf and its theoretical counterpart:

$$g_j(\eta) = \frac{1}{n} \sum_{i=1}^{n} u_j^{W_i} - \exp\left\{ b[(1-c)^a - (1-cu_j)^a] \right\}, \qquad \forall\ j = 1, ..., J,$$

where $\eta = (a, b, c)'$ is the vector of parameters of the GPIG distribution. Then, our strategy will again be to "minimize" these differences, for each j. However, before announcing the GMM estimator, a further complication is the choice of the arguments in the above equation. Indeed, compared to the frequency component, the severity variable is more heavy-tailed. This means that if we take an arbitrary real argument $u \in [0, 1]$, the value of u^{W_i} may be either too large,

or too small, depending on its modulus. A simple solution to this numerical stability issue is to use complex numbers with unitary modulus such as $u_j = e^{\sqrt{-1}j\xi}, j = 1, .., J$, where ξ is a real number. In other words, instead of using the pgf of the severity distribution, we use the associated characteristic function. See Carrasco and Florens (2000) for an example of using characteristic function for model estimation.

Then the GMM estimator is obtained by minimizing the quadratic error:

$$\hat{\eta}_n = \arg\min_{\eta} \overline{g(\eta)}' W g(\eta),$$

where $\theta = (a, b, c)'$ denotes the model parameters, W is a symmetric positive definite weighting matrix, $g = (g_1, ..., g_J)'$ is a vector function of dimension J, and $\overline{g(\eta)}'$ is the complex conjugate. Similar to the frequency component, we use two-step GMM estimation procedure. Again, under standard regularity conditions, the GMM estimator is consistent, so long as the number of moment conditions J is larger than the number of parameters (i.e., 3).

5 Application

In this section, we apply our model proposed in Section 3 to the PRC database discussed in Section 2. We will first discuss the frequency model results in Section 5.1, and then move on to the severity model in Section 5.2. In this section, we will be mainly focused on events segmented by four breach types as defined in Panel A of Table 1. For completeness, in Appendix C we also report the estimation result when the events are alternatively segmented by the organization types, and we find consistent results.

5.1 The frequency model

Table 3 reports the estimation results of the frequency model. The parameters of the frequency model include, among others, a covariance matrix Σ. To ensure that its estimate is symmetric positive definite, we alternatively re-parameterize Σ using the Cholesky decomposition, $\Sigma = LL'$, where square matrix L is lower-triangular. Since this decomposition is unique, L is an equivalent re-parameterization of Σ, whose lower-triangular entries are no longer subject to domain restrictions.[27]

[27] Then the corresponding standard errors of the estimator $\hat{\Sigma}$ are calculated through the Delta Method.

Table 3: **GMM estimation results for loss frequency.** This table summarizes parameters of the negative binomial autoregressive process for the frequency component. Standard errors are collected in parentheses. $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. Σ is symmetric so only lower triangle is reported.

Para	Value	Para	Value	Para	Value
α_{DISC}	0.112^{***}	$\beta_{\text{INSD,HACK}}$	0.034^{***}	$\sigma_{\text{INSD,INSD}}$	0.076^{***}
	(0.031)		(0.005)		(0.021)
α_{HACK}	0.053^{***}	$\beta_{\text{INSD,INSD}}$	0.120^{***}	$\sigma_{\text{ELET,ELET}}$	0.910^{***}
	(0.008)		(0.021)		(0.030)
α_{INSD}	0.246^{***}	$\beta_{\text{INSD,ELET}}$	0.000	$\sigma_{\text{HACK,DISC}}$	0.154^{***}
	(0.027)		(0.002)		(0.027)
α_{ELET}	0.000	$\beta_{\text{ELET,DISC}}$	0.196	$\sigma_{\text{INSD,DISC}}$	-0.063^{***}
	(0.006)		(1.523)		(0.014)
$\beta_{\text{DISC,DISC}}$	0.564^{***}	$\beta_{\text{ELET,HACK}}$	0.133	$\sigma_{\text{INSD,HACK}}$	0.068^{**}
	(0.075)		(0.874)		(0.034)
$\beta_{\text{DISC,HACK}}$	0.141^{***}	$\beta_{\text{ELET,INSD}}$	0.227	$\sigma_{\text{ELET,DISC}}$	-0.001
	(0.025)		(1.991)		(0.023)
$\beta_{\text{DISC,INSD}}$	0.585^{***}	$\beta_{\text{ELET,ELET}}$	0.329	$\sigma_{\text{ELET,HACK}}$	0.148^{***}
	(0.069)		(2.027)		(0.021)
$\beta_{\text{DISC,ELET}}$	0.066^{***}	c_{DISC}	0.479^{***}	$\sigma_{\text{ELET,INSD}}$	0.073^{***}
	(0.025)		(0.057)		(0.016)
$\beta_{\text{HACK,DISC}}$	0.446^{***}	c_{HACK}	1.270^{***}	δ_{DISC}	0.029
	(0.025)		(0.048)		(0.263)
$\beta_{\text{HACK,HACK}}$	0.016	c_{INSD}	0.760^{***}	δ_{HACK}	0.001
	(0.012)		(0.080)		(0.117)
$\beta_{\text{HACK,INSD}}$	0.337^{***}	c_{ELET}	0.041	δ_{INSD}	0.182^{***}
	(0.050)		(0.255)		(0.060)
$\beta_{\text{HACK,ELET}}$	0.017	$\sigma_{\text{DISC,DISC}}$	0.285^{***}	δ_{ELET}	0.007
	(0.012)		(0.022)		(0.875)
$\beta_{\text{INSD,DISC}}$	0.000	$\sigma_{\text{HACK,HACK}}$	0.340^{***}	δ_0	3.047^{***}
	(0.010)		(0.043)		(0.142)

Let us first inspect the parameters that measure the ***pairwise*** contagion channel. First, the scale parameters c_i, $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$, vary a lot among different breach types, indicating different levels of pairwise risk. For example, $c_{\text{HACK}} = 1.27$ which is 31 times larger than $c_{\text{ELET}} = 0.041$. Second, the spillover effects between most breach types are strong. For example, the channel DISC $\rightarrow$ INSD reports the highest cross-sectional spillover effect, with $\beta_{\text{DISC,INSD}} = 0.585$. The spillover effects from HACK to DISC and INSD are also high, with $\beta_{\text{HACK,DISC}} = 0.446$ and $\beta_{\text{HACK,INSD}} = 0.337$, respectively. This is also true for the spillovers from ELET to DISC and INSD, with $\beta_{\text{ELET,DISC}} = 0.196$ and $\beta_{\text{ELET,INSD}} = 0.227$, respectively. By contrast, INSD does not have much spillover effects across sections. In

addition, the within section spillover effects are strong for breach types DISC and INSD, with $\beta_{\text{DISC,DISC}} = 0.564$ and $\beta_{\text{INSD,INSD}} = 0.120$, respectively, but quite weak for HACK and ELET. Note that for ELET, $\beta_{\text{ELET,ELET}} = 0.329$ is economically large but statistically insignificant. Finally, $\delta_{\text{INSD}} = 0.182$, and is statistically significant, indicating a large propensity of its idiosyncratic risk that cannot be interpreted by pairwise or systemic contagion channels.

Figure 3 illustrates the pairwise contagion channel. Each circle represents one of the four types of data breach, with its radius proportional to the scale parameter c_i, $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$. Lines between different circles represents cross-sectional spillover effects, with the line widths proportional to $\beta_{i,j}, i, j = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$. Dashed lines represent zero spillover effects.

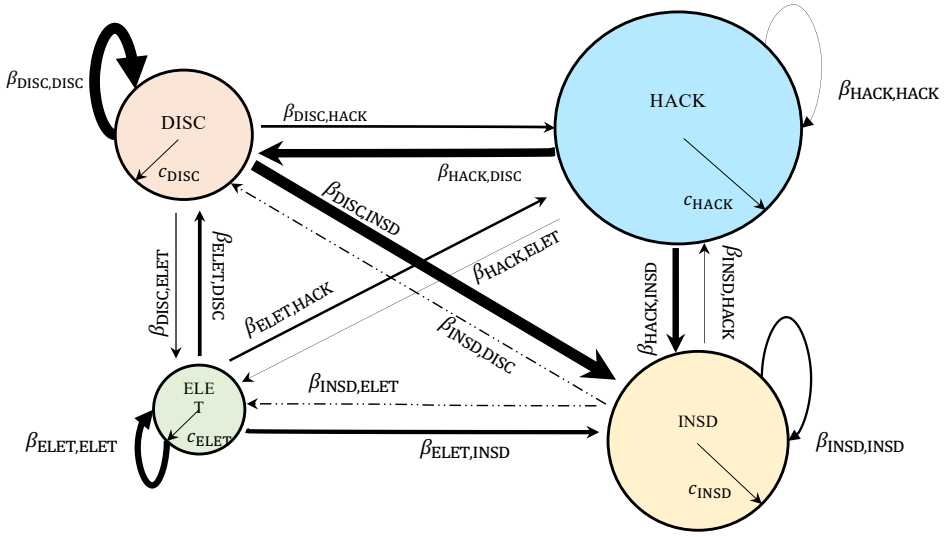


Figure 3: **Illustration of pairwise contagion channel.** Widths of arrow lines are proportional to $\beta_{i,j}, i, j = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$, that measure the spillover effects between each individual breach type. Radius of each circle is proportional to the scale parameter c_i, $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$, measuring idiosyncratic risk. Dashed lines represent zero spillover effects.

Next, we move on to inspect the ***systemic*** contagion channel. The contributions of DISC, HACK, and INSD to the systemic risk factor are large and statistically significant, with $\alpha_{\text{DISC}} = 0.112$, $\alpha_{\text{HACK}} = 0.053$, and $\alpha_{\text{INSD}} = 0.246$. ELET does not have impact on the systemic risk. The systemic risk has a strong impact on all of the four individual processes. Indeed, the diagonal terms of Σ are all positive and statistically significant. In particular, $\sigma_{\text{DISC}} = 0.285$, $\sigma_{\text{HACK}} = 0.340$, $\sigma_{\text{INSD}} = 0.076$, and $\sigma_{\text{ELET}} = 0.910$. We can see that the systemic risk factor

has extremely strong effect on breach type ELET, which is around 3 times of that of DISC or HACK, and around 12 times of that of INSD.

Figure 4 visualizes the systemic contagion channel. The circle in the center represents the systemic risk factor, and the four circles around it represent the four types of data breach. Arrows directing towards systemic risk factor represent the contribution of each breach type to the systemic risk factor, with their line widths proportional to the magnitudes of α_i, $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$. Arrows heading towards small circles stand for the impact of systemic risk on individual processes, with their widths proportional to σ_{ii}, $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$. To facilitate comparison with the pairwise contagion channel, we have used the same proportionality constant for the widths of the arrows in the two figures. Therefore, by comparing Figures 3 and 4, we see that both types of contagion are non-negligible.

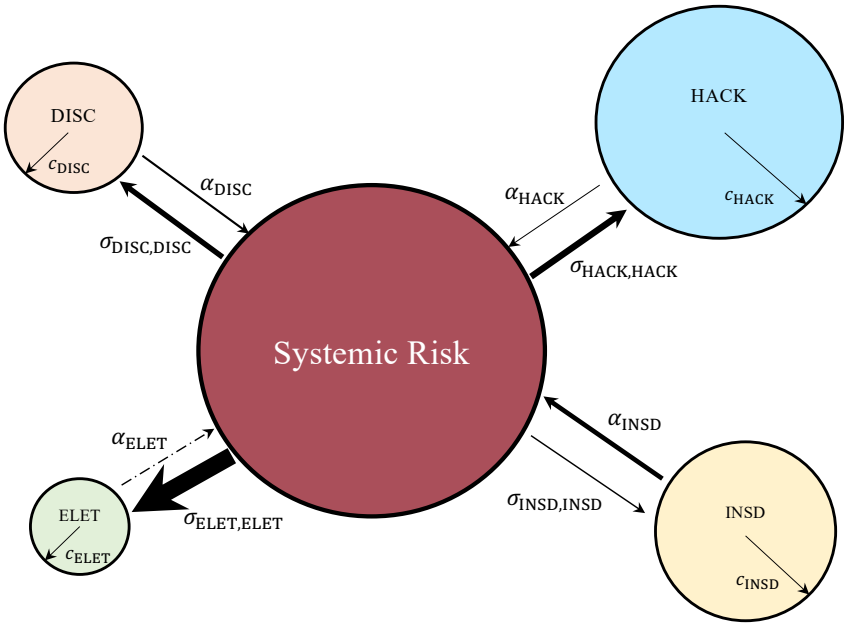


Figure 4: **Illustration of systemic contagion channel.** The circle in the center represents the systemic risk factor, and the circles around it represent idiosyncratic risks. Widths of arrow lines that are directing towards systemic (idiosyncratic) risks are proportional to the magnitudes of α_i (σ_{ii}), $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$.

In addition, from the estimate of Σ, and formula (2), we can also compute the conditional correlation matrix between different components of X_t, the latent factor that captures the systemic contagion channel. This latter is equal to:

$$\begin{array}{c} \begin{matrix} \text{DISC} & \text{HACK} & \text{INSD} & \text{ELET} \end{matrix} \\ \begin{pmatrix} 1 & 0.68 & 0.01 & 0 \\ * & 1 & 0.14 & 0.02 \\ * & * & 1 & 0.23 \\ * & * & * & 1 \end{pmatrix} \begin{matrix} \text{DISC} \\ \text{HACK} \\ \text{INSD} \\ \text{ELET} \end{matrix} \end{array},$$

indicating that different systemic factors are indeed quite correlated, especially between $X_{\text{DISC},t}$ and $X_{\text{HACK},t}$.

Finally, we can also compute matrix M defined in Equation (10):

$$M = (m_{i,j}) = \begin{array}{c} \begin{matrix} \text{DISC} & \text{HACK} & \text{INSD} & \text{ELET} \end{matrix} \\ \begin{pmatrix} 0.33 & 0.10 & 0.42 & 0.03 \\ 0.64 & 0.06 & 0.60 & 0.02 \\ 0.02 & 0.03 & 0.13 & 0.00 \\ 0.21 & 0.10 & 0.46 & 0.01 \end{pmatrix} \begin{matrix} \text{DISC} \\ \text{HACK} \\ \text{INSD} \\ \text{ELET} \end{matrix} \end{array}.$$

Recall that $m_{i,j}$ can be interpreted as the average number of contagious events of type j generated by a previous event of type i, and hence the contagion effect is quite significant between most pairs (i, j), $i, j = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$. Furthermore, the spectral radius is equal to 0.56, which is quite close to the first-order autocorrelation coefficients of the four individual count processes. In other words, our model captures the serial correlation quite well.

5.1.1 Comparing with other frequency models in the literature

To further evaluate the proposed NBAR model for cyber risk modeling, we compare its performance with other existing frequency models that have been widely used in cyber risk literature. We consider two main families of model comparison approaches, with the first based on likelihood function and information criteria, and the second based on the mean square error (MSE).

Likelihood and information criteria based model comparison This approach is based on computing the closed-form likelihood function of the four-dimensional NBAR process. More pre-

cisely, the log-likelihood function is:[28]

$$\ell = \sum_{t=2}^{T} \log \ell(y_t|y_{t-1}),$$

where the conditional density $\ell(y_t|y_{t-1})$ is:

$$\begin{aligned}\ell(y_t|y_{t-1}) = \sum_{i=0}^{y_{1t}} \sum_{j=0}^{y_{2t}} \sum_{k=0}^{y_{3t}} \sum_{h=0}^{y_{4t}} & \mathbb{E}\left[\frac{e^{-Z_{1t}} Z_{1t}^{y_{1t}-i}}{(y_{1t}-i)!}\middle| y_{t-1}\right] \mathbb{E}\left[\frac{e^{-Z_{2t}} Z_{2t}^{y_{2t}-j}}{(y_{2t}-j)!}\middle| y_{t-1}\right] \\ & \cdot \mathbb{E}\left[\frac{e^{-Z_{3t}} Z_{3t}^{y_{3t}-k}}{(y_{3t}-k)!}\middle| y_{t-1}\right] \mathbb{E}\left[\frac{e^{-Z_{4t}} Z_{4t}^{y_{4t}-h}}{(y_{4t}-h)!}\middle| y_{t-1}\right] \\ & \cdot \mathbb{E}\left[\frac{X_{1t}^i X_{2t}^j X_{3t}^k X_{4t}^h e^{-X_{1t}-X_{2t}-X_{3t}-X_{4t}}}{i!j!k!h!}\middle| y_{t-1}\right].\end{aligned}$$

Because for each $j = 1, \cdots, 4$, the intensity Z_{jt} has conditional gamma distribution given Y_{t-1}, we have:

$$\mathbb{E}[e^{-Z_{jt}} Z_{jt}^{y_{jt}-i}] = \frac{c_j^{(y_{jt}-i)}}{(1+c_j)^{(\delta_j+\beta_j' y_{t-1}+y_{jt}-i)}} \frac{\Gamma(\delta_j+\beta_j' y_{t-1}+y_{jt}-i)}{\Gamma(\delta_j+\beta_j' y_{t-1})}.$$

Finally, because $(X_{1t}, X_{2t}, X_{3t}, X_{4t})$ follows Wishart-gamma distribution given Y_{t-1}, we have:

$$\begin{aligned}& \mathbb{E}[X_{1t}^i X_{2t}^j X_{3t}^k X_{4t}^h e^{-X_{1t}-X_{2t}-X_{3t}-X_{4t}}|y_{t-1}] \\ &= (-1)^{i+j+k+l} \frac{\partial^{i+j+k+h}}{\partial s_1^i \partial s_2^j \partial s_3^k \partial s_4^h} \left[\frac{1}{\left|\mathrm{Id}_4 + 2\mathrm{Diag}(s_1,s_2,s_3,s_4)\Sigma\right|^{\delta_0/2+\alpha' y_{t-1}}}\right]\Bigg|_{s_1=s_2=s_3=s_4=1}.\end{aligned}$$

Finally, based on the log-likelihood estimated, information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be derived. The information criteria include a penalty term for the model parsimony, and therefore are more suitable for model comparison and selection. Formally, the AIC and BIC are defined respec-

[28] Note, that even though the log-likelihood function has a closed form, in the estimation section, we did not use this result for estimation, but resort instead to GMM. This is due to the fact that the GMM is much more computationally feasible for the estimation of our multivariate NBAR process.

tively as:

$$\text{AIC} = -2\log\left(\hat{\ell}\right) + 2p,$$
$$\text{BIC} = -2\log\left(\hat{\ell}\right) + p \cdot \log(N),$$

where $\hat{\ell}$ is the estimated value of the log-likelihood function, p is the number of parameters in the model, and N is the total sample size. Models with lower values of AIC and BIC are preferred.

MSE based model comparison One can also compare directly the quality of the fit, as well as that of the prediction of a model by calculating the MSE. For the multivariate NBAR model proposed in this paper, the in-sample and out-of-sample MSEs are easily computed according to Equation (5).

Model comparison results Let us now conduct a horse race between the proposed multivariate NBAR model and the negative binomial model. The later model has been considered by various research (see, e.g., Eling and Jung, 2018; Edwards et al., 2016; Eling and Loperfido, 2017), and has been confirmed superior in cyber risk frequency modeling. Table 4 reports the model comparison results. Panel A presents the estimated likelihood, AIC, and BIC, which measure the model goodness-of-fit. Panel B and Panel C report the in-sample and out-of-sample MSE, respectively. We can see that based on various criteria, our proposed multivariate NBAR model performs better than the negative binomial benchmark.

5.2 The severity model

Table 5 reports the parameter estimates of the severity models, for each of the four event types. We see that for all breach types, the estimated parameter c goes to 1. This indicates that our severity loss data belong to the limiting case that $c \to 1$, and $\mathbb{E}[W]$ goes to infinity. This corresponds to the discrete stable distribution. The estimation results confirm the heavy-tail phenomenon highlighted in Section 2.

Table 4: **Model comparison.** This table compares in-sample and out-of-sample performance of NBAR model with the classic negative binomial model.

Panel A: Goodness-of-fit					
	LL	AIC	BIC		
NB	-6194.95	12405.90	12454.18		
NBAR	-5941.18	11960.36	12195.74		
Panel B: In-sample prediction MSE					
	DISC	HACK	INSD	ELET	Overall
NB	4.55	11.12	1.46	12.27	7.35
NBAR	3.93	9.04	1.37	10.90	6.31
Panel C: Out-of-sample prediciton MSE					
	DISC	HACK	INSD	ELET	Overall
NB	4.24	8.38	0.69	16.44	7.44
NBAR	1.76	1.80	0.53	10.12	3.55

Table 5: **GMM estimation results for loss severity.** This table summarizes parameters of the GPIG distribution for the severity component. Each column reports the parameters specific to one of the four event types. Standard errors are collected in parentheses. $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

	DISC	HACK	INSD	ELET
a	0.114	0.279^{***}	0.290^{***}	0.108
	(0.413)	(0.018)	(0.060)	(0.323)
b	2.402^{***}	4.810^{***}	3.139^{**}	3.448^{***}
	(0.618)	(0.694)	(1.340)	(0.120)
c	1.000^{***}	1.000^{***}	1.000^{***}	1.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)

5.3 Combining frequency and severity models into a collective risk model

In this section, we apply our proposed multivariate count-based frequency-severity framework to cyber risk insurance pricing and risk measurement. For each index $i = \text{DISC}, \text{HACK}, \text{INSD}, \text{ELET}$, we are interested in S_{it}, the aggregate loss due to data breaches of type i at time t, given by:

$$S_{it} = \sum_{j=1}^{Y_{it}} L_{ji}, \tag{22}$$

where Y_{it} is the frequency component that follows a multivariate NBAR process and L_{ji} is the monetized severity component.

To convert the number of breach records to dollar losses, we follow the relationship proposed

by Jacobs (2014),

$$L_{ji} = \exp\{7.68 + 0.76\log(W_{ji})\}, \tag{23}$$

where W_{ji} is the number of breach records of the j th event, where $j = 1, 2, \ldots, Y_{it}$. Moreover, these W_{ji}, j varying, are i.i.d. following the GPIG distribution associated to the $i-$th event type. Furthermore, we assume that the frequency and severity components are independent in eq. (22).

In this collective risk model, we are mostly interested in the conditional expectation of the economic losses:

$$\mathbb{E}[S_{it}|\underline{Y_{t-1}}] = \mathbb{E}[Y_{i,t}|\underline{Y_{t-1}}]e^{7.68}\mathbb{E}\left[W_i^{0.76}\right], \tag{24}$$

where the frequency prediction $\mathbb{E}[Y_{i,t}|\underline{Y_{t-1}}]$ can be obtained from eq. (5).

Thus it suffices to compute $\mathbb{E}[W^{0.76}]$ for a variable W following a GPIG distribution. This can be done using two approaches. First, we could follow Section 2.1 of Zhu and Joe (2009) to draw from the GPIG distribution. This simulation method, however, is quite complicated. Moreover, given that the distribution of W is likely heavy-tailed, a plain Monte Carlo simulation with a reasonable number of draws might lead to a large variance. An alternative method is to remark that, just as the integer moments of W can be obtained by taking successive derivatives of its Laplace transform, its fractional moment involves the fractional derivative of the Laplace transform formula for W. More precisely, we have:

Proposition 3. For any exponent β between 0 and 1, we have:

$$\mathbb{E}\left[W^{1-\beta}\right] = \frac{1}{\Gamma(\beta)}\int_0^\infty s^{\beta-1}\mathbb{E}\left[e^{-sW}W\right]\mathrm{d}s, \tag{25}$$

where $\mathbb{E}\left[e^{-sW}W\right]$ is minus of the derivative of the Laplace transform $\mathbb{E}\left[e^{-sW}\right]$.

Proof. See Appendix B. □

In the case of GPIG distribution, its Laplace transform is given by eq. (12):

$$\mathbb{E}\left[e^{-sW}\right] = \exp\left\{b\left[(1-c)^a - (1-ce^{-s})^a\right]\right\}, \qquad s > 0,$$

thus

$$\mathbb{E}\left[e^{-sW}W\right] = abc(1-ce^{-s})^{a-1}\exp\left\{b\left[(1-c)^a-(1-ce^{-s})^a\right]-s\right\}, \tag{26}$$

and

$$\mathbb{E}\left[W^{1-\beta}\right] = \frac{1}{\Gamma(\beta)}abce^{b(1-c)^a}\int_0^\infty s^{\beta-1}(1-ce^{-s})^{a-1}e^{-b(1-ce^{-s})^a-s}\mathrm{d}s. \tag{27}$$

By taking $\beta = 0.24$, we get the expression of $\mathbb{E}[W^{0.76}]$. This integral can be easily approximated using routine numerical methods such as Riemann summation.

The same method can be used to compute the conditional variance/covariance of the losses due to various types. For instance, using the total variance formula, we have:

$$\begin{aligned}\mathbb{V}[S_{it}|\underline{Y_{t-1}}] &= \mathbb{V}\big[\mathbb{E}[S_{i,t}|Y_{i,t}]|\underline{Y_{t-1}}\big] + \mathbb{E}\big[\mathbb{V}[S_{i,t}|Y_{i,t}]|\underline{Y_{t-1}}\big] \\ &= e^{7.68}\mathbb{E}[W_i^{0.76}]\mathbb{V}\big[Y_{i,t}|\underline{Y_{t-1}}\big] + e^{2(7.68)}\mathbb{V}[W_i^{0.76}]\mathbb{E}\big[Y_{i,t}|\underline{Y_{t-1}}\big],\end{aligned}$$

where the conditional variance $\mathbb{V}\big[Y_{i,t}|\underline{Y_{t-1}}\big]$ has been derived in eq.(7). Thus it suffices to compute $\mathbb{V}[W_i^{0.76}]$, or equivalently $\mathbb{E}[W_i^{2(0.76)}]$. Similar to (25), it is straightforward to show that:[29]

$$\mathbb{E}\left[W^{2-\beta}\right] = \frac{1}{\Gamma(\beta)}\int_0^\infty s^{\beta-1}\mathbb{E}\left[e^{-sW}W^2\right]\mathrm{d}s,$$

where $\mathbb{E}\left[e^{-sW}W^2\right]$ can be obtained by taking derivative in eq.(26) with respect to s:

$$\begin{aligned}&\mathbb{E}\left[e^{-sW}W^2\right] \\ =&abc(1-ce^{-s})^{a-2}\left[abc(1-ce^{-s})^a e^{-s}+1-ace^{-s}\right]\cdot\exp\left\{b\left[(1-c)^a-(1-ce^{-s})^a\right]-s\right\},\end{aligned}$$

and hence,

$$\begin{aligned}&\mathbb{E}\left[W^{2-\beta}\right] \\ =&\frac{1}{\Gamma(\beta)}abce^{b(1-c)^a}\int_0^\infty s^{\beta-1}(1-ce^{-s})^{a-2}[abc(1-ce^{-s})^a e^{-s}+1-ace^{-s}]e^{-b(1-ce^{-s})^a-s}\mathrm{d}s.\end{aligned} \tag{28}$$

Finally, by taking $\beta = 2-2(0.76) = 0.48$, we get the value of $\mathbb{E}\left[W_i^{2(0.76)}\right]$.

[29] Indeed, it suffices to adapt the proof of Proposition 3, by taking second order derivative in eq. (B.1) instead of first order derivative.

Increasing the forecasting horizon The above calculation is based on a one-week-ahead prediction. In practice, this horizon might be too short, especially for insurance purpose. Let us for instance consider a h-week prediction horizon. Then we simply get:

$$\mathbb{E}[S_{i,t+h}|\underline{Y_t}] = \mathbb{E}[Y_{i,t+h}|\underline{Y_t}]e^{7.68}\mathbb{E}[W_i^{0.76}],$$

where the first term can be obtained from (9), and the second term is the same as in eq. (24). The conditional variance also has a recursive formula. Indeed, for $j, k = 1, ..., K$, we have

$$\begin{aligned}
&\mathbb{C}ov[Y_{j,t+2}, Y_{k,t+2}|Y_t] \\
=&\mathbb{C}ov\Big[\mathbb{E}[Y_{j,t+2}|X_{j,t+2}, Z_{j,t+2}], \mathbb{E}[Y_{k,t+2}|X_{j,t+2}, Z_{j,t+2}]|Y_t\Big] + \mathbb{E}\big[\mathbb{C}ov[Y_{j,t+2}, Y_{k,t+2}|X_{j,t+2}, Z_{j,t+2}]|Y_t\Big] \\
=&\mathbb{C}ov\Big[X_{j,t+2} + Z_{j,t+2}, X_{k,t+2} + Z_{k,t+2}|Y_t\Big] + \mathbb{1}_{j=k}\mathbb{E}\Big[X_{j,t+2} + Z_{j,t+2}|Y_t\Big] \\
=&\mathbb{C}ov\Big[\mathbb{E}[X_{j,t+2} + Z_{j,t+2}|Y_{t+1}], \mathbb{E}[X_{k,t+2} + Z_{k,t+2}|Y_{t+1}]|Y_t\Big] \\
&+ \mathbb{E}\Big[\mathbb{C}ov[X_{j,t+2} + Z_{j,t+2}, X_{k,t+2} + Z_{k,t+2}|Y_{t+1}]|Y_t\Big] + \mathbb{1}_{j=k}\mathbb{E}\Big[\mathbb{E}[X_{j,t+2} + Z_{j,t+2}|Y_{t+1}]|Y_t\Big].
\end{aligned} \tag{29}$$

The first term in (29) is equal to:

$$\begin{aligned}
&\mathbb{C}ov\left[c_j(\delta_j + \beta_j'Y_{t+1}) + 2\sigma_{jj}\left(\frac{\delta_0}{2} + \alpha'Y_{t+1}\right), c_k(\delta_k + \beta_k'Y_{t+1}) + 2\sigma_{kk}\left(\frac{\delta_0}{2} + \alpha'Y_{t+1}\right)|Y_t\right] \\
=&\mathbb{C}ov\left[c_j\beta_j'Y_{t+1} + 2\sigma_{jj}\alpha'Y_{t+1}, c_k\beta_k'Y_{t+1} + 2\sigma_{kk}\alpha'Y_{t+1}|Y_t\right] \\
=&(c_j\beta_j + 2\sigma_{jj}\alpha)'\mathbb{C}ov[Y_{t+1}|Y_t](c_k\beta_k + 2\sigma_{kk}\alpha),
\end{aligned}$$

where $\mathbb{C}ov[Y_{t+1}|Y_t]$ is the covariance matrix of Y_{t+1} given Y_t and is given in eq.(7). The second term in (29) is equal to:

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{C}ov[X_{j,t+2}, X_{k,t+2}|Y_{t+1}]|Y_t\Big] + \mathbb{1}_{j=k}\mathbb{E}[\mathbb{V}[Z_{k,t+2}|Y_{t+1}]|Y_t] \\
=&\mathbb{E}\Big[4\sigma_{ij}^2(\frac{\delta_0}{2} + \alpha'Y_{t-1})|Y_t\Big] + \mathbb{1}_{j=k}\mathbb{E}\Big[c_j^2(\delta_j + \beta_j'Y_{t-1})|Y_t\Big] \\
=&2\sigma_{ij}^2\delta_0 + 4\sigma_{ij}^2\alpha'\mathbb{E}[Y_{t+1}|Y_t] + \mathbb{1}_{j=k}\left(c_j^2\delta_j + c_j^2\beta_j'\mathbb{E}[Y_{t+1}|Y_t]\right),
\end{aligned}$$

where $\mathbb{E}[Y_{t+1}|Y_t]$ is the vector of conditional expectation given by eq.(6). As for the third term in (29), it is equal to:

$$\mathbb{1}_{j=k}\mathbb{E}\Big[c_j(\delta_j+\beta_j'Y_{t+1})+2\sigma_{jj}(\frac{\delta_0}{2}+\alpha'Y_{t+1})|Y_t\Big]=\mathbb{1}_{j=k}\Big[c_j\delta_j+\sigma_{jj}\delta_0+(c_j\beta_j'+2\sigma_{jj}\alpha')\mathbb{E}[Y_{t+1}|Y_t]\Big].$$

Combining the results above, we get the conditional covariance matrix at order 2:

$$\begin{aligned}
&\mathbb{C}ov[Y_{j,t+2},Y_{k,t+2}|Y_t]\\
=&(c_j\beta_j+2\sigma_{jj}\alpha)'\mathbb{C}ov[Y_{t+1}|Y_t](c_k\beta_k+2\sigma_{kk}\alpha)+2\sigma_{ij}^2\delta_0+4\sigma_{ij}^2\alpha'\mathbb{E}[Y_{t+1}|Y_t]\\
&+\mathbb{1}_{j=k}\left(c_j^2\delta_j+c_j^2\beta_j'\mathbb{E}[Y_{t+1}|Y_t]\right)+\mathbb{1}_{j=k}\Big[c_j\delta_j+\sigma_{jj}\delta_0+(c_j\beta_j'+2\sigma_{jj}\alpha')\mathbb{E}[Y_{t+1}|Y_t]\Big],
\end{aligned}$$

which is linear in $\mathbb{E}[Y_{t+1}|Y_t]$ and $\mathbb{C}ov[Y_{t+1}|Y_t]$ at order 1.

Lastly, by the same argument, at higher orders $h\geq 3$ we have:

$$\begin{aligned}
&\mathbb{C}ov[Y_{j,t+h},Y_{k,t+h}|Y_t]\\
=&(c_j\beta_j+2\sigma_{jj}\alpha)'\mathbb{C}ov[Y_{t+h-1}|Y_t](c_k\beta_k+2\sigma_{kk}\alpha)+2\sigma_{ij}^2\delta_0+4\sigma_{ij}^2\alpha'\mathbb{E}[Y_{t+h-1}|Y_t] \qquad (30)\\
&+\mathbb{1}_{j=k}\left(c_j^2\delta_j+c_j^2\beta_j'\mathbb{E}[Y_{t+h-1}|Y_t]\right)+\mathbb{1}_{j=k}\Big[c_j\delta_j+\sigma_{jj}\delta_0+(c_j\beta_j'+2\sigma_{jj}\alpha')\mathbb{E}[Y_{t+h-1}|Y_t]\Big], \qquad (31)
\end{aligned}$$

which is linear in the conditional expectation $\mathbb{E}[Y_{t+h-1}|Y_t]$ and covariance $\mathbb{C}ov[Y_{t+h-1}|Y_t]$ at order $h-1$. Thus, the conditional variance/covariance at order h can be easily obtained through the linear recursion (31).

5.4 Numerical application

In this subsection, we show the numerical results. We consider two premium principles, including expectation premium principle ($\pi=(1+\theta)\mathbb{E}(S_{it})$) and standard deviation premium principle ($\pi=\mathbb{E}(S_{it})+\theta\sqrt{\mathbb{V}(S_{it})}$), where S_{it} is the aggregate loss due to breach type i at time t, and θ is the loading parameter. Table 6 summarizes the h-week-ahead ($1\leq h\leq 4$) cyber insurance pricing with $\theta=20\%$. Panel A displays results for expectation premium principle and Panel B displays results for standard deviation premium principle. As explained above, these premiums can be computed using Proposition 3. For the benchmark model, we consider classic aggregate models with discrete frequency and continuous severity. It has been shown in the literature that negative binomial model fits the best to the frequency component of the

PRC data (see, e.g, Edwards et al., 2016; Eling and Loperfido, 2017), and lognormal fits the best to the PRC severity data (see, e.g., Edwards et al., 2016; Eling and Loperfido, 2017; Eling and Jung, 2018). Therefore, we use them as the benchmark.

From Table 6, we have several interesting observations. First, it is immediately shown that our proposed framework could achieve a dynamic prediction, while the classic collective risk model that has been widely discussed in the literature can only perform a static prediction. Second, comparing across different breach types, we see that ELET has the highest premiums than the other categories (based on both premium principles), and this is consistent with both our proposed framework and the benchmark model. Third, our proposed multivariate count-based framework can capture the heavy tail property much better than the benchmark model, with all standard deviation premiums larger than those of the benchmark model for all breach types.

Table 6: **Cyber risk insurance pricing results.** This table summarizes pricing results for h-week-ahead cyber insurance ($1 \leq h \leq 4$). Results for count process are based on our proposed multivariate count-based frequency-severity framework, and results for the benchmark are based on classic aggregate models with discrete frequency and continuous severity. All results are in million dollars.

			$h=1$	$h=2$	$h=3$	$h=4$
Expectation Premium	Count process	DISC	49.02	65.40	71.89	74.72
		HACK	39.84	50.61	58.31	60.60
		INSD	9.19	10.87	11.40	11.66
		ELET	91.90	92.76	93.85	94.41
	Benchmark	DISC		11.08		
		HACK		10.83		
		INSD		2.45		
		ELET		17.87		
Standard Deviation Premium	Count process	DISC	64.19	81.47	88.18	91.09
		HACK	60.38	72.81	81.47	84.01
		INSD	20.70	23.25	24.04	24.41
		ELET	124.74	125.68	126.87	127.49
	Benchmark	DISC		18.08		
		HACK		16.15		
		INSD		4.35		
		ELET		23.92		

6 Concluding remarks

In this paper, we propose a new frequency-severity framework for the analysis of cyber risk. For

the frequency model, we have introduced a multivariate count time series called the multivariate NBAR process. For the severity model, we have proposed the GPIG distribution, as well as its special case, the discrete stable distribution. Both models are new to the insurance literature, and we have illustrated their usefulness for cyber risk. These models are compatible with stylized facts observed on the cyber risk data, and are also quite flexible and tractable. In particular, the frequency model we propose in this paper pioneers the introduction of two separate mechanisms of contagion, including the ***pairwise*** contagion channel and the ***systemic*** contagion channel. Our results also highlight the extremely heavy tail property of the loss severity based on the estimation results of the GPIG distribution.

Given the scarcity of proprietary databases, our analysis is based on the public PRC data. However, our methodology remains highly useful to other commercial databases, including those which provide the economic loss. Indeed, the frequency part of our model, as well as the associated GMM-based estimation method, can be directly applied to any other alternative databases. As for the severity part of our mode, it has been reported by Eling and Wirfs (2019) that the (continuously-valued) severity distribution fitted using SAS OpRisk database is also of infinite mean. A natural extension to our model for that dataset is the (one-sided) α-stable distribution, which has a positive support and allows to capture infinite mean. The α-stable distribution has a tractable Laplace transform. In collective risk models, severity distributions with closed-form Laplace transforms are particularly interesting, since so long as the frequency model also has simple Laplace transform (which is the case under our frequency model), the Laplace transform of the resulting aggregate loss variable inherits tractability. This avoids the needs of computing numerically the convolution of other competing continuous distributions such as log-normal or Pareto (Nguyen and Robert, 2015; Furman et al., 2020). Because of its simple Laplace transform, the parameter estimation of the α-stable distribution could most naturally be conducted using GMM. As a consequence, this paper has also laid the econometric foundation for such an empirical analysis, when competing databases become more easily accessible in the future.

Finally, even though the models newly introduced in this paper have been shown to be very suitable for cyber risk, this latter is by no means the only possible application. Rather, as more and more "high frequency" data become available in insurance and many of them share similar complicated dynamic dependence structure as the cyber risk, the same modeling

approach could become potentially useful for many other emerging risks. For example, because of the climate change, there is a rising need to analyze the dependence between the occurrences of different climate related hazards, or between the frequencies of the same natural hazard at different regions. These tasks are essential, since such risk correlation might have an adverse impact on insurers' traditional diversification plan. However, the literature on discrete-time hazard frequency modeling has so far largely focused on the univariate or bivariate case only (see, e.g. Boudreault and Charpentier, 2011; Livsey et al., 2018) due to the curse of dimensionality of higher-dimensional models. The model introduced in our paper constitutes a serious competitor thanks to its nice trade-off between flexibility and tractability.

A Proof of Proposition 2

Inequality (18) is equivalent to:

$$(1-c)^a - 1 > -ac(1-c)^{a-1} \tag{A.1}$$

Let us define function $f(c) = (1-c)^a - 1$, where $c \in (0,1)$. We have:

$$f(0) = 0, \quad f'(c) = -a(1-c)^{a-1}, \quad f''(c) = a(a-1)(1-c)^{a-2} \leq 0.$$

Thus $f'(c)$ is decreasing. Then by Taylor's theorem, we have

$$\frac{(1-c)^a - 1}{c} = \frac{f(c) - f(0)}{c} > f'(c) = -a(1-c)^{a-1}.$$

Hence inequality (A.1). For inequality (19), it suffices to remark that:

$$\frac{\mathbb{P}[X=1]}{\mathbb{P}[X=0]} = abc < abc(1-c)^{a-1} = \frac{p_1^*}{p_0^*}. \tag{A.2}$$

B Proof of Proposition 3

By the definition of the Gamma function, we have:

$$w^{-\beta} = \frac{1}{\Gamma(\beta)} \int_0^\infty s^{\beta-1} e^{-sw} \mathrm{d}s, \qquad \forall x > 0.$$

Multiplying both sides by e^{-tw}, replacing the constant w by the random variable W, and taking expectation yields:

$$\mathbb{E}[W^{-\beta}\exp(-tW)] = \frac{1}{\Gamma(\beta)}\int_0^\infty s^{\beta-1}\mathbb{E}[e^{-(t+s)W}]\mathrm{d}s,$$

where we can interchange the order of the expectation and integration due to the Fubini's theorem for positive functions, so long as the right hand side is finite.

Taking derivative with respect to t on both sides yields:

$$\mathbb{E}[W^{1-\beta}\exp(-tW)] = \frac{1}{\Gamma(\beta)}\int_0^\infty s^{\beta-1}\mathbb{E}[e^{-(t+s)W}W]\mathrm{d}s, \tag{B.1}$$

Finally, we set t to zero in the above equation and obtain Equation (25).

C Estimation results for events classified by organization types

This section reports estimation results, when the events are segmented not by event type, but by organizations that report the event. Such a categorization follows the definition of Privacy Rights Clearinghouse (PRC), and has been considered in Eling and Jung (2018). In particular, "EDU" stands for educational institutions; "BSF" stands for financial and insurance services; "BSE" stands for non-financial business that combines retail/merchant and other business parties; "GOV" stands for government & military; "MED" stands for healthcare, medical providers and medical insurance services.

The main conclusions we can draw from the estimation results are similar to those in Section 5. For instance, in Table C.1, we see that most of the coefficients of matrix B, vectors c, α, as well as covariance matrix Σ are positive, indicating that both transmission channels are statistically significant. Finally, in Table C.2, we see that the parameter c is again estimated to be indistinguishable from 1, implying that the number of recorded breaches is extremely heavy tailed.

Table C.1: **GMM estimation results for loss frequency.** This table summarizes parameters of the negative binomial autoregressive process for the frequency component. Cyber data is classified according to organization types. Standard errors are collected in parentheses. $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. Σ is symmetric so only lower triangle is reported.

Para	Value	Para	Value	Para	Value	Para	Value
α_{EDU}	0.134^{*}	$\beta_{\text{BSE,GOV}}$	0.235	$\beta_{\text{GOV,MED}}$	0.199^{***}	$\sigma_{\text{BSF,BSE}}$	0.052
	(0.074)		(0.168)		(0.036)		(0.048)
α_{BSE}	0.312^{***}	$\beta_{\text{BSF,EDU}}$	0.221^{***}	$\beta_{\text{GOV,GOV}}$	0.263^{*}	$\sigma_{\text{MED,EDU}}$	0.000
	(0.059)		(0.074)		(0.164)		(0.009)
α_{BSF}	0.147^{*}	$\beta_{\text{BSF,BSE}}$	0.125^{*}	c_{EDU}	0.581^{***}	$\sigma_{\text{MED,BSE}}$	0.079^{***}
	(0.086)		(0.073)		(0.015)		(0.014)
α_{MED}	0.138^{***}	$\beta_{\text{BSF,BSF}}$	0.18^{***}	c_{BSE}	0.327^{***}	$\sigma_{\text{MED,BSF}}$	0.034^{***}
	(0.019)		(0.062)		(0.034)		(0.011)
α_{GOV}	0.164^{*}	$\beta_{\text{BSF,MED}}$	0.125^{***}	c_{BSF}	0.354^{***}	$\sigma_{\text{GOV,EDU}}$	0.000
	(0.091)		(0.040)		(0.060)		(0.011)
$\beta_{\text{EDU,EDU}}$	0.437^{***}	$\beta_{\text{BSF,GOV}}$	0.203^{**}	c_{MED}	0.011	$\sigma_{\text{GOV,BSE}}$	0.019
	(0.044)		(0.095)		(0.273)		(0.015)
$\beta_{\text{EDU,BSE}}$	0.041^{*}	$\beta_{\text{MED,EDU}}$	0.220	c_{GOV}	0.250^{***}	$\sigma_{\text{GOV,BSF}}$	0.006
	(0.026)		(9.333)		(0.053)		(0.013)
$\beta_{\text{EDU,BSF}}$	0.287^{***}	$\beta_{\text{MED,BSE}}$	0.199	$\sigma_{\text{EDU,EDU}}$	0.000	$\sigma_{\text{GOV,MED}}$	-0.094^{***}
	(0.073)		(8.121)		(0.023)		(0.013)
$\beta_{\text{EDU,MED}}$	0.144^{***}	$\beta_{\text{MED,BSF}}$	0.235	$\sigma_{\text{BSE,BSE}}$	0.146^{***}	δ_{EDU}	0.273^{***}
	(0.010)		(8.850)		(0.013)		(0.034)
$\beta_{\text{EDU,GOV}}$	0.122	$\beta_{\text{MED,MED}}$	0.129	$\sigma_{\text{BSE,BSF}}$	0.045	δ_{BSE}	0.000
	(0.080)		(3.465)		(0.049)		(0.344)
$\beta_{\text{BSE,EDU}}$	0.175^{**}	$\beta_{\text{MED,GOV}}$	0.227	$\sigma_{\text{MED,MED}}$	0.696^{***}	δ_{BSF}	0.068
	(0.072)		(12.746)		(0.045)		(0.274)
$\beta_{\text{BSE,BSE}}$	0.222^{***}	$\beta_{\text{GOV,EDU}}$	0.216^{**}	$\sigma_{\text{GOV,GOV}}$	0.088^{**}	δ_{MED}	0.000
	(0.067)		(0.086)		(0.041)		(17.619)
$\beta_{\text{BSE,BSF}}$	0.239^{***}	$\beta_{\text{GOV,BSE}}$	0.215^{***}	$\sigma_{\text{BSE,EDU}}$	0.000	δ_{GOV}	0.015
	(0.070)		(0.062)		(0.002)		(0.502)
$\beta_{\text{BSE,MED}}$	0.235^{***}	$\beta_{\text{GOV,BSF}}$	0.237^{***}	$\sigma_{\text{BSF,EDU}}$	0.000	δ_0	4.01^{***}
	(0.031)		(0.090)		(0.038)		(0.433)

Table C.2: **GMM estimation results for loss severity.** This table summarizes parameters of the GPIG distribution for the severity component. Cyber data is classified according to organization types. Standard errors are collected in parentheses. $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

	EDU	BSE	BSF	MED	GOV
a	0.297^{***}	0.010^{***}	0.124^{**}	0.245	0.304^{***}
	(0.057)	(0.002)	(0.062)	(0.378)	(0.033)
b	8.594^{***}	0.621^{***}	1.389^{***}	7.703	6.511^{***}
	(3.136)	(0.106)	(0.245)	(14.755)	(1.645)
c	1.000^{***}	1.000^{***}	1.000^{***}	1.000^{***}	1.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

References

Aït-Sahalia, Y., Laeven, R. J., and Pelizzon, L. (2014). Mutual excitation in Eurozone sovereign CDS. *Journal of Econometrics*, 183(2):151–167.

AM Best (2019). Cyber insurance: Profitability less certain as new risks emerge. Technical report, Best's Market Segment Report.

Bailey, T., Del Miglio, A., and Richter, W. (2014). The rising strategic risks of cyberattacks. Technical report, McKinsey & Company.

Bessy-Roland, Y., Boumezoued, A., and Hillairet, C. (2021). Multivariate Hawkes process for cyber insurance. *Annals of Actuarial Science*, 15(1):14–39.

Biener, C., Eling, M., and Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. *Geneva Papers on Risk and Insurance-Issues and Practice*, 40(1):131–158.

Böhme, R. and Kataria, G. (2006). Models and measures for correlation in cyber-insurance. In *WEIS*, volume 2, page 3.

Boudreault, M. and Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv preprint arXiv:1112.0929*.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75(2):305–332.

Carrasco, M. and Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16(6):797–834.

Catania, L. and Di Mari, R. (2021). Hierarchical Markov-switching models for multivariate integer-valued time-series. *Journal of Econometrics*, 221(1):118–137.

Cebula, J. L. and Young, L. R. (2010). A taxonomy of operational cyber security risks. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.

Chavez-Demoulin, V., Davison, A. C., and McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234.

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776.

Christoph, G. and Schreiber, K. (1998). Discrete stable random variables. *Statistics & Probability Letters*, 37(3):243–247.

Cummins, J. D., Lewis, C. M., and Wei, R. (2006). The market value impact of operational loss events for US banks and insurers. *Journal of Banking & Finance*, 30(10):2605–2634.

Darolles, S., Gourieroux, C., and Jasiak, J. (2006). Structural Laplace transform and compound autoregressive models. *Journal of Time Series Analysis*, 27(4):477–503.

Darolles, S., Le Fol, G., Lu, Y., and Sun, R. (2019). Bivariate integer-autoregressive process with an application to mutual fund flows. *Journal of Multivariate Analysis*, 173:181–203.

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V., and Ravishanker, N. (2021). Count time series: A methodological review. *forthcoming Journal of the American Statistical Association*, pages 1–15.

Denuit, M. and Lu, Y. (2021). Wishart-gamma random effects models with applications to nonlife insurance. *Journal of Risk and Insurance*, 88(2):443–481.

Edwards, B., Hofmeyr, S., and Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14.

El-Shaarawi, A. H., Zhu, R., and Joe, H. (2011). Modelling species abundance using the Poisson–Tweedie family. *Environmetrics*, 22(2):152–164.

Eling, M. and Jung, K. (2018). Copula approaches for modeling cross-sectional dependence of data breach losses. *Insurance: Mathematics and Economics*, 82:167–180.

Eling, M. and Loperfido, N. (2017). Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, 75:126–136.

Eling, M. and Schnell, W. (2020). Capital requirements for cyber risk and cyber risk insurance: An analysis of solvency II, the US Risk-based capital standards, and the Swiss Solvency Test. *North American Actuarial Journal*, 24(3):370–392.

Eling, M. and Wirfs, J. (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research*, 272(3):1109–1119.

Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, 48(A):367–378.

Fahrenwaldt, M. A., Weber, S., and Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin*, 48(3):1175–1218.

Falco, G., Eling, M., Jablanski, D., Weber, M., Miller, V., Gordon, L. A., Wang, S. S., Schmit, J., Thomas, R., Elvedi, M., et al. (2019). Cyber risk research impeded by disciplinary barriers. *Science*, 366(6469):1066–1069.

Farkas, S., Lopez, O., and Thomas, M. (2021). Cyber claim analysis using Generalized Pareto regression trees with applications to insurance. *Insurance: Mathematics and Economics*, 98:92–105.

FitchRatings (2021). Sharply rising cyber insurance claims signal further risk challenges.

Fokianos, K., Støve, B., Tjøstheim, D., and Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471–499.

Furman, E., Hackmann, D., and Kuznetsov, A. (2020). On log-normal convolutions: An analytical–numerical method with applications to economic capital determination. *Insurance: Mathematics and Economics*, 90:120–134.

Gartner (2021). Gartner forecasts worldwide security and risk management spending to exceed $150 billion in 2021.

Gatzert, N. and Kolb, A. (2014). Risk measurement and management of operational risk in insurance companies from an enterprise perspective. *Journal of risk and insurance*, 81(3):683–708.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37(2):475–515.

Gouriéroux, C. and Lu, Y. (2019). Negative binomial autoregressive process with stochastic intensity. *Journal of Time Series Analysis*, 40(2):225–247.

Gourieroux, C., Monfort, A., and Polimenis, V. (2006). Affine models for credit risk analysis. *Journal of Financial Econometrics*, 4(3):494–530.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Heinen, A. and Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14(4):564–583.

Hillairet, C. and Lopez, O. (2021). Propagation of cyber incidents in an insurance portfolio: Counting processes combined with compartmental epidemiological models. *forthcoming Scandinavian Actuarial Journal*, pages 1–24.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396.

IBM (2020). Cost of a data breach report. Technical report, IBM Security.

Jacobs, J. (2014). Analyzing ponemon cost of data breach.

Jevtić, P. and Lanchier, N. (2020). Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based lan topology. *Insurance: Mathematics and Economics*, 91:209–223.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(2):127–145.

Jørgensen, B., Lundbye-Christensen, S., Song, P.-K., and Sun, L. (1999). A state space model for multivariate longitudinal count data. *Biometrika*, 86(1):169–181.

Jung, K. (2021). Extreme data breach losses: An alternative approach to estimating probable maximum loss for data breach risk. *forthcoming North American Actuarial Journal*, pages 1–24.

Jung, R. C., Liesenfeld, R., and Richard, J.-F. (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29(1):73–85.

Kendall, M. G. (1975). *Rank correlation methods*. Charles Griffin, London, 4th ed. edition.

Kim, E., Gardner, D., Deshpande, S., Contu, R., Kish, D., and Canales, C. (2018). Forecast analysis: Information security, worldwide, 2Q18 update. *Gartner Research*.

Livsey, J., Lund, R., Kechagias, S., and Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *Annals of Applied Statistics*, 12(1):408–431.

Lu, Y. (2018). Dynamic frailty count process in insurance: A unified framework for estimation, pricing, and forecasting. *Journal of Risk and Insurance*, 85(4):1083–1102.

Lu, Y. (2021). The predictive distributions of thinning-based count processes. *Scandinavian Journal of Statistics*, 48(1):42–67.

Maillart, T. and Sornette, D. (2010). Heavy-tailed distribution of cyber-risks. *European Physical Journal B*, 75(3):357–364.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 13(3):245–259.

Mátyás, L., editor (1999). *Generalized Method of Moments Estimation*. Cambridge University Press, Cambridge.

Nguyen, Q. H. and Robert, C. Y. (2015). Series expansions for convolutions of pareto distributions. *Statistics & Risk Modeling*, 32(1):49–72.

OECD (2017). Enhancing the role of insurance in cyber risk management. Technical report, OECD Publishing, Paris.

OECD (2019). Insurance business written in the reporting country. *OECD Insurance Statistics (database)*.

Pedeli, X. and Karlis, D. (2013). On composite likelihood estimation of a multivariate INAR (1) model. *Journal of Time Series Analysis*, 34(2):206–220.

Peng, C., Xu, M., Xu, S., and Hu, T. (2017). Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, 44(14):2534–2563.

Peng, C., Xu, M., Xu, S., and Hu, T. (2018). Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, 45(15):2718–2740.

RMS-CCRS (2016). Managing cyber insurance accumulation risk. Technical report, Risk Management Solutions, Inc. and Cambridge Centre for Risk Studies, Cambridge University.

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7(5):893–899.

Sun, H., Xu, M., and Zhao, P. (2020). Modeling malicious hacking data breach risks. *forthcoming North American Actuarial Journal*, pages 1–19.

Swiss Re (2017). Cyber: Getting to grips with a complex risk. *Sigma*, 1(1).

Tanoue, Y., Kawada, A., and Yamashita, S. (2017). Forecasting loss given default of bank loans with multi-stage model. *International Journal of Forecasting*, 33(2):513–522.

The Geneva Association (2016). Ten key questions on cyber risk and cyber risk insurance. Technical report, The Geneva Association, Zurich.

Trivedi, P. and Zimmer, D. (2017). A note on identification of bivariate copulas for discrete count data. *Econometrics*, 5(1):10.

Wang, L., Akritas, M. G., and Van Keilegom, I. (2008). An anova-type nonparametric diagnostic test for heteroscedastic regression models. *Journal of Nonparametric Statistics*, 20(5):365–382.

Wang, S. and Panjer, H. (1993). Critical starting points for stable evaluation of mixed Poisson probabilities. *Insurance: Mathematics and Economics*, 13(3):287–297.

Wheatley, S., Hofmann, A., and Sornette, D. (2021). Addressing insurance of data breach cyber risks in the catastrophe framework. *Geneva Papers on Risk and Insurance-Issues and Practice*, 46(1):53–78.

Wheatley, S., Maillart, T., and Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *European Physical Journal B*, 89(1):1–12.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 1987(3-4):113–127.

Xu, M. and Hua, L. (2019). Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal*, 23(2):220–249.

Xu, M., Hua, L., and Xu, S. (2017). A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, 59(4):508–520.

Xu, M. and Zhang, Y. (2021). Data breach CAT bonds: Modeling and pricing. *forthcoming North American Actuarial Journal*, pages 1–19.

Zhu, R. and Joe, H. (2009). Modelling heavy-tailed count data using a generalised Poisson-inverse Gaussian family. *Statistics & Probability Letters*, 79(15):1695–1703.