

Is accumulation risk in cyber systematically underestimated?

Gabriela Zeller* Matthias Scherer†

March 30, 2023

Abstract

Many insurers have started to underwrite cyber in recent years. In parallel, they developed their first actuarial models to cope with this new type of risk. On the portfolio level, two major challenges hereby are the adequate modelling of the dependence structure among cyber losses and the lack of suitable data based on which the model is calibrated. The purpose of this article is to highlight the importance of taking a holistic approach to cyber. In particular, we argue that actuarial modelling should not be viewed stand-alone, but rather as an integral part of an interconnected value chain with other processes such as cyber-risk assessment and cyber-claims settlement. We illustrate that otherwise, i.e. if these data-collection processes are not aligned with the actuarial (dependence) model, naïve data collection necessarily leads to a dangerous underestimation of accumulation risk. We illustrate the detrimental effects on the assessment of the dependence structure and portfolio risk by using a simple mathematical model for dependence through common vulnerabilities. The study concludes by highlighting the practical implications for insurers.

Keywords: Cyber Risk; Cyber Insurance; Accumulation Risk; Poisson process.

1 Introduction

Cyber insurance still is a relatively new, but steadily expanding market.¹ Insurers who have recently entered the market and started to establish their cyber portfolios, exploiting the ongoing growth in demand, are becoming increasingly aware of the challenges associated with

*Chair of Mathematical Finance, Technical University of Munich, Parkring 11, 85748 Garching-Hochbrück, Germany; Electronic address: gabi.zeller@tum.de; Corresponding author

†Chair of Mathematical Finance, Technical University of Munich, Parkring 11, 85748 Garching-Hochbrück, Germany; Electronic address: scherer@tum.de

¹In 2015, the global market size was estimated at approximately \$2 billion in premium, with US business accounting for around 90%. A rapid market growth was projected, with total premium reaching \$20+ billion by 2025 ([7]). This estimate currently still seems within reach, with a global market size of around \$7 billion in 2020 ([26]).

insuring cyber risk. These include the dynamically evolving threat landscape, interdependence of risks, heavy-tailed loss severities, and scarcity of reliable data to calibrate (nascent) actuarial models. Particularly the last point is repeated like a mantra; and indeed, while there are growing databases on cyber incidents and their consequences², they often do not contain the information necessary for the various tasks of an actuary. In fact, the best data source which can be adjusted to contain all details to calibrate an insurer’s individual model is the insurer’s own claims-settlement department. While an increasing number of claims in cyber insurance strain insurers’ profitability margins, from the statistical point of view they should be welcomed as the detailed and reliable data whose lack is so frequently lamented. To make full use of the data collected in-house, however, the processes and systems around the underwriting of a cyber portfolio need to be aligned using a holistic approach, where risk assessment, product design, actuarial modelling, and claims settlement are treated as complementary activities interconnected by feedback loops.

In this article, we aim at illustrating the importance for insurers of using the current moment – namely when starting to underwrite cyber risk – to contemplate and establish data-collection processes in risk assessment and claims settlement which allow them to actually use the collected data to calibrate and refine their actuarial models continuously. Sections 2.1 and 2.2, respectively, address the *cyber insurance value chain* in detail to illustrate the above mentioned interconnections and to introduce one particular approach to modelling dependence in cyber, namely via common vulnerabilities.

In Section 3 we introduce a (purposely simplified) mathematical model capturing such a dependence structure to illustrate that straightforward, naïve data collection necessarily leads to accumulation risk being systematically underestimated, both in the statistical and colloquial sense. We show that while this does not necessarily imply erroneous pricing of individual contracts, it may lead to a dangerous underestimation of dependence and portfolio risk. This is illustrated by comparing the common risk measures Value-at-Risk and Expected Shortfall for the total incident number in the portfolio as well as the joint loss arrival rate for any two companies in the portfolio.

Section 4 concludes and highlights the practical implications of this study for insurers.

2 Two Challenges for Cyber Insurance

2.1 A holistic approach to cyber-insurance underwriting

In recent years, various academic papers and numerous empirical studies have been devoted to proposing stochastic models for cyber risk.³ Likewise, the establishment of cyber insurance as a new business line has occupied many insurers and industry subsidiaries such as brokers, see e.g. [6]. Whenever a new insurance line is introduced, the central tasks for actuaries will be technical pricing of the to-be-insured risks and risk management of the resulting

²See e.g. PRC [1] for a publicly available dataset on data breaches and e.g. the commercial providers Advisen [2] and SAS for more specialized datasets.

³For example, game-theoretic models based on a highly stylised understanding of the IT landscape ([17, 35]) or analyses of publicly available cyber loss data to propose frequency and severity distributions ([14, 15]), to name opposite ends of the modelling spectrum.

portfolio (or more precisely in cyber, risk management of an established portfolio, which now additionally contains risks from cyber policies). Underwriting and pricing risks can be done based on expert judgement for each risk individually⁴ or – more commonly – based on a chosen mathematical model. In other words, actuaries have to devise an answer to the question: “*How (do we choose) to model cyber risk?*” Equally important, however, and often overlooked by academic papers, is the observation that it is not reasonable for actuaries to come up with a (no matter how accurate) answer to the above question in the isolation of an actuarial department. Instead, the chosen mathematical model needs to be simultaneously based on and itself be the basis of the business processes surrounding actuarial modelling along the entire economic insurance value chain. The development, calibration, and back-testing of an actuarial model are only sensible if they are based on information and data from risk assessment, product design, and claims settlement, as detailed below and illustrated in Figure 1.

- **Product design:** Before even starting to devise an actuarial model, a clear-cut definition and taxonomy of cyber risk(s) needs to be established in order to determine which aspects of *cyber* are deemed insurable (anything else should be excluded from the coverage by contract design) and which coverage components a cyber insurance policy should consist of. This product design process naturally needs to be revised regularly with the involvement of legal and market experts, as the cyber threat landscape as well as prospective clients’ coverage needs evolve dynamically.
- **Risk assessment:** The risk-assessment process serves to elicit information deemed relevant to estimate a prospective policyholder’s susceptibility to cyber risk. For cyber insurance, this process should naturally include an assessment of the client’s IT infrastructure and existing cyber-security provisions. For an accurate assessment of such technical systems, cooperation with IT security experts is indispensable. However, how to adequately include extensive qualitative knowledge about an IT system’s vulnerabilities and security into a stochastic model is a complex, unresolved issue in itself. Nevertheless, the questions asked and information gathered from prospective policyholders during the risk-assessment process should depend on the actuarial model that is subsequently used for pricing of individual contracts and risk management of the cyber portfolio.
- **Actuarial modelling:** The actuarial modelling step aims at developing a stochastic model which allows an estimation of the distribution of each policy’s and the overall portfolio’s loss from cyber risk. This serves as the basis for (technical) pricing and risk management. The model should be calibrated – and ideally back-tested – using adequate data (once available) and expert judgement. In summary, the choice of stochastic model depends on product design (which types of cyber losses are to be modelled) and in order to calibrate and develop it further, adequate data must be gathered through risk assessment and claims settlement.
- **Claims settlement:** Claims settlement deals with incoming claims from cyber losses in existing policies. In practice, this task is often treated completely disjoint from the above-mentioned processes (except product design), and typically conducted by legal experts

⁴This is indeed common e.g. for very large risks in industrial lines.

whose main concern is to understand the intricacies of each individual claim well enough to judge whether and to which extent it is covered by the components of the policy. The manner of data collection and storage is mostly dictated by legal (and efficiency) concerns. For cyber it is relevant to stress that technical expertise cannot be expected in a classical claims-settlement department. However, this is a crucial shortcoming: The information that needs to be collected in order to make claims data usable for model calibration is dictated by the choice of model. Vice versa, additional information collected may uncover flaws or omissions of the actuarial model and support its continuing development. Therefore, it is important to collect historical claims information with the underlying actuarial model in mind. In cyber, it is well-established consensus that any actuarial model needs to take **dependence** between cyber losses into account. The exact choice of dependence model is of course an insurer's individual decision⁵, but it is clear that if one strives to calibrate such a model based on data, the model choice needs to be reflected in the data-collection process from the insurer's own claims experience.

Depending on the reader's own practical experience, interconnection of the above processes and cooperation between all stakeholders may sound like a utopia or a matter of course. We agree that for established business lines, either may be the case, depending on whether systems and processes were set up and continuously monitored intentionally or rather were allowed to *grow historically*. It is clear that as cyber insurance is just being established, now is the moment to intentionally set up this value chain in a way that enables insurers to cope with the dynamic challenges of this new and continuously evolving risk type in the future.

2.2 Dependence in cyber via common vulnerabilities

It is uncontested that a core actuarial challenge in cyber risk is the failure of the independence assumption between claim occurrences, which underlies the diversification principle in insurance. Due to increasing interconnectivity, businesses, systems, and supply chains become ever more dependent on functional IT infrastructure and crucially, more interdependent. Therefore, including the modelling of dependence in an actuarial model for cyber risk is indispensable. The actuarial literature discusses several approaches for this, most commonly using epidemic spreading on networks / graphs (e.g. [17, 35]), based on (marked / self- or cross-exciting) point processes (e.g. [10, 28, 36]), or employing copula approaches (e.g. [20, 24, 29]).

Regardless of the concrete modelling approach, dependence between cyber losses is worrisome for insurers as it may entail *accumulation risk*, which can be defined e.g. as the

*risk of large aggregate losses from a single event or peril due to the concentration of insured risk exposed to that single event or peril.*⁶

Of course, accumulation risk as a concern is not limited to cyber insurance; other lines of business typically confronted with exposure concentrated to a single event are lines subject

⁵We will advocate for modelling common vulnerabilities as the source of dependence in cyber in the coming sections, but the exact choice of dependence modelling is irrelevant for this argument.

⁶Compare the definition of *risk exposure accumulation* by Casualty Actuarial Society (<https://www.casact.org/>).

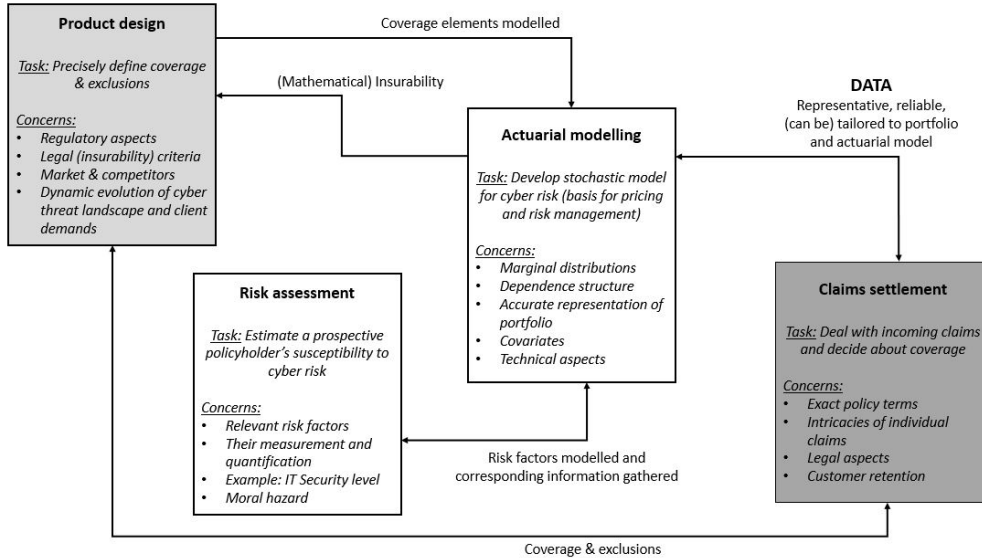


Figure 1: The diagram illustrates the interconnections between different tasks in a holistic insurance value chain. While actuaries are typically mainly involved in risk assessment and actuarial modelling, there are crucial connections to other areas which must not be overlooked. In particular, the necessity to create awareness that meaningful data, which can (and should) be tailored to the chosen actuarial model, is being collected daily in the claims-settlement department (usually by a completely disjoint group of experts, who do not have actuarial modelling aspects on their agenda of primary concerns) should be emphasized.

to natural catastrophes⁷ or marine insurance (see e.g. [4]).

Following the classical decomposition of *risk* into a combination of threat, vulnerability, and impact (see e.g. [23]), a cyber threat only manifests itself as an incident (with potential monetary impact) if there is a corresponding vulnerability in the target system. Therefore, we postulate that any cyber incident is caused by the exploitation of a vulnerability in the company’s system, where it can be distinguished between *symptomatic* and *systemic*⁸ vulnerabilities (see [12, 11]), the former affecting a single company while the latter affect multiple companies simultaneously. Commonly cited examples of systemic vulnerabilities are the usage of the same operating system, cloud service provider, or payment system, affiliation with the same industry sector, or dependence on the same supplier.

Example 1. *We give two recent examples of common vulnerabilities which prominently exposed many companies to a cyber threat simultaneously. The following information and more technical details on both examples can be found in the report [33]. These examples serve to illustrate that in some cases, it might be quite obvious for an insurer to determine from incoming claims data that several cyber claims are rooted in the same common vulnerability,*

⁷For example, Hurricane Katrina has been named as the most expensive event ever to the insurance industry world-wide, see [3].

⁸We remark that some authors (see the recent survey paper [9]) employ a slightly diverging nomenclature: They denote dependency of cyber risks from common vulnerabilities as *systematic risk* and, in turn, understand *systemic risk* to mean cyber risk due to contagion effects in interconnected networks.

whereas in other cases this is very difficult to detect.

- **Microsoft Exchange:** *In the first quarter of 2021, threat actors exploited four zero-day vulnerabilities in Microsoft Exchange Server. The attacks drew widespread media attention due to the high number of affected companies (estimates of 60.000 victims globally, see [34]) within a short time frame, enabled by the ubiquitous use and accessibility of Exchange Servers at organizations world-wide and by their ability to be chained with other vulnerabilities. Due to the massive media coverage, leading to high awareness among companies, and the relatively clear time frame (the attacks had begun in January and were rampant during the first quarter of 2021), it was relatively easy for insurers to identify whether incoming cyber claims during (or slightly after) this time frame were rooted in one of the Microsoft Exchange vulnerabilities.*
- **Print Spooler / Print Nightmare:** *In the third quarter of 2021, several zero-day vulnerabilities were disclosed in Windows Print Spooler, another widely used service in Windows environments. As mentioned in [33], the same service was already exploited in 2010 in the so-called Stuxnet attacks. Stuxnet was a malicious worm consisting of a layered attack, where Windows systems were infected first (through zero-day vulnerabilities), but not the eventual target; i.e. the infection would have usually stayed undetected in the Windows system and sought to propagate to certain (Siemens) PLCs (see, e.g., [18, 32]). These 2010 attacks were not immediately connected to an insurance context. However, if an analogous mechanism (e.g. through the recent Print Spooler vulnerabilities) were to cause cyber insurance claims, it would certainly be hard to attribute all claims to the same common vulnerability for two reasons: First, the eventual target system where the (economic) impact is caused differs from the system affected by the common vulnerability and second, the time frame is much less clear than in the previous example, as the delay between exploitation of the vulnerability and economic impact is somewhat arbitrary.*

In any case, in order to calibrate a model that uses common vulnerabilities as the source of dependence, an insurer needs to collect at least some information about the root cause for each claim to be able to estimate the dependence structure correctly. We now give a very general overview of how information on common vulnerabilities would be reflected in the insurer’s risk modelling process, before introducing a more concrete, slightly simplified mathematical model in Section 3.

2.3 Notation

Assume that an insurer’s portfolio consists of $K \in \mathbb{N}$ companies. From the viewpoint of each company, indexed $i \in \{1, \dots, K\}$, cyber incidents arrive according to a simple point process with corresponding counting process $(N^{(i)}(t))_{t \geq 0} = \left(|\{k \in \mathbb{N} : t_k^{(i)} \in [0, t]\}| \right)_{t \geq 0}$, in the simplest case a homogeneous Poisson process with rate $\lambda^i > 0$. This rate may differ between companies (i.e. some are assumed to be more frequently affected than others) and the main focus of cyber risk assessment (e.g. via a questionnaire, see [19] for a blueprint, or a more extensive audit for larger risks) is to gather information about characteristics which are considered relevant to determine a prospective policyholder’s rate (classical covariates are e.g. company size, type and amount of data stored, types of business activities, see e.g.

[15, 30, 31]).

As the λ^i are naturally unknown, the insurer usually estimates them given past claims experience of *similar* policyholders (depending on the portfolio size, more or less homogeneous groups would be considered *similar*). The overall arrival of incoming incidents to company i is actually composed of several (assumed independent and Poisson) arrival processes (from idiosyncratic incidents and common events), i.e. the overall Poisson rate for company $i \in \{1, \dots, K\}$ decomposes into

$$\lambda^i = \lambda^{i,\text{idio}} + \sum_{s \in S_i^*} \lambda^{s,\text{syst}} > 0, \quad (1)$$

where $\lambda^{i,\text{idio}} \geq 0$ is the rate of idiosyncratic incidents arriving to company i , possibly modelled as some function of the covariates⁹, $S_i^* \subseteq \{1, \dots, S\}$ is the subset of S known systemic risk factors (any common factor through which multiple companies in the portfolio could be affected simultaneously) present at company i and $\lambda^{s,\text{syst}} \geq 0$ is the overall occurrence rate of an event due to exploitation of systemic risk factor $s \in \{1, \dots, S\}$. In this modelling step, several “pitfalls” could occur:

- (1) If questions about relevant covariates are omitted during risk assessment (i.e. because their influence on the frequency of cyber incidents is unknown), this may introduce a bias when estimating $\lambda^{i,\text{idio}}$ (in either direction, i.e. over-/underestimation depending on the covariates).
- (2) If certain systemic risk factors are unknown and therefore not inquired about during risk assessment (e.g. no question about the choice of operating system or cloud service provider) for some or all companies, a systematic underestimation of the true rates is introduced, as the set S , resp. subsets S_i^* , do not contain all possible events.

The errors (1) and (2) should be mitigated by refining risk assessment procedures continuously based on expert input and evaluation of claims data. This leads to the main point of inquiry in this article: Given (correct) assumptions about covariates and systemic risk factors, the goal is to enable the insurer to estimate the corresponding rates, both idiosyncratic and systemic, using historical claims data. As the insurer monitors incoming claims over a policy year $[0, T]$, where typically $T = 1$, in addition to client-related data and basic claims-related data, usually a description of the incident (i.e. the order of occurrences that lead to a monetary loss) is provided by the client. This is unstructured data, and depending on the case could e.g. be given in the form of a phone conversation or e-mail report to an insurance agent or via a scanned PDF containing a report of an IT forensics expert. This information is typically reviewed by the insurance agent in order to decide whether the claim is covered, but may not or only in abbreviated form be entered into the insurer’s claims database. This means that information allowing claims to be identified as stemming from the same systemic vulnerability is often not available or (fully or partly) discarded. In the following, we illustrate the detrimental effect of this omission of information about the extent of systemic events on the estimation of dependence and portfolio risk.

⁹For example, fitting a standard GLM or GAM here would be common practice.

3 Mathematical model

To quantify the effect we have introduced and discussed on a qualitative level in Section 2, we now construct a simple mathematical model which captures common events (‘shocks’) and allows to analyze the effect of underestimating the extent of joint events.

3.1 An exchangeable portfolio model and the modelling of missing information

We assume that the insurer’s portfolio consists of $K \in \mathbb{N}$ homogeneous companies and let $\emptyset \subset I \subseteq \{1, \dots, K\}$ denote a non-empty subset of the portfolio affected by a common event. Assume that cyber events (to any set I) arrive according to independent, homogeneous Poisson processes. In theory, each subset I could potentially have a different arrival rate of common events, leading to the prohibitive complexity of needing to estimate $2^K - 1$ rates. To avoid the curse of dimensionality, we make the following assumption.

Assumption 1 (Exchangeability: Equal rates for subsets of equal size). *Assume that arrival rates only depend on the number of companies in the subset, i.e. the insurer aims at estimating a vector of K arrival rates $\boldsymbol{\lambda} := (\lambda^{|I|=1}, \dots, \lambda^{|I|=K})$, where $\lambda^{|I|=k}$ denotes the arrival rate of events affecting any subset of size $k \in \{1, \dots, K\}$.*

We denote as model (M) the model given these ‘true’ rates $\boldsymbol{\lambda}$.¹⁰ Assumption 1 leads to homogeneous marginal arrival rates λ^i , $i \in \{1, \dots, K\}$, for each company of

$$\lambda^i = \sum_{k=1}^K \frac{\lambda^{|I|=k}}{\binom{K}{k}} \binom{K-1}{k-1} = \sum_{k=1}^K \frac{k}{K} \lambda^{|I|=k} = \underbrace{\frac{\lambda^{|I|=1}}{K}}_{\text{idiosyncratic incidents}} + \underbrace{\sum_{k=2}^K \frac{k}{K} \lambda^{|I|=k}}_{\text{incidents from common events}}. \quad (2)$$

Note that (2) is a simplified formalisation of (1).

It is well-known that the maximum likelihood estimator of the rate of a homogeneous Poisson process is given by the sample mean (see e.g. [13]) over the observation period, i.e. in our case each estimator $\hat{\lambda}^{|I|=k}$ is given by the mean total number of observed events affecting precisely k companies¹¹, i.e. for $L > 0$ observed policy years

$$\hat{\lambda}^{|I|=k} = \frac{1}{L} \sum_{\ell=1}^L \hat{n}_\ell^{|I|=k},$$

¹⁰Note that model (M) describes a setting where the first claim-arrival times, denoted $\tau = (\tau_1, \dots, \tau_K)$, of the companies in the portfolio follow an *exchangeable Marshall-Olkin distribution*, see [22], p. 122ff. Note that in contrast to [22], we denote by $\lambda^{|I|=k}$ the arrival rate of the Poisson process that is essentially the superimposed process of all arrival processes to subsets of size k , i.e. the rate for every particular subset of size k would be (independently of the subset) given by $\lambda_k := \frac{\lambda^{|I|=k}}{\binom{K}{k}}$. For example, for $k = 1$, $\lambda^{|I|=1}$ describes the overall rate of events affecting one single firm. As the model is exchangeable, each firm is equally likely to be affected by such an event, i.e. from the viewpoint of each of the K firms, these events arrive with rate $\lambda_1 = \frac{\lambda^{|I|=1}}{K}$.

¹¹For simplicity, we assume policy years of length $T = 1$, during which the portfolio does not change.

where $\hat{n}_\ell^{|I|=k}$ is the number of observed events to subsets of size k during policy year (or simulation run) $\ell \in \{1, \dots, L\}$.

Assumption 2 (Missing information on common events). *Assume that, independently for each common event to a subset of any size $|I| \geq 2$ and independently for each company in the subset, i.e. $i \in I$, the probability that the arrival at this company is correctly identified as belonging to the common event (affecting all companies in I) is given by $p \in [0, 1]$.¹²*

Example 2. *To illustrate Assumption 2, consider the following situation: A vulnerability in a commonly used software could be exploited, leading to hackers gaining access to confidential data which allowed them to defraud several companies throughout the policy year. After the policy year, when historical claims data is analyzed, all incidents in the database are first considered independent. Those incidents where detailed information is available, in this case that the original cause of the loss was the exploit of the common vulnerability, are then identified as belonging to a common event. If originally five companies were affected in this way, but only for three of them the required information was available, instead of (correctly) counting one observed event on a subset of five companies (contribution to the estimator $\hat{\lambda}^{|I|=5}$), the insurer would (incorrectly) count one event on a subset of three companies and two independent incidents (contribution to the estimators $\hat{\lambda}^{|I|=3}$ and twice to $\hat{\lambda}^{|I|=1}$).*

Mathematically, Assumption 2 means that the Poisson arrival processes to subsets of size $|I| = k \geq 2$ are subject to *thinning* (with probability $(1 - p^k)$) and *superposition* of $(K - k)$ other Poisson arrival processes.

Definition 1 (Model (\widetilde{M}) - missing information). *Assumption 2 leads to a different model, denoted (\widetilde{M}) , with Poisson arrival rates denoted $\widetilde{\lambda} := (\widetilde{\lambda}^{|I|=1}, \dots, \widetilde{\lambda}^{|I|=K})$ given by*

$$\widetilde{\lambda}^{|I|=1} = \lambda^{|I|=1} + \sum_{i=2}^K \lambda^{|I|=i} \left[i(f_{Bin}(0; i, p) + f_{Bin}(1; i, p)) + \sum_{j=2}^{\max(i-1, 2)} (i-j)f_{Bin}(j; i, p) \right], \quad (3)$$

$$\widetilde{\lambda}^{|I|=k} = \sum_{i=k}^K \lambda^{|I|=i} f_{Bin}(k; i, p), \quad k \in \{2, \dots, K\}, \quad (4)$$

where $f_{Bin}(k; i, p) = \binom{i}{k} p^k (1-p)^{i-k}$ is the p.m.f. of a Binomial distribution.

Remark 1 (Interpretation of the rates $\widetilde{\lambda}$). *The rates $\widetilde{\lambda}$ can be interpreted as follows:*

- For $k = K$, the rate in the model with missing information is given by

$$\widetilde{\lambda}^{|I|=K} = \lambda^{|I|=K} f_{Bin}(K; K, p) = \lambda^{|I|=K} p^K,$$

i.e. the original rate thinned by the probability that all (of the K independently investigated) incidents are identified correctly. Note that for $p \in [0, 1)$, $\widetilde{\lambda}^{|I|=K} < \lambda^{|I|=K}$, i.e. the rate of events that jointly affect the whole portfolio is obviously lowered.

¹²A straightforward generalisation would be to assume different detection probabilities for different event sizes, i.e. a vector $\mathbf{p} := (p^{|I|=2}, \dots, p^{|I|=K})$. Intuitively, this may e.g. be used to represent the assumption that incidents from larger events are more likely to be detected, as such events are often subject to public coverage (see e.g. the *Microsoft Exchange* example above) and therefore insurers may already be alert to check if recorded claims belong to this same root cause.

- For $1 < k < K$, the rate in the model with missing information is given by the sum of the original rate for $i = k$ thinned by the probability of classifying all k incidents correctly (summand for $i = k$) and the rates resulting from the probabilities of misclassifying events to more than k firms incorrectly such that they are counted as events to k firms (summands for $i > k$); compare Example 2. $\tilde{\lambda}^{|I|=k}$ can thus be higher or lower than $\lambda^{|I|=k}$, depending on λ and p . However, in general, the cumulative rate of ‘small’ events (i.e. all events up to any size k) does not decrease, i.e.

$$\sum_{i=1}^k \tilde{\lambda}^{|I|=i} \geq \sum_{i=1}^k \lambda^{|I|=i}, \quad \forall k \in \{1, \dots, K\}.$$

- The rate for idiosyncratic incidents in model (\widetilde{M}) is given by the sum of the original rate (these incidents are never “misclassified”) and all the “fallout” from classifying common events incorrectly: If for an event to a subset of size i , none or only one of the firms are classified correctly, all i incidents will be counted as idiosyncratic (first part in square bracket in (3)); if $j \geq 2$ firms are attributed correctly, the remaining $i - j$ are classified as idiosyncratic (second part in square bracket in (3)). Therefore, for $p \in [0, 1)$, it holds $\tilde{\lambda}^{|I|=1} > \lambda^{|I|=1}$, i.e. the rate of idiosyncratic incidents is increased.

Lemma 1 (Marginal rates remain unchanged). *The marginal arrival rates for each company stay unchanged between model (M) and model (\widetilde{M}) , i.e.*

$$\tilde{\lambda}^i = \lambda^i = \sum_{k=1}^K \frac{k}{K} \lambda^{|I|=k}, \quad i \in \{1, \dots, K\}.$$

Proof. Intuitively, the statement is clear, as an incorrect (non-)identification of common events does not lead to missing a claim, but to wrongly attributing its cause. A formal proof is given in Appendix A. \square

The interpretation of Lemma 1 is of high practical relevance: For pricing of (cyber) insurance policies, usually only the individual loss distribution of a company is taken into account. As the marginal arrival rates stay unchanged, prices for all individual insurance contracts would stay unchanged (i.e. ‘correct’) between models (M) and (\widetilde{M}) . This means that omitting information about common events would not lead to mispricing of individual policies. This identity of marginal rates is dangerous, as the crucial oversight of underestimating the extent of common events would not be evident as affecting (average) profitability, but only in a (worst-case) scenario that an unexpectedly large loss (exceeding the estimated risk measure, typically Value-at-Risk, which may be much smaller in model (\widetilde{M}) than the actual one in model (M) , see next section) manifests.

3.2 Implications for dependence- and risk-measurement

Portfolio Value-at-Risk

Despite the marginal rates staying unchanged when moving from (M) to (\widetilde{M}) , see Lemma 1, omitting information about common events may have dangerous implications for risk man-

agement. We first illustrate how it may lead to an underestimation of portfolio risk, measured e.g. by Value-at-Risk, denoted $\mathbf{VaR}_{1-\gamma}$, of the total incident number in the portfolio in a policy year.¹³ $\mathbf{VaR}_{1-\gamma}$ for a r.v. X in an actuarial context (where positive values denote losses) is defined as

$$\mathbf{VaR}_{1-\gamma}(X) = \inf \{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq 1 - \gamma\}, \quad \gamma \in (0, 1). \quad (5)$$

Note that the overall incident number in a portfolio of size K follows a compound Poisson distribution, i.e.

$$S(T) := \sum_{i=1}^{N(T)} Z_i, \quad \text{where } N(T) \sim \text{Poi}\left(T \sum_{k=1}^K \lambda^{|I|=k}\right),$$

$$\{Z_i\}_{i \in \mathbb{N}} \text{ i.i.d. with } \mathbb{P}(Z_i = k) = \frac{\lambda^{|I|=k}}{\sum_{k=1}^K \lambda^{|I|=k}}, \quad \forall k \in \{1, \dots, K\}.$$

The rate $\left(\sum_{k=1}^K \lambda^{|I|=k}\right)$ corresponds to the overall Poisson arrival rate of events (of any size), and $\{Z_i\}_{i \in \mathbb{N}}$ correspond to the associated “jump sizes” of the total incident number, i.e. the number of companies affected in the i^{th} event. Therefore, we can use the *Panjer recursion* formula (based on [27]) to compute the probability mass function (p.m.f.) and corresponding cumulative distribution function (c.d.f.) and Value-at-Risk (as in Equation (5)) of the total incident number in a policy year under models (M) and (\widetilde{M}) for chosen $\boldsymbol{\lambda}$ and $p \in [0, 1]$. We choose an exemplary set of rates for a portfolio of size $K = 10$ as given in Table 1, where $\boldsymbol{\lambda}$ again denotes the rates of an original model (M) and $\widetilde{\boldsymbol{\lambda}}$ the rates of the corresponding model (\widetilde{M}) resulting from Assumption 2.

Figure 2a displays the p.m.f. under model (M) and highlights the comparison of $\mathbf{VaR}_{0.995}$ for $p = 1$ (full information, i.e. original rates), $p = 0.5$ (partial information about common events, compare Table 2), and $p = 0$ (no information about common events, i.e. complete independence assumption). Figure 2b compares $\mathbf{VaR}_{1-\gamma}$ for $(1 - \gamma) \in \{0.95, 0.995\}$ and $p \in [0, 1]$, based on the c.d.f. of total incident numbers under the rates $\boldsymbol{\lambda}$ and $\widetilde{\boldsymbol{\lambda}}$. This small example already highlights the importance of gathering (full!) information about the origins of cyber incidents, as otherwise the portfolio risk will be drastically underestimated.

Finally, let us mention an observation that can be made by considering the p.m.f. (and corresponding c.d.f.) for different $p \in [0, 1]$, as exemplarily depicted in Figure 3: When moving from (M) to (\widetilde{M}) , no events / incidents are missed completely, thus the c.d.f.s of the total incident number in the portfolio are **not** ordered in the sense of *usual stochastic order*, i.e. it does not hold that for all $x \geq 0 : F_{S_{\widetilde{M}}(T)}(x) \geq F_{S_M(T)}(x)$, where $S_M(T)$ (resp. $S_{\widetilde{M}}(T)$) denotes the total incident number under model (M) (resp. (\widetilde{M})).

We have observed, however, from the results illustrated in Table 2 and Figure 2, that this ordering of c.d.f.s does hold for certain large values of x . Figure 3b shows that indeed it holds *exactly* for large values of x , more precisely $x > x_0$ for some $x_0 \geq 0$, i.e. the so-called *single-crossing condition* or *cut-off criterion* (see e.g. [25]) is fulfilled here. This is meaningful as

¹³For the sake of simplicity, we only consider incident numbers here, as of course the results would not be qualitatively different if for an insurance application, one were to equip each incident with a (random) monetary loss size.

Model	p	$ I = 1$	$ I = 2$	$ I = 3$	$ I = 4$	$ I = 5$	$ I = 6$	$ I = 7$	$ I = 8$	$ I = 9$	$ I = 10$
(M)	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(\widetilde{M})	0.5	29.49	1.93	1.77	1.45	1.00	0.55	0.23	0.07	0.01	0.0010
(\overline{M})	0	55.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 1: Original rates and resulting rates for $p = 0.5$ (i.e. for each event affecting a subset of at least two firms jointly, the incident at each firm is attributed correctly to this event with probability $p = 0.5$ and otherwise incorrectly seen as independent as a result of not being able to identify the common root cause) and $p = 0$. By partially omitting information about common events, the resulting idiosyncratic rates are much increased, rates of smaller common events (here up to $|I| = 4$) are also increased, whereas rates of larger common events (here from $|I| = 6$ on) are lowered.

Model	p	λ^i ($i \in \{1, \dots, K\}$)	$\mathbb{E}[S(T)]$	$\mathbf{VaR}_{0.95}(S(T))$	$\mathbf{VaR}_{0.99}(S(T))$	$\mathbf{VaR}_{0.995}(S(T))$
(M)	1	5.5	55	90	107	113
(\widetilde{M})	0.5	5.5	55	76	86	90
(\overline{M})	0	5.5	55	68	74	76

Table 2: Resulting marginal rates (homogeneous for all companies), expected total incident numbers, and risk measures $\mathbf{VaR}_{1-\gamma}(S(T))$ at three levels for $p \in \{0, 0.5, 1\}$ and $T = 1$. Crucially, marginal rates and thus expected incident numbers $\mathbb{E}[S(T)]$ do not change (by Lemma 1 and linearity), while $\mathbf{VaR}_{1-\gamma}(S(T))$ at all chosen levels is systematically lowered when common event information is partly or fully disregarded.

it is a sufficient condition for another (weaker) type of stochastic order, so-called *increasing convex order*, which has an important connection to the class of coherent risk measures; this will be addressed more generally in a subsequent section.

Quantifying dependence by joint loss arrival rate

From a practical viewpoint, the illustrations of the last section already emphasize the detrimental effects of missing information about common events. Theoretically, there are different quantities one might use to assess the extent of “missed / overlooked dependence” in model (\widetilde{M}) compared to the true model (M) . From a risk management perspective, it is clear that simultaneous losses by multiple policyholders carry potentially greater risk than independent, diversifiable losses. Therefore, one might look at the instantaneous rate of two policyholders $i, j \in \{1, \dots, K\}$, $i \neq j$, simultaneously experiencing a cyber claim.¹⁴ As arrivals of cyber incidents to policyholder $i \in \{1, \dots, K\}$ follow a Poisson process with rate λ^i (see (2)), the first arrival time, denoted τ_i , follows an exponential distribution and for small $T > 0$ it holds by a first-order Taylor expansion

$$\mathbb{P}(\tau_i \leq T) = 1 - e^{-\lambda^i T} \approx 1 - (1 - \lambda^i T) = \lambda^i T \iff \frac{1}{T} \approx \frac{\lambda^i}{\mathbb{P}(\tau_i \leq T)}.$$

¹⁴As we are assuming an *exchangeable* model, w.l.o.g. $i = 1$, $j = 2$.

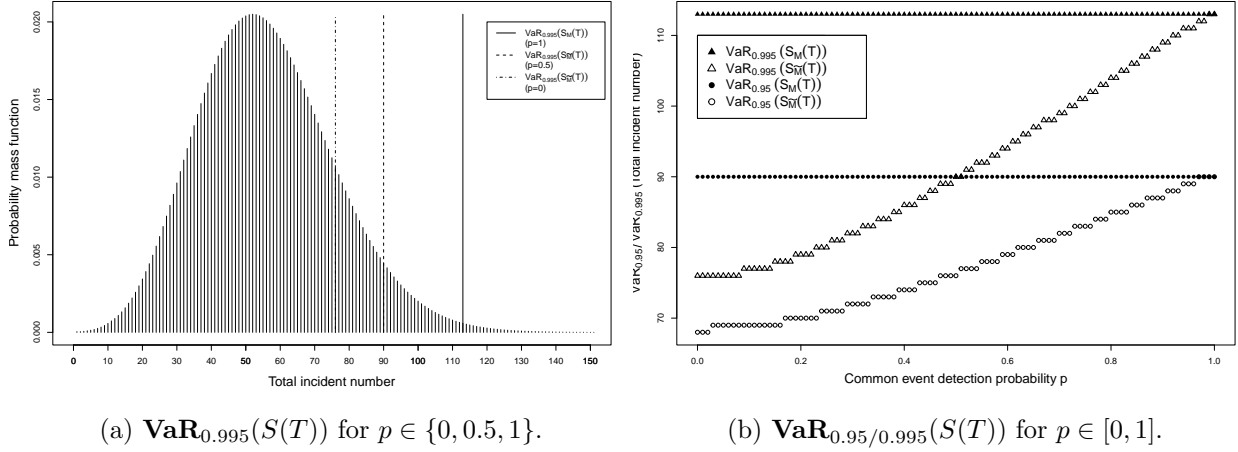


Figure 2: Panel 2a shows the p.m.f. of the total incident number for parameters as in Table 1 and again $T = 1$. The solid vertical line depicts the corresponding $\mathbf{VaR}_{0.995}$ if full information about common events is available ($p = 1$), i.e. all incidents are classified correctly. The dashed lines depict analogously $\mathbf{VaR}_{0.995}$ for partial information ($p = 0.5$, i.e. for each event on average half of the resulting incidents are attributed correctly), and no information ($p = 0$, i.e. all incidents regarded as idiosyncratic) about common events. In both latter cases, the true risk is clearly underestimated (compare $\mathbf{VaR}_{0.995}$ for $p = 0$ with the ‘true’ underlying distribution!). Panel 2b shows \mathbf{VaR} . for $(1 - \gamma) \in \{0.95, 0.995\}$ and $p \in [0, 1]$ (in steps of $\Delta = 0.01$), based on underlying rates λ and $\bar{\lambda}$. As expected, the lower the probability p of correctly identifying a common root cause, the more severe is the resulting underestimation of the risk.

This implies for the instantaneous joint loss arrival rate

$$\lim_{T \searrow 0} \frac{\mathbb{P}(\tau_i \leq T, \tau_j \leq T)}{T} \approx \lim_{T \searrow 0} \frac{\lambda^i \mathbb{P}(\tau_i \leq T, \tau_j \leq T)}{\mathbb{P}(\tau_i \leq T)} = \lambda^i \lim_{T \searrow 0} \mathbb{P}(\tau_j \leq T \mid \tau_i \leq T) = \lambda^i \text{LTD}_C, \quad (6)$$

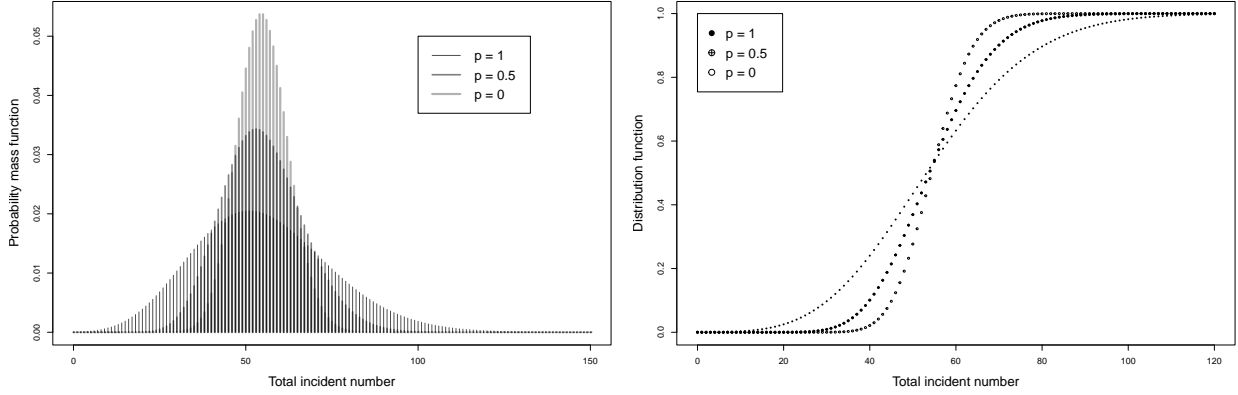
where τ_i, τ_j are the first arrival times of a cyber claim to policyholders i and j , respectively, and LTD_C denotes the *lower tail dependence coefficient* of the bivariate copula C of (τ_i, τ_j) . We know (see [22], p. 122ff) that by Assumption 1 the survival copula of the random vector of all K first claim-arrival times, (τ_1, \dots, τ_K) , is an *exchangeable Marshall–Olkin (eMO)* survival copula, and its two-margins (i.e. the survival copula of (τ_i, τ_j)) are bivariate *Cuadras–Augé* copulas with parameter α given by¹⁵:

$$\alpha = 1 - \frac{\sum_{i=1}^{K-1} \binom{K-2}{i-1} \frac{1}{\binom{K}{i}} \lambda^{|I|=i}}{\sum_{i=1}^K \binom{K-1}{i-1} \frac{1}{\binom{K}{i}} \lambda^{|I|=i}} = 1 - \frac{\sum_{i=1}^{K-1} \binom{K-2}{i-1} \lambda_i}{\sum_{i=1}^K \binom{K-1}{i-1} \lambda_i}. \quad (7)$$

From (7), some interpretation of α is immediately visible:

- Comonotonicity occurs iff only common events to the whole portfolio occur, i.e. $\alpha = 1 \iff \lambda_K > 0, \lambda_i = 0 \forall i \in \{1, \dots, K-1\}$;
- Independence occurs iff only idiosyncratic incidents occur, i.e. $\alpha = 0 \iff \lambda_1 > 0, \lambda_i = 0 \forall i \in \{2, \dots, K\}$.

¹⁵See the previous footnote on the relation of $\lambda^{|I|=i}$ and λ_i .



(a) Probability mass function of $S(T)$ for $p \in \{0, 0.5, 1\}$.

(b) Cumulative distribution function of $S(T)$ for $p \in \{0, 0.5, 1\}$.

Figure 3: Panel 3a shows the p.m.f. of the total incident number for rates λ as in Table 1, $T = 1$, and resulting rates $\tilde{\lambda}$ for $p \in \{0, 0.5\}$. Figure 3b analogously plots the c.d.f.s, illustrating that while the c.d.f.s are not ordered in the sense $F_{S_{\tilde{M}}(T)}(x) \geq F_{S_M(T)}(x)$, $\forall x \geq 0$, there is a threshold value x_0 s.t. this ordering holds (exactly) for large values $x > x_0 \geq 0$, i.e. the so-called *single-crossing condition* is fulfilled here. In the actuarial context, one is typically interested in high quantiles of the loss distribution ($\mathbf{VaR}_{1-\gamma}$ for $(1 - \gamma)$ close to 1), i.e. the region where in this case it holds for the quantile functions $F_{S_{\tilde{M}}(T)}^{\leftarrow}(1 - \gamma) \leq F_{S_M(T)}^{\leftarrow}(1 - \gamma)$, leading to the observations for the portfolio risk measure discussed in this section.

Definition 2 (Bivariate Cuadras–Augé copula, [22], p. 9). For $\alpha \in [0, 1]$, let $C_\alpha : [0, 1]^2 \mapsto [0, 1]$ be defined by

$$C_\alpha(u_1, u_2) := \min\{u_1, u_2\} \max\{u_1, u_2\}^{1-\alpha}, \quad u_1, u_2 \in [0, 1].$$

Remark 2 (Tail dependence coefficients of Cuadras–Augé (survival) copula ([22], p. 34f)). For a bivariate Cuadras–Augé copula C_α , the tail dependence coefficients are given by

$$UTD_{C_\alpha} = \alpha, \quad LTD_{C_\alpha} = \mathbf{1}_{\{\alpha=1\}}.$$

Note that in general for a copula C and its survival copula \hat{C} , it holds (provided existence) that $UTD_C = LTD_{\hat{C}}$ and $LTD_C = UTD_{\hat{C}}$, respectively.

This means for the comparison of the instantaneous joint loss arrival rate in (6), we are interested in comparing the parameter α (as in (7)) for models (M) and (\tilde{M}) .

Remark 3 ($LTD_{\hat{C}_\alpha}$ for constant λ). Assume $\lambda^{|I|=i} \equiv \bar{\lambda} > 0$, $\forall i \in \{1, \dots, K\}$. Then, in model (M) the lower tail dependence coefficient of the bivariate copula of (τ_i, τ_j) is given by

$$LTD_{\hat{C}_\alpha} = \alpha = \frac{2}{3},$$

and the instantaneous joint loss arrival rate in (6) is given by

$$\lim_{T \searrow 0} \frac{\mathbb{P}(\tau_i \leq T, \tau_j \leq T)}{T} = \lambda^i \alpha = \frac{\bar{\lambda}(K+1)}{2} \cdot \frac{2}{3} = \frac{\bar{\lambda}(K+1)}{3}.$$

Lemma 2 (Relation of $LTD_{\tilde{C}_\alpha}$ for models (M) and (\widetilde{M})). *Let (M) be an exchangeable model as in Assumption 1 with any vector of arrival rates $\boldsymbol{\lambda}$ and let (\widetilde{M}) be the corresponding model according to Definition 1. Let α and $\tilde{\alpha}$ be the respective parameters of the bivariate survival copulas of (any two) first-arrival times (τ_i, τ_j) as given in (7). Then, it holds that $\tilde{\alpha} \leq \alpha$ and more specifically, under Assumption 2,*

$$\tilde{\alpha} = p^2 \alpha$$

for any $p \in [0, 1]$.

Lemma 2 implies that in model (\widetilde{M}) , by omitting information about common events according to Assumption 2, the instantaneous joint loss arrival rate for any two companies in the portfolio is underestimated by a factor of p^2 , which intuitively makes sense, as this factor indicates the probability of independently not overlooking a joint event in two companies.

Stochastic ordering and coherent risk measures

Above, we have observed exemplarily that the portfolio risk when measured by Value-at-Risk (at ‘relevant’ levels in an actuarial context, see the remark about the single-crossing condition above and illustration in Figure 3b) is underestimated in a model with missing information (\widetilde{M}) compared to an original model (M) . Another important risk measure is *Expected Shortfall* (at level $(1 - \gamma)$), in the following denoted $\mathbf{ES}_{1-\gamma}(X)$ for a r.v. X in the actuarial context, defined as (see e.g. [5]):

$$\mathbf{ES}_{1-\gamma}(X) = \frac{1}{\gamma} \int_{1-\gamma}^1 \mathbf{VaR}_z(X) dz, \quad (8)$$

where $\mathbf{VaR}_z(X)$ is defined in (5). It is well-known that $\mathbf{ES}_{1-\gamma}$ possesses in a certain sense preferable analytical properties compared to $\mathbf{VaR}_{1-\gamma}$, in particular $\mathbf{ES}_{1-\gamma}$ is a *coherent* risk measure.¹⁶¹⁷ The fact of $\mathbf{ES}_{1-\gamma}$ being coherent allows to draw some interesting theoretical conclusions for the present study presented below in Corollary 1. As a basis, we use the more general observation on the stochastic ordering of compound Poisson random variables summarized in the following theorem.

Theorem 1 (Increasing convex order for specific compound Poisson distributions). *Let $L > 0$ and $\ell \in \mathbb{N}$ and consider two independent homogeneous Poisson processes with intensities*

¹⁶See the seminal work of [8] for the definition and properties of coherent risk measures and e.g. [16] for a collection of proofs of the coherence of expected shortfall.

¹⁷Note that the term ‘expected shortfall’ is often simply used interchangeably with ‘average / tail / conditional Value-at-Risk’ or ‘tail conditional expectation’, which are in turn usually used synonymously. In an actuarial context, the most well-known definition is $\mathbf{TVaR}_{1-\gamma}(X) = \mathbb{E}[X|X \geq \mathbf{VaR}_{1-\gamma}(X)]$, i.e. the expected loss given that a loss at least equal to the Value-at-Risk occurs. However, many equivalencies between the above risk measures, and in particular the coherence of the risk measures other than $\mathbf{ES}_{1-\gamma}$ as defined in (8), only hold if X follows a continuous distribution; see [5] for a detailed discussion. As in the context of this work, discrete underlying distributions (of incident numbers) occur, we therefore only consider $\mathbf{ES}_{1-\gamma}$.

$\lambda > 0$ and $\tilde{\lambda} := \ell \lambda > 0$, denoted $N(t) := (N(t))_{t \geq 0}$ and $\tilde{N}(t)$, respectively. For any fixed $T > 0$, let

$$S(T) := \sum_{i=1}^{N(T)} L = L N(T) \quad \text{and} \quad \tilde{S}(T) = \sum_{i=1}^{\tilde{N}(T)} \frac{L}{\ell} = \frac{L}{\ell} \tilde{N}(T).$$

Then, $\mathbb{E}[S(T)] = \mathbb{E}[\tilde{S}(T)]$ and

$$S(T) \geq_{icx} \tilde{S}(T), \tag{9}$$

where \geq_{icx} denotes ‘increasing convex order’.

Proof. See Appendix A.¹⁸ □

Remark 4 (Notes to Theorem 1).

1. Note that $S(T) \geq_{icx} \tilde{S}(T)$ and $\mathbb{E}[S(T)] = \mathbb{E}[\tilde{S}(T)]$ is equivalent to $S(T) \geq_{cx} \tilde{S}(T)$ (‘convex order’), see [25], Theorem 1.5.3.
2. In actuarial science, a perhaps more common, synonymous name for ‘increasing convex order’ (\geq_{icx}) is ‘stop-loss order’ (\geq_{sl}), which stems from an important characterization of \geq_{icx} by the so-called stop-loss transforms (see [25], Theorem 1.5.7):

$$X \leq_{icx} Y \iff \mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+] \quad \forall t \in \mathbb{R}. \tag{10}$$

3. Note that $S(T)$ and $\tilde{S}(T)$ can be interpreted as two collective risk models with equal expected total claims amount $\mathbb{E}[S(T)] = \mathbb{E}[\tilde{S}(T)]$, where
 - ◇ $S(T)$ is the total claims amount from a model with relatively few, large losses (of deterministic size $L > 0$), and
 - ◇ $\tilde{S}(T)$ is the total claims amount from a model with relatively many, small losses (of deterministic size $0 < \frac{L}{\ell} < L$).

Thus, Theorem 1 states that the model with on average many (independent) small losses is preferable (‘less risky’) in the sense of increasing convex order compared to a model with equal expected claims amount and on average few (independent) large losses.

Corollary 1 (Expected Shortfall for models (M) and (\tilde{M})). Let $\mathbf{ES}_{1-\gamma}(\cdot)$ denote Expected Shortfall as in (8) and let $S_M(T)$ and $S_{\tilde{M}}(T)$ denote the total incident number in the portfolio under models (M) and (\tilde{M}) , respectively, until a fixed time $T > 0$. Then, for any $T > 0$ and any $\gamma \in (0, 1)$, it holds

$$\mathbf{ES}_{1-\gamma}(S_M(T)) \geq \mathbf{ES}_{1-\gamma}(S_{\tilde{M}}(T)). \tag{11}$$

Proof. See Appendix A. □

This implies that by omitting information about common events, the portfolio risk is necessarily underestimated when using expected shortfall (or any other coherent risk measure).

¹⁸Somewhat surprising to us, we did not find the (or a correspondent) statement of the theorem in the literature, hence, for completeness we provide an elementary proof in the Appendix.

4 Conclusion

When insurers started to develop actuarial models for cyber risk, they soon emphasized that one major challenge is the lack of adequate data to calibrate and backtest their models. Many classical actuarial models are based on the assumption of independence between losses and historical data is mainly used to draw inference about individual policyholders' loss distributions (i.e. the parameters of their loss frequency and severity distribution for a certain risk). Indeed, this is sufficient in markets where the claims are independent. Risk assessment and claims settlement therefore usually take into account this individual client-specific information. However, in the case of cyber, collecting such individual information alone is not sufficient, as not only parameters of the individual (marginal) loss distributions, but also those of an adequate model of dependence, have to be calibrated. This is only possible if information about dependence between historical claims, i.e. that losses may have stemmed from the same cause, is systematically collected.

This article has used a stylized mathematical model to highlight the effects on portfolio risk measurement if information on common events is fully or partly discarded. In practice, and we have to raise a big warning sign here, the resulting underestimation of accumulation risk would only become evident too late, namely once a (to-be-avoided) extreme portfolio loss has occurred.

The urgent practical implications for insurers are evident: As outlined in Section 2.1, actuarial modelling of cyber cannot be regarded as an isolated challenge, but as one interconnected step in the insurance value chain. Actuaries therefore must be in continuous exchange with other stakeholders, in particular legal experts (regarding insurability of cyber, product design, and requirements on the collection of claims settlement data) and information security experts. The central importance of the latter group for the actuarial modelling of cyber can hardly be overstated; their expertise is essential in tackling important challenges such as how to include an extensive qualitative assessment of a company's IT landscape, including existing security provisions, into a stochastic actuarial model.

Only continuous interdisciplinary cooperation will allow to develop a holistic approach which allows insurers to proactively steer their cyber underwriting activities without exposing themselves to potentially starkly underestimated levels of accumulation risk.

Acknowledgement

This work was conducted at ERGO Center of Excellence in Insurance at Technical University of Munich and the authors would like to thank ERGO Group AG for supporting this research and providing access to insurance industry expertise.

References

- [1] <https://privacyrights.org/data-breaches>.
- [2] <https://www.advisenltd.com/data/cyber-loss-data/>.

- [3] *Hurricane Katrina 10.* <https://www.agcs.allianz.com/news-and-insights/reports/lessons-learned-from-hurricane-katrina.html>, 2015.
- [4] *Tianjin Blast Could Be Largest Marine Insurance Loss Ever.* <https://maritime-executive.com/article/tianjin-blast-could-be-largest-marine-insurance-loss-ever>, 2016.
- [5] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.
- [6] Advisen. *2018 Cyber Guide: The Ultimate Guide to Cyber Service Providers.* <https://www.advisenltd.com/media/reports/cyber-guide/>.
- [7] Allianz Global Corporate & Specialty. *A Guide to Cyber Risk: Managing the Impact of Increasing Interconnectivity.* <https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-CyberRisk-report.pdf>, 2015.
- [8] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [9] K. Awiszus, T. Knispel, I. Penner, G. Svindland, A. Voß, and S. Weber. Modeling and pricing cyber insurance: Idiosyncratic, systematic, and systemic risks. *European Actuarial Journal*, 2023.
- [10] Baldwin, A., Gheyas, I., Ioannidis, C., Pym, D., and Williams, J. Contagion in cyber security attacks. *Journal of the Operational Research Society*, 68(7):780–791, 2017.
- [11] Bandyopadhyay, T., Mookerjee, V., and Rao, R. Why IT managers don’t go for cyber-insurance products. *Communications of the ACM*, 52(11):68, 2009.
- [12] Böhme, R., Laube, S., and Riek, M. A fundamental approach to cyber risk analysis. *Variance*, 11(2), 2018.
- [13] Daley, D.J. and Vere-Jones, D. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods.* Springer, 2. ed. edition, 2003.
- [14] Edwards, B., Hofmeyr, S., and Forrest, S. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14, 2016.
- [15] Eling, M. and Wirfs, J.H. What are the actual costs of cyber risk events? *European Journal of Operational Research*, 272(3):1109–1119, 2019.
- [16] P. Embrechts and R. Wang. Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling*, 3(1), 2015.
- [17] Fahrenwaldt, M., Weber, S., and Weske, K. Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin*, 48(3):1175–1218, 2018.
- [18] Murchu L. Falliere, N. and E. Chien. W32.stuxnet dossier. *Symantec*, 2010.

- [19] Gesamtverband der Deutschen Versicherungswirtschaft e.V. *Unverbindlicher Muster-Fragebogen zur Risikoerfassung im Rahmen von Cyber-Versicherungen für kleine und mittelständische Unternehmen. (Unverbindliche Bekanntgabe des Gesamtverbandes der Deutschen Versicherungswirtschaft e.V. (GDV) zur fakultativen Verwendung. Abweichende Vereinbarungen sind möglich.)*, 2019.
- [20] Herath, H. and Herath, T. Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies*, 2(1), 2011.
- [21] J. Kingman. *Poisson processes*, volume 3 of *Oxford studies in probability*. Clarendon Press, Oxford, 1993.
- [22] Mai, J. and Scherer, M. *Simulating copulas: Stochastic models, sampling algorithms, and applications*, volume vol. 6 of *Series in quantitative finance*. World Scientific Publishing, New Jersey and London and Singapore, 2nd edition edition, 2017.
- [23] Marotta, A., Martinelli, F., Nanni, S., Orlando, A., and Yautsiukhin, A. Cyber-insurance survey. *Computer Science Review*, 24:35–61, 2017.
- [24] Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., and Sadhukhan, S. Cyber-risk decision models: To insure it or not? *Decision Support Systems*, 56:11–26, 2013.
- [25] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*, volume 389. Wiley, 2002.
- [26] Munich Re. Cyber insurance: Risks and trends 2020, 2020.
- [27] H. Panjer. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin: The Journal of the IAA*, 12(1):22–26, 1981.
- [28] Peng, C., Xu, M., Xu, S., and Hu, T. Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, 44(14):2534–2563, 2017.
- [29] Peng, C., Xu, M., Xu, S., and Hu, T. Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, 45(15):2718–2740, 2018.
- [30] S. Romanosky. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, 2(2):121–135, 2016.
- [31] Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. Content analysis of cyber insurance policies: how do carriers price cyber risk? *Journal of Cybersecurity*, 5(1):117, 2019.
- [32] B. Schneier. The story behind the Stuxnet virus. *Forbes*, 07.10.2010.
- [33] tenable. tenable’s 2021 threat landscape retrospective. *tenable Research*, 2021.
- [34] W. Turton and J. Robertson. Microsoft attack blamed on china morphs into global crisis. *Bloomberg*, 07.03.2021.

- [35] Xu, M., Da, G., and Xu, S. Cyber epidemic models with dependences. *Internet Mathematics*, 11(1):62–92, 2015.
- [36] G. Zeller and M. Scherer. A comprehensive model for cyber risk based on marked point processes and its application to insurance. *European Actuarial Journal*, 12(1):33–85, 2022.

A Appendix A

Proof of Lemma 1

Proof of Lemma 1. Starting from Definition 1, we observe that the new marginal rates for any $\ell \in \{1, \dots, K\}$ are given by

$$\begin{aligned}
\tilde{\lambda}^\ell &= \sum_{i=1}^K \frac{i}{K} \tilde{\lambda}^{|I|=i} = \frac{1}{K} \lambda^{|I|=1} + \frac{1}{K} \left[\sum_{i=2}^K \lambda^{|I|=i} \left[i(f_{\text{Bin}}(0; i, p) + f_{\text{Bin}}(1; i, p)) \right. \right. \\
&\quad \left. \left. + \sum_{j=2}^{\max(i-1, 2)} (i-j) f_{\text{Bin}}(j; i, p) \right] \right] + \sum_{i=2}^K \frac{i}{K} \sum_{j=i}^K \lambda^{|I|=j} f_{\text{Bin}}(i; j, p) \\
&= \frac{1}{K} \lambda^{|I|=1} + \frac{1}{K} \underbrace{\left[\sum_{i=2}^K \lambda^{|I|=i} i (f_{\text{Bin}}(0; i, p) + f_{\text{Bin}}(1; i, p)) \right]}_{(S1)} \\
&\quad + \underbrace{\sum_{i=2}^K \lambda^{|I|=i} \sum_{j=2}^{\max(i-1, 2)} (i-j) f_{\text{Bin}}(j; i, p)}_{(S2)} + \underbrace{\sum_{i=2}^K i \sum_{j=i}^K \lambda^{|I|=j} f_{\text{Bin}}(i; j, p)}_{(S3)}.
\end{aligned}$$

It remains to show that the sum in the square bracket equals $\sum_{j=2}^K j \lambda^{|I|=j}$. Reversing the order of summation in (S3) and renaming $i \leftrightarrow j$ in the remaining terms yields

$$\begin{aligned}
\left[(S1) + (S2) + (S3) \right] &= \sum_{j=2}^K \lambda^{|I|=j} j (f_{\text{Bin}}(0; j, p) + f_{\text{Bin}}(1; j, p)) \\
&+ \sum_{j=2}^K \lambda^{|I|=j} \sum_{i=2}^{\max(j-1, 2)} (j-i) f_{\text{Bin}}(i; j, p) + \sum_{j=2}^K \lambda^{|I|=j} \sum_{i=2}^j i f_{\text{Bin}}(i; j, p) \\
&= \sum_{j=2}^K \lambda^{|I|=j} j (f_{\text{Bin}}(0; j, p) + f_{\text{Bin}}(1; j, p)) + \sum_{j=2}^K \lambda^{|I|=j} \left(\sum_{i=2}^{j-1} j f_{\text{Bin}}(i; j, p) + j f_{\text{Bin}}(j; j, p) \right) \\
&= \sum_{j=2}^K j \lambda^{|I|=j} \underbrace{\sum_{i=0}^j f_{\text{Bin}}(i; j, p)}_{=1} = \sum_{j=2}^K j \lambda^{|I|=j}.
\end{aligned}$$

□

Proof of Remark 3

Proof of Remark 3. Note that due to the properties of the Binomial coefficient, it holds that

$$\begin{aligned}\frac{\binom{K-1}{i}}{\binom{K}{i+1}} &= \frac{i+1}{K}, \\ \frac{\binom{K-2}{i}}{\binom{K}{i+1}} &= \frac{\frac{K-1-i}{K-1} \binom{K-1}{i}}{\binom{K}{i+1}} = \frac{K-1-i}{K-1} \cdot \frac{i+1}{K} = \frac{(K-(i+1))(i+1)}{K(K-1)}.\end{aligned}$$

Inserting this into the expression in (7) yields

$$\begin{aligned}\alpha &= 1 - \frac{\sum_{i=0}^{K-2} \binom{K-2}{i} \frac{1}{\binom{K}{i+1}} \lambda^{|I|=i+1}}{\sum_{i=0}^{K-1} \binom{K-1}{i} \frac{1}{\binom{K}{i+1}} \lambda^{|I|=i+1}} = 1 - \frac{\frac{1}{K(K-1)} \bar{\lambda} \sum_{i=0}^{K-2} (K-(i+1))(i+1)}{\frac{1}{K} \bar{\lambda} \sum_{i=0}^{K-1} (i+1)} \\ &= 1 - \frac{1}{K-1} \frac{\sum_{i=1}^{K-1} (K-i)i}{\sum_{i=1}^K i} = 1 - \frac{1}{K-1} \frac{\frac{1}{6} K(K+1)(K-1)}{\frac{1}{2} K(K+1)} = 1 - \frac{2}{6} = \frac{2}{3}.\end{aligned}$$

For the marginal rates λ^i in (2), it holds

$$\lambda^i = \sum_{k=1}^K \frac{k}{K} \lambda^{|I|=k} = \frac{\bar{\lambda}}{K} \sum_{k=1}^K k = \frac{\bar{\lambda}}{K} \frac{K(K+1)}{2} = \frac{\bar{\lambda}(K+1)}{2},$$

implying the remark. □

Proof of Lemma 2

Proof of Lemma 2. By definition, α and $\tilde{\alpha}$ are given by

$$\alpha = 1 - \frac{\sum_{i=1}^{K-1} \binom{K-2}{i-1} \lambda_i}{\sum_{i=1}^K \binom{K-1}{i-1} \lambda_i} =: 1 - \frac{Z_\alpha}{N_\alpha}, \quad \tilde{\alpha} = 1 - \frac{\sum_{i=1}^{K-1} \binom{K-2}{i-1} \tilde{\lambda}_i}{\sum_{i=1}^K \binom{K-1}{i-1} \tilde{\lambda}_i} =: 1 - \frac{Z_{\tilde{\alpha}}}{N_{\tilde{\alpha}}},$$

where $\lambda_i = \frac{\lambda^{|I|=i}}{\binom{K}{i}}$ and $\tilde{\lambda}_i = \frac{\tilde{\lambda}^{|I|=i}}{\binom{K}{i}}$.

We use the following properties of the Binomial coefficient and the Binomial distribution

$$\binom{K-1}{i-1} = \frac{i}{K} \binom{K}{i}, \tag{BIN1}$$

$$\binom{K-2}{i-1} = \frac{K-i}{K-1} \binom{K-1}{i-1} = \frac{K-i}{K-1} \frac{i}{K} \binom{K}{i}, \tag{BIN2}$$

$$\binom{K-2}{i-2} = \binom{K-1}{i-1} - \binom{K-2}{i-1}. \tag{BIN3}$$

$$X \sim \text{Binom}(K, p) \implies \mathbb{E}[X] = Kp, \tag{BIN4}$$

$$X \sim \text{Binom}(K, p) \implies \mathbb{E}[X^2] = Kp(1-p) + K^2p^2. \tag{BIN5}$$

This implies the following auxiliary result:

$$N_\alpha - Z_\alpha = \sum_{i=1}^K \binom{K-1}{i-1} \lambda_i - \sum_{i=1}^{K-1} \binom{K-2}{i-1} \lambda_i \stackrel{\text{(BIN3)}}{=} \lambda_K + \sum_{i=2}^{K-1} \binom{K-2}{i-2} \lambda_i = \sum_{i=2}^K \binom{K-2}{i-2} \lambda_i. \quad (12)$$

Furthermore, it holds that $N_\alpha = N_{\tilde{\alpha}}$, as

$$N_{\tilde{\alpha}} = \sum_{i=1}^K \binom{K-1}{i-1} \tilde{\lambda}_i = \sum_{i=1}^K \binom{K-1}{i-1} \frac{\tilde{\lambda}^{|I|=i}}{\binom{K}{i}} \stackrel{\text{(BIN1)}}{=} \sum_{i=1}^K \frac{i}{K} \tilde{\lambda}^{|I|=i} \stackrel{\text{Lemma 1}}{=} \sum_{i=1}^K \frac{i}{K} \lambda^{|I|=i} = N_\alpha. \quad (13)$$

We will show that for $Z_{\tilde{\alpha}}$ it holds that

$$Z_{\tilde{\alpha}} = \lambda_1 + \sum_{i=2}^K \lambda_i \left[\binom{K-1}{i-1} - \binom{K-2}{i-2} p^2 \right]. \quad (*)$$

This implies the claim, as one can rewrite

$$\begin{aligned} Z_{\tilde{\alpha}} &= \lambda_1 + \sum_{i=2}^K \lambda_i \left[\binom{K-1}{i-1} - \binom{K-2}{i-2} p^2 \right] = \sum_{i=1}^K \binom{K-1}{i-1} \lambda_i - p^2 \sum_{i=2}^K \binom{K-2}{i-2} \lambda_i \\ &\stackrel{(12),(13)}{=} N_\alpha - p^2 (N_\alpha - Z_\alpha). \end{aligned} \quad (14)$$

From this it follows

$$\tilde{\alpha} = 1 - \frac{Z_{\tilde{\alpha}}}{N_{\tilde{\alpha}}} \stackrel{(13),(14)}{=} 1 - \frac{N_\alpha - p^2(N_\alpha - Z_\alpha)}{N_\alpha} = 1 - \left[1 - p^2 \left(1 - \frac{Z_\alpha}{N_\alpha} \right) \right] = p^2 \alpha.$$

To show (*), we rewrite (3) as

$$\begin{aligned} \tilde{\lambda}^{|I|=1} &= \lambda^{|I|=1} + \sum_{i=2}^K \lambda^{|I|=i} \left[i(f_{\text{Bin}}(0; i, p) + f_{\text{Bin}}(1; i, p)) + \sum_{j=2}^{\max(i-1, 2)} (i-j) f_{\text{Bin}}(j; i, p) \right] \\ &= \lambda^{|I|=1} + \lambda^{|I|=2} 2 \underbrace{(f_{\text{Bin}}(0; 2, p) + f_{\text{Bin}}(1; 2, p))}_{(1-p^2)} + \sum_{i=3}^K \lambda^{|I|=i} \left[i \underbrace{\sum_{j=0}^{i-1} f_{\text{Bin}}(j; i, p)}_{i(1-p^i)} - \underbrace{\sum_{j=2}^{i-1} j f_{\text{Bin}}(j; i, p)}_{\pm \sum_{j=0,1,i} j f_{\text{Bin}}(j; i, p)} \right] \\ &\stackrel{\text{(BIN4)}}{=} \lambda^{|I|=1} + \lambda^{|I|=2} 2(1-p^2) + \sum_{i=3}^K \lambda^{|I|=i} [i - ip^i - (ip - ip(1-p)^{i-1} - ip^i)] \\ &= \lambda^{|I|=1} + \lambda^{|I|=2} 2(1-p^2) + \sum_{i=3}^K \lambda^{|I|=i} i(1-p + p(1-p)^{i-1}) \\ &= \lambda^{|I|=1} + \sum_{i=2}^K \lambda^{|I|=i} i(1-p + p(1-p)^{i-1}). \end{aligned}$$

Changing to the rates $\tilde{\lambda}_1 = \frac{\tilde{\lambda}^{|\cdot|=1}}{K}$ (LHS) and $\lambda_i = \frac{\lambda^{|\cdot|=i}}{\binom{K}{i}}$ (RHS) yields

$$\tilde{\lambda}_1 = \lambda_1 + \sum_{i=2}^K \frac{i}{K} \binom{K}{i} \lambda_i (1-p + p(1-p)^{i-1}) \stackrel{(BIN1)}{=} \lambda_1 + \sum_{i=2}^K \binom{K-1}{i-1} \lambda_i (1-p + p(1-p)^{i-1}),$$

i.e. for fixed $i \in \{2, \dots, K\}$, the coefficient of λ_i from $\tilde{\lambda}_1$, which appears in $Z_{\tilde{\alpha}}$ with factor $\binom{K-2}{0} = 1$ is given by $\binom{K-1}{i-1} (1-p + p(1-p)^{i-1})$. Analogously, the coefficients of λ_i from $\sum_{j=2}^{K-1} \tilde{\lambda}_j$, scaled by $\binom{K-2}{j-1}$, are illustrated as the column sums in Table 3 and given by

$$\begin{aligned} & \underbrace{\lambda_i \binom{K}{i}}_{\lambda_i \binom{K}{i} = \lambda^{|\cdot|=i}} \sum_{j=2}^i \binom{i}{j} p^j (1-p)^{i-j} \underbrace{\binom{K-2}{j-1}}_{\text{Def. of } Z_{\alpha}} \underbrace{\frac{1}{\binom{K}{j}}}_{\lambda_j = \frac{\lambda^{|\cdot|=j}}{\binom{K}{j}}} \\ & \stackrel{(BIN2)}{=} \lambda_i \binom{K}{i} \sum_{j=2}^i \frac{K-j}{K-1} \frac{j}{K} f_{\text{Bin}}(j; i, p) = \frac{\lambda_i \binom{K}{i}}{(K-1)K} \left[K \sum_{j=2}^i j f_{\text{Bin}}(j; i, p) - \sum_{j=2}^i j^2 f_{\text{Bin}}(j; i, p) \right] \\ & \stackrel{(BIN4), (BIN5)}{=} \frac{\lambda_i \binom{K}{i}}{(K-1)K} \left[K(ip - ip(1-p)^{i-1}) - [ip(1-p) + i^2 p^2 - ip(1-p)^{i-1}] \right] \\ & = \frac{\lambda_i \binom{K}{i}}{(K-1)K} \left[Kip - Kip(1-p)^{i-1} - ip + ip^2 - i^2 p^2 + ip(1-p)^{i-1} \right] \\ & = \frac{\lambda_i \binom{K}{i}}{(K-1)K} \left[(K-1)ip - (K-1)ip(1-p)^{i-1} - (i-1)ip^2 \right] \\ & = \lambda_i \left[\binom{K}{i} \frac{i}{K} p - \binom{K}{i} \frac{i}{K} p(1-p)^{i-1} - \binom{K}{i} \frac{i(i-1)}{K(K-1)} p^2 \right] \\ & \stackrel{(BIN1)}{=} \lambda_i \left[\binom{K-1}{i-1} (p - p(1-p)^{i-1}) - \binom{K-2}{i-2} p^2 \right]. \end{aligned}$$

Thus, adding the coefficients of λ_i from $\tilde{\lambda}_1$ and $\sum_{j=2}^K \tilde{\lambda}_j \binom{K-2}{j-1}$ for each fixed $i \in \{2, \dots, K-1\}$ yields

$$\binom{K-1}{i-1} (1-p + p(1-p)^{i-1}) + \binom{K-1}{i-1} (p - p(1-p)^{i-1}) - \binom{K-2}{i-2} p^2 = \binom{K-1}{i-1} - \binom{K-2}{i-2} p^2,$$

which implies (*) and therefore the claim. \square

Proof of Theorem 1

Proof of Theorem 1.

Step 1: Increasing convex order for some discrete random variables

For an integer $K > 0$, consider a Bernoulli r.v. $Z \sim \text{Ber}(p)$, $p \in [0, 1]$ and K i.i.d. copies of it denoted Z_i , $i \in \{1, \dots, K\}$.

Furthermore, consider the r.v.s X and Y defined as follows:

$$\begin{aligned} X &= K Z, \\ Y &= \sum_{i=1}^K k_i Z_i, \quad i \in \{1, \dots, K\}, \end{aligned} \tag{15}$$

where $\mathbf{k} := (k_i)_{i \in \{1, \dots, K\}}$ is an \mathbb{N}_0^K -vector s.t. $\forall i : k_i \in \{0, \dots, K\}$ with $\sum_{i=1}^K k_i = \|\mathbf{k}\|_1 = K$. Assume w.l.o.g. $k_i \geq k_{i+1}$, $\forall i \in \{1, \dots, K-1\}$, and let $i^* := |\{k_i : k_i > 0\}|$, then the first i^* entries of \mathbf{k} represent a *partition* of K (and the remaining entries equal 0).

It is obvious that for any r.v. Y as above

$$\mathbb{E}[Y] = \mathbb{E}[X] = Kp,$$

and we will now show that for any such Y it holds that

$$Y \leq_{icx} X$$

by using the following sufficient condition (the so-called *cut criterion* or *crossing condition*, see e.g. [25], p. 23): If for two r.v.s X and Y with c.d.f.s F_X and F_Y respectively, it holds that $\mathbb{E}[Y] \leq \mathbb{E}[X]$ and in addition, there exists $t_0 \in \mathbb{R}$ s.t.

$$\begin{aligned} F_Y(t) &\leq F_X(t) \quad \forall t < t_0, \\ F_Y(t) &\geq F_X(t) \quad \forall t \geq t_0, \end{aligned} \tag{16}$$

then this implies $Y \leq_{icx} X$.

Let us in the following exclude the trivial cases $p \in \{0, 1\}$ and $\mathbf{k} = (K, 0, \dots, 0)$ as they lead to $F_X = F_Y$. Note that in all non-degenerate cases we have $i^* > 1$.

Then, for r.v.s X and Y as defined in (15), there exists $t_0 \in [1, K-1]$ s.t. the single-crossing condition is fulfilled:

For $t < 0$ and $t \geq K$, obviously $F_X(t) = F_Y(t)$.

For $t \in [0, 1)$, we use that $p \in (0, 1)$ and $i^* > 1$ to see

$$F_Y(t) = \mathbb{P}(Y = 0) = (1-p)^{i^*} < 1-p = \mathbb{P}(X = 0) = F_X(t).$$

For $t \in (K-1, K)$, again with $p \in (0, 1)$ and $i^* > 1$,

$$F_Y(t) = \mathbb{P}(Y \leq K-1) = 1 - \mathbb{P}(Y = K) = 1 - p^{i^*} > 1-p = \mathbb{P}(X = 0) = F_X(t).$$

Lastly, note that

- $t \mapsto F_X(t)$ is constant for $t \in (0, K-1]$ at the level $F_X(t) \equiv 1-p$.
- $F_Y(t)$ is monotone increasing (being a c.d.f.) for $t \in (0, K-1]$ with (non-negative) jumps at some of the $\{1, \dots, K-1\}$ and $F_Y(0+) = (1-p)^{i^*} < 1-p < 1-p^{i^*} = F_Y(K-1)$.

Thus, due to the monotonicity of F_Y , there must be a *unique* $t_0 \in [1, K-1]$ fulfilling (16).

Step 2: Implication for (compound) Poisson process setting

Now, fix a time horizon $T > 0$ and consider two independent homogeneous Poisson processes $N(t) := (N(t))_{t \geq 0}$ with rate $\lambda > 0$ and $\tilde{N}(t) := (\tilde{N}(t))_{t \geq 0}$ with rate $\ell\lambda > 0$, $\ell \in \mathbb{N}$. As $\tilde{N}(t)$ can be understood (in the sense of being equal in distribution) as the superposition of ℓ independent Poisson processes $\tilde{N}_j(t)$, $j \in \{1, \dots, \ell\}$, all of them with rate $\lambda > 0$ (see e.g. [21], p. 16), one can write $S(T)$ and $\tilde{S}(T)$ as

$$\begin{aligned} S(T) &= \sum_{i=1}^{N(T)} L = L N(T), \\ \tilde{S}(T) &= \sum_{i=1}^{\tilde{N}(T)} \frac{L}{\ell} \stackrel{D}{=} \frac{L}{\ell} \sum_{j=1}^{\ell} \tilde{N}_j(T). \end{aligned}$$

Due to the properties of the homogeneous Poisson process and by Wald's equation, it follows immediately that

$$\begin{aligned} N(T), \tilde{N}_j(T) &\sim \text{Poi}(\lambda T), \quad j \in \{1, \dots, \ell\}, \\ \mathbb{E}[N(T)] &= \mathbb{E}[\tilde{N}_j(T)] = \lambda T, \quad j \in \{1, \dots, \ell\}, \\ \mathbb{E}[S(T)] &= \mathbb{E}[\tilde{S}(T)] = \lambda T L, \end{aligned}$$

where $\text{Poi}(\lambda)$ denotes the Poisson distribution with density $f_{\text{Poi}(\lambda)}(k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $k \in \mathbb{N}_0$, $\lambda > 0$. Now, consider the following random variables:

$$\begin{aligned} X^i &= L \mathbf{1}_{\{N(T) \geq i\}} = \begin{cases} L & \text{if } N(T) \geq i, \\ 0 & \text{else,} \end{cases} \implies X^i = \begin{cases} L & \text{w.p. } 1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j), \\ 0 & \text{w.p. } \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j), \end{cases} \\ Y_j^i &= \frac{L}{\ell} \mathbf{1}_{\{\tilde{N}_j(T) \geq i\}} = \begin{cases} \frac{L}{\ell} & \text{if } \tilde{N}_j(T) \geq i, \\ 0 & \text{else,} \end{cases} \implies Y_j^i = \begin{cases} \frac{L}{\ell} & \text{w.p. } 1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j), \\ 0 & \text{w.p. } \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j), \end{cases} \quad j \in \{1, \dots, \ell\}. \end{aligned}$$

Note that X^i denotes the size of the i^{th} jump of the Poisson process $N(t)$ if it occurs until time T (of deterministic size $L > 0$ if the process jumps at least i times until time T , and of size 0 else), and analogously the ℓ independent random variables Y_j^i denote the sizes of the i^{th} jump of each of the independent Poisson processes $\tilde{N}_j(t)$ if they occur until time T . As the Y_j^i , $j \in \{1, \dots, \ell\}$, are independent, one can derive the density of their sum, denoted Y^i , from arguments borrowed from the Binomial law:

$$Y^i := \sum_{j=1}^{\ell} Y_j^i = \begin{cases} L & \text{w.p. } \left(1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)\right)^\ell, \\ \frac{\ell-1}{\ell} L & \text{w.p. } \binom{\ell}{\ell-1} \left(1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)\right)^{\ell-1} \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j), \\ \dots & \\ \frac{1}{\ell} L & \text{w.p. } \binom{\ell}{1} \left(1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)\right) \left(\sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)\right)^{\ell-1}, \\ 0 & \text{w.p. } \left(\sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)\right)^\ell. \end{cases}$$

Note that this illustrates the fundamental difference between the two considered cases (process $N(t)$ vs. superposition of ℓ processes $\tilde{N}_j(t)$): In the notation of a collective risk model, if the claim occurrences are driven by the process $N(t)$ (corresponding to relatively few events) and claim sizes are relatively large (i.e. of size L), either a large total claims amount occurs

or no claim at all occurs for each jump. On the contrary, if claim occurrences are driven by the independent processes $\tilde{N}_j(t)$ or equivalently their superposition $\tilde{N}(t)$ (relatively many events) and claim sizes are relatively small (i.e. of size $\frac{L}{\ell}$), for a large total claims amount of size L from all the first (second, third, ...) jumps to occur, all ℓ processes $\tilde{N}_j(t)$ independently need to jump at least once (twice, three times, ...); equivalently, ℓ independent jumps need to occur before time T in the superimposed process $\tilde{N}(t)$. Likewise, to obtain no claim at all from the i^{th} jumps, any of the processes $\tilde{N}_j(t)$ independently must not jump more than $(i-1)$ times; or equivalently, the superimposed process may not jump more than $(i-1)\ell$ times until T . Therefore, the probability of both large (i.e. size L) and no (size 0) total claims amounts is reduced, and probability mass is shifted to the intermediate cases that some (but not all or none) of the independent processes observe at least i jumps. As

$$\mathbb{E}[X^i] = \mathbb{E}[Y^i] = L \left(1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j) \right)$$

– note that the weights for Y^i are akin to the density of a Binomial distribution with $N = \ell$, $p = 1 - \sum_{j=0}^{i-1} f_{\text{Poi}(\lambda T)}(j)$ – for any $i \in \mathbb{N}$ the discrete random variables X^i and Y^i are akin to X and Y from the first part of the proof, X^i being a Bernoulli r.v. with positive mass only on the largest admissible value L and Y^i following a discrete density supported on the set of values $\{0, \frac{L}{\ell}, \dots, \frac{(\ell-1)L}{\ell}, L\}$ with equal expectation. It follows from the above derivations that $X^i \geq_{icx} Y^i$, $i \in \mathbb{N}$. As (increasing) convex order is preserved under summation (this follows immediately from the transitivity of \leq_{icx}), this implies the statement of the theorem as

$$S(T) = \sum_{i \in \mathbb{N}} X^i \geq_{icx} \sum_{i \in \mathbb{N}} Y^i = \tilde{S}(T).$$

Note that it is straightforward to again extend the result to a case where not all deterministic jump sizes corresponding to the ℓ arrival processes $\tilde{N}_\ell(t)$ are equally of size $\frac{L}{\ell}$, but instead one replaces them by a collection $\{L_i\}_{i \in \{1, \dots, \ell\}}$, such that $L_i > 0, \forall i \in \{1, \dots, \ell\}$, and $\sum L_i = L$. \square

Proof of Corollary 1

Proof of Corollary 1. It is a well-known result that for any two integrable r.v. X and Y , convex order is equivalent to the ordering of expected shortfall at all levels q , i.e.

$$Y \leq_{cx} X \iff \mathbf{ES}_q(Y) \leq \mathbf{ES}_q(X), \quad \forall q \in (0, 1),$$

see e.g. [16] and the references therein. Therefore, the statement of the corollary is equivalent to showing $S_M(T) \geq_{cx} S_{\tilde{M}}(T)$. As from Lemma 1 (and linearity) it follows that $\mathbb{E}[S_M(T)] = \mathbb{E}[S_{\tilde{M}}(T)]$, it is sufficient to show $S_M(T) \geq_{icx} S_{\tilde{M}}(T)$ (see first point of Remark 4).

This follows immediately from Theorem 1: Recall that in model (M) , the arrival rates for events of size $k \in \{1, \dots, K\}$ are given by $\boldsymbol{\lambda} := (\lambda^{|I|=1}, \dots, \lambda^{|I|=K})$ and that all arrivals are independent (from arrivals of events of the same or any other size). The total incident

number until time T can therefore again be written as a sum of K independent compound Poisson r.v.s:

$$S_M(T) \stackrel{D}{=} \sum_{k=1}^K kN_k(T),$$

where $N_k(t) := (N_k(t))_{t \geq 0}$, $k \in \{1, \dots, K\}$, are independent Poisson processes with rates $\lambda^{|I|=k}$. In turn, for any k , the process $N_k(t)$ can (artificially) be understood as the superposition of $(k+1)$ independent Poisson processes $N_{k,j}(t) := (N_{k,j}(t))_{t \geq 0}$, $j \in \{0, \dots, k\}$, with rates $\lambda^{|I|=k} f_{\text{Bin}}(j; k, p)$, where in model (M) each of the arrivals of each of these processes is associated with a jump of size k .

Then, the total incident number from events of size k until time $T > 0$, denoted $S_k(T)$, and events of all sizes, denoted $S(T)$, are given by the following compound Poisson r.v.s, respectively:

$$S_k(T) = \sum_{j=0}^k kN_{k,j}(T) \stackrel{D}{=} kN_k(T), \quad S(T) = \sum_{k=1}^K S_k(T) = \sum_{k=1}^K \sum_{j=0}^k kN_{k,j}(T),$$

where $\stackrel{D}{=}$ denotes equality in distribution.

In model (M) , for any $k \in \{1, \dots, K\}$ each Poisson arrival process $N_{j,k}(t)$ is associated with jumps of size k . In model (\widetilde{M}) , each arrival process $N_{j,k}(t)$ is replaced by several independent processes with equal Poisson rate, but associated with smaller jump sizes (which sum up to k), as represented in Table 4. From Theorem 1, it follows immediately that the compound incident number in the second column (model (M)) of each row dominates in increasing convex order the compound incident number of the corresponding processes in the third column (model (\widetilde{M})). By summing over all rows (recall that \leq_{icx} is preserved under summation), the same holds for the compound incident number from each process $N_k(t)$, $k \in \{1, \dots, K\}$, in model (M) as compared to the overall compound incident number from all the corresponding independent processes in model (\widetilde{M}) . By summing over all $k \in \{1, \dots, K\}$, it follows that $S_M(T) \geq_{icx} S_{\widetilde{M}}(T)$ for any fixed $T > 0$ and thus the statement of the corollary. \square

	λ_2	λ_3	\dots	λ_{K-2}	λ_{K-1}	λ_K	$\binom{K}{K} \lambda_K p^K \frac{\binom{K}{K}}{\binom{K}{K}}$	$\binom{K-2}{j-1}$
$\tilde{\lambda}_K$								
$\tilde{\lambda}_{K-1}$					$\binom{K-1}{K-1} \lambda_{K-1} p^{K-1} \frac{\binom{K}{K-1}}{\binom{K}{K-1}} +$	$\binom{K}{K-1} \lambda_K p^{K-1} (1-p) \frac{\binom{K}{K}}{\binom{K}{K-1}}$		$\binom{K-2}{K-2}$
$\tilde{\lambda}_{K-2}$				$\binom{K-2}{K-2} \lambda_{K-2} p^{K-2} \frac{\binom{K}{K-2}}{\binom{K}{K-2}} +$	$\binom{K-1}{K-2} \lambda_{K-1} p^{K-2} (1-p) \frac{\binom{K}{K-1}}{\binom{K}{K-2}} +$	$\binom{K}{K-2} \lambda_K p^{K-2} (1-p)^2 \frac{\binom{K}{K}}{\binom{K}{K-2}}$		$\binom{K-2}{K-3}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
$\tilde{\lambda}_3$		$\binom{3}{3} \lambda_3 p^3 \frac{\binom{K}{3}}{\binom{K}{3}} +$	$\dots +$	$\binom{K-2}{3} \lambda_{K-2} p^3 (1-p)^{(K-2)-3} \frac{\binom{K}{K-2}}{\binom{K}{3}} +$	$\binom{K-1}{3} \lambda_{K-1} p^3 (1-p)^{(K-1)-3} \frac{\binom{K}{K-1}}{\binom{K}{3}} +$	$\binom{K}{3} \lambda_K p^3 (1-p)^{K-3} \frac{\binom{K}{K}}{\binom{K}{3}}$		$\binom{K-2}{2}$
$\tilde{\lambda}_2$	$\binom{2}{2} \lambda_2 p^2 \frac{\binom{K}{2}}{\binom{K}{2}} +$	$\binom{3}{2} \lambda_3 p^2 (1-p) \frac{\binom{K}{3}}{\binom{K}{2}} +$	$\dots +$	$\binom{K-2}{2} \lambda_{K-2} p^2 (1-p)^{(K-2)-2} \frac{\binom{K}{K-2}}{\binom{K}{2}} +$	$\binom{K-1}{2} \lambda_{K-1} p^2 (1-p)^{(K-1)-2} \frac{\binom{K}{K-1}}{\binom{K}{2}} +$	$\binom{K}{2} \lambda_K p^2 (1-p)^{K-2} \frac{\binom{K}{K}}{\binom{K}{2}}$		$\binom{K-2}{1}$

Table 3: The table illustrates the calculation of the coefficient of each λ_i , $i \in \{2, \dots, K-1\}$, in $Z_{\tilde{\alpha}}$. In each row, $\tilde{\lambda}_j$ is calculated based on the definition of model (\tilde{M}) given in (4), where $\tilde{\lambda}_j = \frac{\tilde{\lambda}^{|I|=j}}{\binom{K}{j}}$ and $\lambda_i = \frac{\lambda^{|I|=i}}{\binom{K}{i}}$ are substituted (leading to the last fraction of Binomial coefficients in each entry). The coefficient of each λ_i , $i \in \{2, \dots, K-1\}$, in $Z_{\tilde{\alpha}}$ is given by the scalar product of the i^{th} column and the very last column which lists the factors $\binom{K-2}{j-1}$ from the definition of $Z_{\tilde{\alpha}}$.

	Model (M): (Poisson rate, jump size)	Model (\widetilde{M}): (Poisson rate, jump size)	Interpretation
$N_{k,k}(T)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k; k, p), k)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k; k, p), k)$	All k joint arrivals recognized \implies process with jump size k is “replaced by” process with jump size k .
$N_{k,k-1}(T)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-1; k, p), k)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-1; k, p), k-1)$ $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-1; k, p), 1)$	$k-1$ joint arrivals recognized \implies process with jump size k is replaced by two independent processes with jump sizes $k-1$ and 1, respectively.
$N_{k,k-2}(T)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-2; k, p), k)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-2; k, p), k-2)$ $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-2; k, p), 1)$ $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(k-2; k, p), 1)$	$k-2$ joint arrivals recognized \implies process with jump size k is replaced by three independent processes with jump sizes $k-2$, 1, and 1, respectively.
\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots
$N_{k,2}(T)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(2; k, p), k)$	$(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(2; k, p), k-2)$ $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(2; k, p), 1)$ $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(2; k, p), 1)$ \dots $(\lambda^{ \mathcal{I} =k} T f_{\text{Bin}}(2; k, p), 1)$	2 joint arrivals recognized \implies process with jump size k is replaced by independent processes with jump sizes 2 (one process) and 1 ($k-2$ processes), respectively.
$N_{k,1}(T) + N_{k,0}(T)$	$(\lambda^{ \mathcal{I} =k} T (f_{\text{Bin}}(1; k, p) + f_{\text{Bin}}(0; k, p)), k)$	$(\lambda^{ \mathcal{I} =k} T (f_{\text{Bin}}(1; k, p) + f_{\text{Bin}}(0; k, p)), 1)$ \dots \dots $(\lambda^{ \mathcal{I} =k} T (f_{\text{Bin}}(1; k, p) + f_{\text{Bin}}(0; k, p)), 1)$	1 or 0 joint arrivals recognized \implies process with jump size k is replaced by k independent processes, each with jump size 1.

Table 4: Comparison of the compound Poisson processes corresponding to models (M) and (\widetilde{M}) for any fixed $k \in \{1, \dots, K\}$: The arrival process $N_k(t)$ for events of size k , i.e. associated with jumps of size k , can be replaced by $(k+1)$ independent processes with thinned rates according to the weights of a Binomial distribution. According to Definition 1, in model (\widetilde{M}), these processes are replaced by several independent processes, associated with smaller jump sizes adding up to k .