Aggregate Cyber-Risk Management in the IoT Age *Cautionary Statistics for (Re)Insurers and Likes*

Ranjan Pal, *Member, IEEE,* Ziyuan Huang, *Student Member, IEEE,* Xinlong Yin, *Student Member, IEEE,* Sergey Lototsky, *Member, IEEE* Swades De, *Senior Member, IEEE* Sasu Tarkoma, *Senior Member, IEEE* Mingyan Liu, *Fellow, IEEE,* Jon Crowcroft, *Fellow, IEEE,* and Nishanth Sastry, *Senior Member, IEEE*

Abstract*—IoT-driven smart societies are modern service-networked ecosystems, whose proper functioning is hugely based on the success of supply chain relationships. Robust security is still a big challenge in such ecosystems, catalyzed primarily by naive cyber-security practices (e.g., setting default IoT device passwords) on behalf of the ecosystem managers, i.e., users and organizations. This has recently led to some catastrophic malware-driven DDoS and ransomware attacks (e.g., the *Mirai* and *WannaCry* attacks). Consequently, markets for commercial third party cyber-risk management services (e.g., cyber-insurance) are steadily but sluggishly gaining traction with the rapid increase of IoT deployment in society, and provides a channel for ecosystem managers to transfer residual cyber-risk post attack events. Current empirical studies have shown that such residual cyber-risks affecting smart societies are often *heavy-tailed* in nature and exhibit *tail dependencies*. This is both, a major concern for a profit-minded cyber-risk management firm that might normally need to cover multiple such dependent cyber-risks from different sectors (e.g., manufacturing, energy) in a service-networked ecosystem, and a good intuition behind the sluggish market growth of cyber-risk management products. In this paper, we provide (i) a rigorous general theory to elicit conditions on (tail-dependent) heavy-tailed cyber-risk distributions under which a risk management firm might find it (non)sustainable to provide aggregate cyber-risk coverage services for smart societies, and (ii) a real-data driven numerical study to validate claims made in theory assuming *boundedly rational* cyber-risk managers, alongside providing ideas to boost markets that aggregate dependent cyber-risks with heavy-tails. *To the best of our knowledge, this is the only complete general theory till date on the feasibility of aggregate cyber-risk management.

***Index Terms*—aggregate cyber-risk, heavy-tail, tail-dependency**

I. Introduction

IoT-driven smart cities are examples of service networked ecosystems that are popularly on the rise around the globe, with major cities like Singapore, Dubai, Barcelona, and Amsterdam being working examples. The proper functioning of such cities is hugely based on the success of supply chain relationships from diverse sectors such as automobiles, electronics, energy, finance, aerospace, etc. In the IoT age, these relationships are often realized via large scale systemic network linkages (see Figure 1.1. in [1]) that operate via the interplay of IoT hardware (e.g., sensors, actuators, cameras), application software (e.g., Oracle for DBMS support, cloud service software), and IoT firmware.

Currently, robust IoT security is a challenge [2] with a significant fraction of users controlling IoT systems being naive about effective cyber-security practices (e.g., the use of non-default device passwords, periodic patch updates). Consequently a cyber-attack exploiting a software vulnerability can have a catastrophic cascading service disruption effect that could amount to losses in billions of dollars across various service sectors. Recent examples of such cyber-attacks include the *Mirai* DDoS (2016), *NotPetya* ransomware (2017), and *WannaCry* ransomware (2017) attacks, which wrecked havoc among firms in various industries across the globe, resulting in huge financial losses due to service interruption (see [1] for more examples). As a result of such large losses, a certain section of society overall could be negatively impacted and experience psychological depression and affected lifestyles.

As instruments to cover cyber-losses in society, markets for commercial third-party services (e.g., cyber-insurance) are steadily but sluggishly gaining traction with the rapid increase of societal IoT deployment, and provides a channel for members (individuals and organizations) to transfer residual cyber-risk post cyber-attack events. The primary benefits of commercial cyber-loss management services have been recently cited in detail by the authors in Biener et.al. [3], and include (i) indemnification of loss events, (ii) helping corporations estimate cost of cyber-risk, and (iii) improve cyber-security [4][5][6][7]. The steady rise in market requirement for such services primarily arises from a combination of (a) the naivety of user security practices, (b) the non fool-proof nature of technical security solutions to remove cyber-risk [8], (c) higher board level concerns in organizations post notable cyber-breach incidents (e.g., Sony, Target, WannaCry) and their negative effect on stock prices [9] [10], and (d) the growing perception of cyber-risk in the digital society [11].

Despite the promised potential for commercial cyber-risk management services, the markets have been too sluggish for our liking. The yearly estimates of cyber-loss approximately amount to USD 600 billion globally (1% of US GDP) [1], whereas the cumulative global public and private sector spendings on cyber-security amount only to USD 174 billion [12]. In addition, the total yearly market for cyber-insurance

R. Pal. M. Liu, Z. Huang, and X. Yin are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA. E-mail:{palr, mingyan, ziyuanh, connory}@umich.edu

S. Lototsky is with the Department of Mathematics, University of Southern California, USA. E-mail: lototsky@umich.edu

J. Crowcroft is with the Computer Laboratory, University of Cambridge, UK, and the Alan Turing Institute, UK, E-mail: jac22@cam.ac.uk

S. De is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, India. E-mail: swadesd@ee.iitd.ac.in

S. Tarkoma is with the Department of Computer Science, University of Helsinki, Finland. E-mail: sasu.tarkoma@helsinki.fi

N. Sastry is with the Department of Informatics, King's College London, UK. E-mail: nishanth.sastry@kcl.ac.uk

services - the most popular form of commercial third party commercial cyber-risk management offerings, approximates to a paltry USD 6 billion globally [12], compared to the amount of net cyber-loss. The primary reasons for such a low (but increasing) market penetration are (a) misunderstanding and lack of coverage awareness by the demand side (users and organizations) [12], (b) unavailability of quality plus quantity data on cyber-risks and demand side cyber-hygiene behavior, that contribute to policy pricing nuances [13] [14] [12], and (c) the empirical evidence of certain cyber-risk distributions being heavy-tailed and tail-dependent [3] [15] [16] [17], that makes profit-minded risk-averse cyber-insurers go low on confidence to expand coverage markets, where coverage is on an aggregate sum of such heavy-tailed cyber-risks.

A. Research Motivation

It is obvious that the ushering pervasive IoT age with 100s of IoT devices per home/organization will bring forth the need for businesses and homes to increasingly buy coverage CRM solutions like cyber-insurance. This is simply because the cyber-attack space will be broad enough in the digital terrain for humans to always prevent being security-hacked by smart adversaries. As a result, any coverage CRM solution provider will face aggregate cyber-risks from its clients. The idea of spreading aggregate cyber-risk among multiple risk managers (e.g., cyber (re)insurers) is gaining traction [1] [18] [19] for IoT-driven smart society settings whereby insurers covering aggregate cyber-risk of organizations in a given sector (e.g., manufacturing) wish to spread that risk among insurers of firms that are higher up in the supply chain (e.g., energy companies). However (a) there is **no formal analysis** on the effectiveness of this idea for *general individual cyber-risk distributions*, and (b) there may be significant differences in the cyber and non-cyber re-insurance settings - benefits of non-systemic outcomes in the latter (as qualitatively stated in [18]) may not apply to the former (see Section IV for more details). Consequently, without a formal analysis, aggregate cyber-risk managers may not have the confidence to scale their service markets [20]. *Our main goal in this paper is to devise a foundational methodology that analyzes the effect of individual heavy-tailed and tail-dependent cyber-risks on the effectiveness of aggregate cyber-risk management markets.*

B. Research Contributions

We make the following research contributions in this paper.

1) We prove that spreading *catastrophic* heavy-tailed cyber-risks that are identical and independently distributed (i.i.d.), i.e., not tail-dependent, *is not* an effective practice for aggregate cyber-risk managers. However, spreading i.i.d. heavy-tailed cyber-risks that are *not catastrophic* is an effective practice for aggregate cyber-risk managers. While this latter point has long been believed and empirically validated in the cyber-insurance research literature, the former point is a surprising new facet that we unravel in this paper via theory (see Section II).
2) We prove that spreading *catastrophic* and *curtailed* heavy-tailed cyber-risks that are (non) identical and independently distributed (i.i.d.), i.e., not tail-dependent, *is not* an effective practice for aggregate cyber-risk managers. (see Section III).
3) We show that spreading *catastrophic and tail-dependent* heavy-tailed cyber-risks *is not* an effective practice for aggregate cyber-risk managers. Though this result has been empirically established in the past for some heavy-tailed distributions (and also somewhat intuitive from the results of Section II), *there exists no formal proof for general heavy-tailed cyber-risk distributions, leave alone catastrophic heavy-tailed distributions* (see Section IV).
4) We experimentally validate our theory using a real-world cyber-breach data set by (a) relaxing the constraint of dealing with *stable* heavy-tailed cyber-risk distributions (see online Appendix A for details) needed for tractable analyses (as in Sections II, III, IV), and (b) assuming risk managers to be boundedly, i.e., not be perfectly rational in interpreting the extent of cyber-risk, as is usual in practice (see Section V).

Our proposed research presents a foundational methodology to analyze the effectiveness of spreading catastrophic heavy-tailed and tail-dependent cyber-risks. *To the best of our knowledge, this is the only complete general theory till date on the feasibility of aggregate cyber-risk management, and is invariant of specific threat models that eventually induce cyber-risk distributions.* Though the empirical occurrence of catastrophic cyber-risks is uncommon, it is a matter of time we start encountering them relatively more frequently in the IoT age (see Chapters 1.2, 1.3 in [1]). *A basic primer of important statistical and econometric concepts used in the paper is provided in online Appendix A*, and a table of important notations in the paper is presented in Table I.

C. Contribution Impact on Society and Technology

Our research contributions stated thus far are primarily targeted towards the advancement in the economics and econometrics of cyber-risk management in the IoT age through the solution of open research issues - the main focus of our research. However, each of these contributions have a *direct impact* on IoT security improvement, and its consequent positive impact on society.

To start with, according to data sources, the global number of connected devices has already reached 22 billion at the end of 2018 - more than half of which belong to enterprise IoT [21], and will grow to 29 billion by 2022 [22]. Moreover, worldwide spending on IoT is projected to reach a significant 1.2 trillion USD by 2022 with the number of Internet-connected devices being projected to reach a whopping 125 billion by 2030 [23]. *A thing common to nearly all IoT devices is the poor cyber-hygiene associated with their use (e.g., default passwords) - a primary reason being the scale of such devices in operation and the disproportionate human effort (that is likely to continue) needed to strengthen basic security in such devices [1].* This increasingly becoming common knowledge would push organizations and individual households to consider investing in third-party cyber-risk management (CRM) solutions as a necessary risk management step in the upcoming pervasive IoT age.

Contribution #1 states that cyber-risk "buyers" (i.e., the CRM firms) need to develop regulated pricing policies for their CRM solutions. These solutions will enable end-users to voluntarily (incentive compatibly) "look after" to a considerable degree, the security hygiene (and hence cyber-risk exposure) of IoT devices under their control. Consequently, such steps will prevent each end-user (individual household or organization) to be a source of a cyber-risk distribution that is heavy-tailed, i.e., catastrophic. This will allow CRM solution markets to scale and flourish, and improve cyber-security in society. **Contribution #2** reflects the same things for the CRM solution buyers as that from Contribution #1, but additionally warns the 'risk-buyer' side to put increasing focus on pricing policies that prevent IoT-controlled sources (organizations or individual households) to be a root of catastrophic cyber-risk distributions. The increased focus needed due to the fact that statistical curtailement of such cyber-risks (unlike that in Contribution #1) will also not allow CRM markets to scale and flourish - thereby having a negative effect on society as a whole. **Contribution #3** reflects similar learnings for both the CRM solution provider and the buyer sides, as that from Contributions#'s 1 and 2. **Contribution #4** clearly states that when CRM solution providers suffer from practical and subjective behavioral biases in appropriately assessing cyber-risk extent [1], it should not aggregate cyber-risk of catastrophic nature - thereby implying, similar to that in Contribution #'s 1-3, that solution pricing policies should be designed in a way so as to incentivize CRM solution buyers to invest enough efforts in cyber-security so as not to be a source for catastrophic cyber-risks. Finally, while appropriate CRM pricing policies might 'nudge' the demand side to improve their cyber-hygiene, all the contributions together indicate the important role of regulators (e.g., the government) to regulate the enforcement of improved security strength in factory settings of IoT devices during/post manufacturing. This will mitigate (a) the negative effect of human "laziness" towards improving cyber-hygiene, and (b) the chances of society dealing with catastrophic risks.

II. (Catastrophic) IID Cyber-Risk Aggregation

One of the key features of risk management (CRM) (e.g., via insurance) in general as a business model is its ability to pool different types of risks, thereby reducing an underwriter's overall risk exposure. This is particularly true for a reinsurer (not necessarily a cyber re-insurer) , who is in a position to significantly diversify its risks, by selling reinsurance contracts to very different front-line insurers who specialize in different sectors (e.g., retail, pharmaceutical, manufacturing, etc.), primarily independent of one another. This means that a reinsurer typically takes on or *aggregates* a fraction of many different risks that are most likely to be independent of one another. However, this independence property may not hold true of some cyber-risks. *In Section II & III, we make a simplistic assumption that cyber-risks aggregated by a aggregate cyber-risk manager are independent, and leave the analysis of tail-dependent cyber-risks for Section IV.* Specifically, in this paper we will often consider the average of n (dependent or independent) cyber-risks $X_1, \cdots, X_n$ arising from different IoT-driven organizations in a smart society, given by $Z_{\underline{w}} = \frac{1}{n}\sum_{i=1}^{n} X_i$, or more generally, the weighted average given a fraction of each cyber-risk $w = [w_1, \cdots, w_n]$: $Z_w = \sum_{i=1}^{n} w_i X_i$. In what follows, in this section we will first examine, for increasing cyber-risk spread (variance), the distribution resulting from aggregating catastrophic cyber-risks, whose first and second moments are undefined. We will then generalize this result and examine the standard VaR risk measure (*see online Appendix A for a definition and a valid rationale for using the VaR metric*) as a result of aggregating n cyber-risks (catastrophic or otherwise).

Symbol	Description
$VaR_q(X)$	Value-at-Risk (VaR) of X at level q
$S_\alpha(\sigma, \beta, \mu)$	stable and heavy-tailed distribution characterized by the index of stability α, scale parameter σ, symmetry index β, and location parameter μ
$\overline{\mathcal{CS}}(r)$	class of symmetric distributions that are convolutions of $S_\alpha(\sigma, 0, 0)$ distributions with $r \leq \alpha < 2$ and $\sigma > 0$
$\underline{\mathcal{CS}}(r)$	class of symmetric distributions that are convolutions of $S_\alpha(\sigma, 0, 0)$ distributions with $0 \leq \alpha < r$ and $\sigma > 0$
$\mathcal{CSLC}$	class of symmetric distributions that are convolutions of symmetric distributions that are either log-concave or stable with exponent $\alpha > 1$
Z_w	aggregated risk with weights w and risk portfolio $X_1, \cdots, X_n$, such that $Z_w = \sum_{i=1}^{n} w_i X_i$
a	length of support of a probability distribution

TABLE I: Table of Notation

A. An intuitive observation

To give some intuition, we begin with a simple comparison of risk spread (standard deviation) between aggregating light-tailed distributions and heavy-tailed distribution. Consider the *Normal* distribution as a representative of the former and the *Levy* [24] and the *Cauchy* distributions as representatives of the latter that are *statistically stable*[25]; the latter exhibit power-law decay with cdf given by $F(-x) \approx x^{-\alpha}, x, \alpha > 0$. For n IID normal $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, their average $\frac{1}{n}\sum_{i=1}^{n} X_i$ is also normally distributed with $\mathcal{N}(\mu, \frac{1}{n}\sigma^2)$. The implication here is that the aggregate risk has a spread (the standard deviation) that grows as $\sqrt{\frac{1}{n}}$ of σ for a given μ, suggesting a decrease in average risk as one spreads over an increasing number of individual risks. Thus in this case higher diversification – the spreading over larger pool of risks – is desirable.

Now consider the Levy distribution denoted by $\mathcal{L}(\mu, \sigma)$, with location parameter μ, scale σ, pdf and cdf is respectively given by

$$\phi(x) = \begin{cases} \sqrt{\frac{\sigma}{2\pi}} e^{\frac{-\sigma}{2(\mu - x)}} (\mu - x)^{\frac{-3}{2}} & \text{if } x < \mu, \\ 0 & \text{if } x \geq \mu, \end{cases}$$

$$F(x) = \begin{cases} \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\frac{-\sigma}{2(\mu - x)}}} e^{-t^2} dt & \text{if } x < \mu, \\ 1 & \text{if } x \geq \mu. \end{cases}$$

A simple algebraic manipulation will suggest that for IID $X_1, \cdots, X_n \sim \mathcal{L}(\mu, \sigma)$, we have $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{L}(\mu, n\sigma)$. In other words, contrary to the normal case, the risk spread as a result of aggregating Levy distributions *increases* linearly in the number of individual risks for a given μ. This suggests that risk aggregation in this case is undesirable.

As another example, consider the Cauchy distribution denoted by $\mathcal{G}(\mu, \sigma)$, with location parameter μ and scale σ, pdf given by

$$\phi(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{(x-\mu)^2}{\sigma^2}\right)},$$

and the corresponding cdf given by

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\sigma}\right).$$

Again, standard results suggest that for IID $X_1, \cdots, X_n \sim \mathcal{G}(\mu, \sigma)$, we have $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{G}(\mu, \sigma)$, meaning that the spread of the aggregate risk is unchanged from the individual risk spread. So in this case risk aggregation does not bring risk reduction benefit; it is neither desirable nor undesirable.

The above suggests that the notion of spreading risks is sound when the underlying individual risks are light-tailed, but casts doubts on the wisdom of doing so when the underlying risks are heavy-tailed. In the remainder of this section we formally establish this result using the VaR risk measure.

B. Aggregating IID catastrophic cyber-risks

We first consider aggregating IID risks X_i from the family $\underline{\mathcal{CS}}(1)$, which are class of distributions that are convolutions of symmetric and stable distributions with characteristic exponent $\alpha < 1$ - those exhibiting an *infinite* mean and variance, and representing catastrophic cyber-risks (see online Appendix A for details). We have the following result regarding VaR performance post cyber-risk aggregation, the proof of which is in online Appendix B .

Theorem 2.1: Consider IID r.v's $X_i \sim \underline{\mathcal{CS}}(1), i = 1, \cdots, n,$ $q \in (0, 1)$*, and* n*-vector of weights* $w, v \in \mathbf{R}_{+}^{n}$*. Then*

1) $VaR_q(Z_w) > VaR_q(Z_v)$ *if* $v \prec w$ *and* v *is not a permutation of* w*; in other words, the function* $VaR_q(Z_w)$ *is strictly Schur-concave in* $w \in \mathbf{R}_{+}^{n}$.
2) *In particular,* $VaR_q(Z_{\bar{w}}) < VaR_q(Z_w) < VaR_q(Z_{\underline{w}})$, $\forall w \in \mathcal{I}_n$ *such that* $w \neq \underline{w}$ *and* w *is not a permutation of* $\bar{w}$.

Theorem Implications - On a practical note, the theorem simply implies that when an aggregate cyber-risk covering agency is faced with covering independent and identical catastrophic cyber-risk distributions, the variance of the combined distribution increases with the number of piled up cyber-risks - *simply a dampening signal for-profit cyber-risk managers to contribute to a sustainable aggregate loss coverage market*.

Now consider the special borderline case $\alpha = 1$ (borderline catastrophic), which corresponds to IID $X_1, \cdots, X_n$ with a symmetric Cauchy distribution $S_1(\sigma, 0, 0)$. In this case, we have for all $w = (w_1, ..., w_n) \in \mathcal{I}_n$, $Z_w = \sum_{i=1}^{n} w_i X_i =_d X_1$. Consequently, $VaR_q(Z_w) = VaR_q(X_1)$ is independent of w and is the same for all portfolios of risk X_i with weights $w \in \mathcal{I}_n$. In other words, in such a case variations in a portfolio has *no effect* on riskiness of its aggregate return. Thus, the symmetric Cauchy distribution with characteristic exponent $\alpha = 1$ is the boundary between extremely heavy-tailed distributions (for which aggregate coverage is statistically not incentive compatible) with infinite first moments, and moderately heavy tailed distributions with finite first moments (aggregate coverage might be sustainable). Similarly, for general weights $w = (w_1, ..., w_n) \in \mathbf{R}_{+}^{n}$, $\alpha = 1$ implies $Z_w = \sum_{i=1}^{n} w_i X_i =_d (\sum_{i=1}^{n} w_i) X_1$. Thus, $VaR_q(Z_w) = (\sum_{i=1}^{n} w_i) VaR_q(X_1)$ is independent of w so long as $\sum_{i=1}^{n} w_i$ is fixed. Consequently, $VaR_q(Z_w)$ is both Schur-convex and Schur-concave in $w \in \mathbf{R}_{+}^{n}$ for IID $X_i \sim S_1(\sigma, 0, 0)$.

C. Aggregating IID non-catastrophic cyber-risks

We now consider aggregating IID risks X_i from the family $\overline{\mathcal{CSLC}}$, which are class of distributions that are convolutions of symmetric distributions that are either log-concave or stable with exponent $\alpha > 1$ - those exhibiting *finite* mean and variance, and representing non-catastrophic heavy-tailed cyber-risks (see online Appendix A for details). We have the next result regarding VaR performance post cyber-risk aggregation, the proof of which is in online Appendix B .

Theorem 2.2: Consider IID r.v's $X_i \sim \overline{\mathcal{CSLC}}, i = 1, \cdots, n,$ $q \in (0, 1)$*, and* n*-vector of weights* $w, v \in \mathbf{R}_{+}^{n}$*. Then*

1) $VaR_q(Z_w) < VaR_q(Z_v)$ *if* $v \prec w$ *and* v *is not a permutation of* w*; in other words, the function* $VaR_q(Z_w)$ *is strictly Schur-convex in* $w \in \mathbf{R}_{+}^{n}$.
2) *In particular,* $VaR_q(Z_{\underline{w}}) < VaR_q(Z_w) < VaR_q(Z_{\overline{w}})$, $\forall w \in \mathcal{I}_n$ *such that* $w \neq \underline{w}$ *and* w *is not a permutation of* $\bar{w}$.

Theorem Implications - On a practical note, the theorem simply implies that when an aggregate cyber-risk covering agency is faced with covering independent and identical non-catastrophic cyber-risk distributions, the variance of the combined distribution does not increase with the number of piled up cyber-risks - *simply an encouraging signal for-profit cyber-risk managers to contribute to a sustainable aggregate loss coverage market*. While this latter point has long been believed and empirically validated in the cyber-insurance research literature, the result from Theorem 2.1 is a surprising new facet that we unravel in this paper via theory.

III. Aggregating Curtailed IID Catastrophic Risks

In this section we analyze what happens when aggregating multiple heavy-tailed risks each of which has been curtailed, *to fit the realistic scenario where cyber-risk managers have upper bounds on coverage.* We also study the role of how the length of the distributional support needed for the analogue to hold depends on the number of cyber-risks in a manager's portfolio and the degree of heavy-tailedness of unbounded cyber-risk distributions. We have the following result, an analogue of Theorem 2.1 for curtailed catastrophic cyber-risks in this regard, the proof of which is in online Appendix B .

Theorem 3.1: Let $n \geq 2$ *and let* $w \in \mathcal{I}_n$ *be a weight vector with* $w_{[1]} \neq 1$*. Let* $X_i, i = 1, \cdots, n$ *be IID r.v.'s* $\sim \underline{\mathcal{CS}}(r)$ *for some* $r \in (0, 1)$ *and their respective a-truncated version given by* Y_i *defined above. Denote* $G(w, z) = P(w_{[1]}X_1 + w_{[2]}X_2 >$

$z) - P(X_1 > z)$, which is positive if $w_{[1]} \neq 1$ (via Theorem 2.1). For any $z > 0$, and all

$$a > \left(\frac{\mathbb{E}[|X_1|^r](n-1)}{2G(w,z)}\right)^{\frac{1}{r}}, \tag{1}$$

the following inequality holds:

$$P(Y_w(a) > z) > P(Y_1(a) > z). \tag{2}$$

Note that $G(w,z)$ reflects that $VaR_q[X_w] > VaR_q[w_{[1]}X_1 + w_{[2]}X_2] > VaR_q[X_1]$.

The **implications of this theorem** are multifarious and are presented in multiple blocks.

Implication 1 - The practical implications of the theorem are analogous to Theorem 2.1 in the case of bounded cyber-risks. More specifically, cyber-risk aggregation coverage continues to be disadvantageous in general for catastrophic truncated heavy-tailed distributions. For $n \geq 2$ and any cyber-risk valuation $z > 0$, there exists n cyber-risks with finite support with the property that the variance return of the aggregate cyber-risk portfolio is riskier than that of the portfolio consisting of a single cyber-risk. From a mathematical viewpoint, Theorems 2.1 and 3.1 indicate that VaR is not sub-additive and, thus, its coherency (see online Appendix A for details) is always violated in the class of extremely heavy-tailed cyber-risks with infinite first moments. More specifically, Theorem 3.1 implies that VaR may also be non-coherent in the world of cyber-risks with bounded distributional support. We just proposed conditions under which it is statistically incentive compatible for a (re)-insurer to spread catastrophic cyber-risks having heavy tails. One could also further study conditions under which it will *not be optimal* to spread risks - in the interest of space, this analysis is provided in online Appendix C and also in [26].

Implication 2 - We note that in *the special case of a cyber-risk portfolio with equal weights*, $\tilde{w}_n = \left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$, we have

$$G(\tilde{w}_n, z) = P\left(\frac{X_1 + X_2}{2} > z\right) - P(X_1 > z). \tag{3}$$

This means that the length of the distributional support reflecting statistical incentive non-compatibility to aggregate cyber-risk coverage in Theorem 3.1 can be taken to be same for all the portfolios with equal weights $\tilde{w}_n$. This holds, obviously, for the whole class of the portfolios w such that $w_{[1]} < \frac{1}{2}$. Furthermore, a similar result holds as well for the class of portfolios w such that $w_{[1]} < 1 - \epsilon$, (and, thus, $w_i < 1\epsilon$ for all i), where $0 < \epsilon < \frac{1}{2}$. As follows from the proof of Theorem 3.1, for all such portfolios w, the theorem holds for $a > \left(\frac{\mathbb{E}[|X_1|^r](n-1)}{2\tilde{G}(w,z)}\right)^{\frac{1}{r}}$, where $\tilde{G}(\epsilon, z) = P\left((1-\epsilon)X_1 + \epsilon X_2 > z)\right) < G(w,z)$. This follows since any vector w with $w_{[1]} < 1 - \epsilon$ is majorized (see basics of majorization in the online Appendix A) by the vector $(1 - \epsilon, \epsilon, 0, .., 0)$.

Implication 3 - From the proof of Theorem 3.1, it follows that, in the special case of portfolios with equal weights $\tilde{w}_n = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ where $n > 2$, the length of the interval of truncation a can be reduced to a smaller value. In such a case, the theorem holds under the restriction $a > \left(\frac{E|X_1|^r(n-1)}{2F_n(z)}\right)^{1/r}$, where

$$F_n(z) = P\left(\frac{\sum_{i=1}^n X_i}{n} > z\right) - P\left(X_1 > z\right) \tag{4}$$

Note that, by Theorem 2.1, $F_n(z) > H(z) = G\left(\tilde{w}_n, z\right)$ for $n \geq 3$. *This suggests that if the support is large compared to the number of cyber-risks to be aggregated, it might be infeasible for an aggregate risk manager to cover the risks. This demonstrates the "unpleasant" properties of VaR as a cyber-risk measure under heavy-tailedness does not arise from the relatively high likelihood of getting very large losses but rather from the fact that there are too few cyber-risks available for the profitable aggregate cyber-risk coverage to work.*

Implication 4 - Theorem 3.1 also shows that, for a specific loss probability q, there exists a sufficiently large a such that the value at risk $VaR_q[Y_w(a)]$ of the return $Y_w(a)$ at level q is greater than the value at risk $VaR_q[Y_1(a)]$ of the return $Y_1(a)$ at the same level: $VaR_q[Y_w(a)] > VaR_q[Y_1(a)]$. *This highlights the dampening factor to the sustainability of covering aggregate heavy-tailed cyber-risks.* One should emphasize that the last inequality between the returns $Y_w(a)$ and $Y_1(a)$ holds for the particular fixed loss probability q and, in the comparisons of the values at risks $VaR_q\left[Y_w(a)\right]$ and $VaR_q\left[Y_1(a)\right]$, the length of the interval needed for the reversals of the stylized facts on the portfolio variation depends on q (similar to the fact that in Theorem 3.1, the length of the distributional support a depends on the value of the disaster level z - denoting the degree of heavy-tailedness). This is the crucial qualitative difference of the results in Theorem 3.1 for bounded/curtailed cyber-risk distributions and their implications for the value at risk, from those given by Theorem 2.1 and Theorem 3.1 for unbounded risks, where the inequalities hold for all $z > 0$ and all $q \in (0, 1)$.

Implication 5 (Case of non-identical distributions) - The analogues of Theorem 2.1 hold for i.i.d. risks $X_1, \ldots, X_n$ that have skewed extremely thick-tailed stable distributions with infinite first moments: $X_i \sim S_{0<\alpha<1}(\sigma, \beta, 0), \alpha \in (0,1)$, $\sigma > 0, \beta \in [-1,1], i = 1, \ldots, n$. As follows from the proof of Theorem 3.1 (see online Appendix B), this implies that complete analogues of the results in the present section for bounded versions of symmetric risks from the classes $\underline{\mathcal{CS}}(r)$ continue to hold for truncated extremely heavy-tailed stable distributions $S_\alpha(\sigma, \beta, 0)$ with $\alpha \in (0,1), \sigma > 0$, and an arbitrary skewness parameter $\beta \in [-1,1]$. In particular, Theorem 3.1 continues to hold for arbitrary skewed risks $X_i \sim S_\alpha(\sigma, \beta, 0)$, $\alpha \in (0,1), \sigma > 0, \beta \in [-1,1]$ if $a > \left(\frac{E|X_1|^r(n-1)}{G(w,z)}\right)^{1/r}$.

Results Overview and Impact on IoT Societies - As a summary of the theory results in this section and the previous one, Figure 1 provides a graphical illustration of the impact of the type and number of cyber-risks on a risk manager's valuation (statistical utility, i.e., decreased VaR) of covering aggregate cyber-risk. The interesting observation is that for cases B and C illustrating curtailed cyber-risks, there is a drop in the utility, i.e., increased VaR, as a function of the number of cyber-risks, in covering aggregate risk, followed

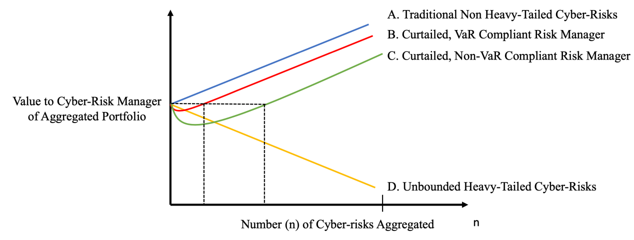


Fig. 1: Conceptual illustration of statistical utility of cyber-risk aggregation as a function of the number of **i.i.d.** cyber-risks. For a given 'n', cyber-risk valuation 'z', there always exists a 'dipping' statistical utility region for cases B and C, and the region expands as a increases (ref. Theorem 3.1). $\alpha < 1$ for B, C, and D.

by an indefinite increase in utility henceforth. The initial drop is due to the tradeoffs from the higher costs of aggregate and increased variance-induced coverage due to a certain threshold 'n' catastrophic cyber-risks versus the benefit received from coverage premiums. Clearly, beyond 'n' cyber-risks the statistical benefits of aggregate cyber-risk coverage outweighs the negatives of increased risk spread. The outcome of cases A and D are intuitively obvious.

In [18], the author rationalizes why aggregate loss coverage services like re-insurance might be sustainable, and not encounter a systemic catastrophe problem. For the general re-insurance setting, he mentions (i) a portfolio of independent risks and geographical diversification, (ii) partial cessation of risk with proper risk screening, and (iii) lack of liability loops, to be the major factors in favor of re-insurance services being sustainable. *However, there are major differences between general re-insurance and cyber re-insurance services, that allows us to closely look at cyber re-insurance service sustainability under universal risk types.* Clearly (i) and (ii) are impractical when major cyber-catastrophes occur and impact IoT societies (e.g., ones caused by the WannaCry and Mirai attacks) (Curve D). In the most optimistic scenarios, Figure 1 illustrates what the size and nature of the coverage portfolio should look like for a cyber re-insurer assuming *limited* coverage liability for i.i.d. heavy-tailed cyber-risks (Curves A-C). However, the challenge still remains to deal with non i.i.d. heavy-tailed cyber-risks such as those posed by *WannaCry* and *Mirai*.

IV. Aggregating Non-IID Heavy-Tailed Risks

Cyber-risks are not only heavy tailed in nature, but are likely to be correlated, i.e., tail-dependent. This is true especially in scenarios of major systemic impact causing cyber-attacks. The likelihood of systemic loss impacts are fairly high in a service-networked smart society [27][1] driven by IoT technologies. In this section we study the effect on VaR on aggregating such cyber-risk types. Statistical correlations and dependencies between distributions are often captured systematically using *copulas* [28][29] (see online Appendix A for a preliminary introduction), that are multivariate functions of marginal distributions outputting dependence values. In our case, the marginal distributions are cyber-risk random variables having a heavy-tail characterized via a power-law distribution family.

To illustrate dependencies between such marginal distributions, we start with the bivariate (generalization to follow) *Eyraud-Farlie-Gumbel-Morgenstern* (EFGM) copula - a power type copula (see online Appendix A for more details) whose marginal distributions obey the power law to reflect heavy-tailed cyber-risk distributions (both catastrophic and otherwise). Let (X_1, X_2) be random variables with the EFGM copula and power-law marginals. Then, for any $x \geq 1$ and for $j = 1, 2$, we have

$$F_j(x) \sim 1 - x^{-\alpha};\ f_j(x) \sim \alpha x^{-\alpha-1}$$

$$H(x_1, x_2) = \Pi_{i=1}^2 F_i(x_i)\left[1 + \gamma\left(1 - F_1(x_1)\right)\left(1 - F_2(x_2)\right)\right]$$
$$h(x_1, x_2) = \Pi_{i=1}^2 f_i(x_i)\left[1 + \gamma\left(1 - 2F_1(x_1)\right)\left(1 - 2F_2(x_2)\right)\right]$$

Let $(\xi_1(\alpha), \xi_2(\alpha))$ be independent random variables from power-law distributions with tail index α, often called independent copies of (X_1, X_2). Our key insight is that in the tail, the behavior of products and powers of power-law densities and distributions of X_j 's is identical to the behavior of their independent copies. This makes it possible to provide asymptotic (with respect to the loss comparisons between the VaR of the aggregated loss and that of a single risk. More specifically, the crucial component of $\mathbb{P}\left(\frac{X_1+X_2}{2} > x\right)$ under the EFGM copula can be written as follows

$$\begin{aligned}
&\int_{\frac{s+t}{2}>x} \alpha^2 s^{-\alpha-1} t^{-\alpha-1}\left(2s^{-\alpha}-1\right)\left(2t^{-\alpha}-1\right) ds dt \\
=&4\alpha^2\mathbb{P}\left(\frac{\xi_1(2\alpha)+\xi_2(2\alpha)}{2} > z\right) - 2\alpha^2\mathbb{P}\left(\frac{\xi_1(2\alpha)+\xi_2(\alpha)}{2} > z\right) \\
&- 2\alpha^2\mathbb{P}\left(\frac{\xi_1(\alpha)+\xi_2(2\alpha)}{2} > z\right) + \alpha^2\mathbb{P}\left(\frac{\xi_1(\alpha)+\xi_2(\alpha)}{2} > z\right)
\end{aligned}$$

where the behavior of the individual summands for large z is driven by the lowest tail index of ξ_j in the spreading portfolio.

We formalize this result in the following theorem (see online Appendix B for a proof), which generalizes to n dependent heavy-tailed random variables $X_1, X_2, \ldots, X_n$ with multivariate EFGM copula and power-law marginals.

Theorem 4.1: For an asymptotically large $z > 0$, and any $n, \alpha > 0$

$$\mathbb{P}\left(\sum_{i=1}^n X_i > zn\right) \sim \mathbb{P}\left(\sum_{i=1}^n \xi_i(\alpha) > zn\right)$$

Theorem Implications - The result suggests that suboptimality of cyber-risk aggregation in the VaR framework for extremely heavy tailed losses carries over from independence to the dependence-capturing EFGM copula. *That is, cyber-risk aggregation increases VaR of dependent extremely heavy tailed risks within this copula family.* It is also easy to see that for dependent losses with the EFGM copula and sufficiently small loss probability q, we have

$$\begin{aligned}
&VaR_q\left(\tfrac{X_1+X_2}{2}\right) < VaR_q\left(X_1\right), &&\text{if} \quad \alpha > 1 \\
&VaR_q\left(\tfrac{X_1+X_2}{2}\right) > \mathrm{VaR}_q\left(X_1\right), &&\text{if} \quad \alpha < 1
\end{aligned}$$

Important generalizations of Theorem 4.1 arise if we consider the wider class of power-type copulas. Most popular members of this class such as the polynomial copula of Drouet Mari and Kotz [30] and the copula with cubic section of Nelsen et al. [31] can be written in the following general form

$$C(u_1,\ldots,u_n) = \sum_{i_1,\ldots,i_n=0,1,\ldots} \gamma_{i_1,i_2,\ldots,i_n} \cdot u_1^{i_1} \cdot u_2^{i_2} \cdot \ldots \cdot u_n^{i_n} \tag{5}$$

for a multiple index $i = (i_1, i_2, \ldots, i_n)$ and a set of corresponding parameters γ_i with appropriate restrictions that make $C(u_1,\ldots,u_n)$ a copula. For example, Drouet Mari and Kotz [30][32] show how to obtain a polynomial copula from function $f = u^k v^q$. The key feature of such copulas is that they and their densities can be expressed as powers of u_j 's. This allows to apply similar arguments as for EFGM. To this end, we have the following theorem, the proof of which is in online Appendix B .

Theorem 4.2: For dependent losses with a power-type copula in (5) and for an asymptotically large $z > 0$, *and any* $n, \alpha > 0$, *the conclusions of Theorem 4.1 hold.*

Theorem Implication - The implications are the same as that of Theorem 4.1.

V. Experimental Evaluation

In this section, we put our theory to a rigorous test using real-world cyber-loss data. We want to study whether aggregating individual cyber-risks from different IoT-driven organizational sources (assumed to show characteristics of real-world cyber-loss) in a smart society increase or decrease a risk manager's VaR/Expected Utility (EU) - the scalar metric for measuring the extent of aggregate cyber-risk. In particular, (a) we relax the mathematical assumption used in theory that cyber-loss distributions are stable - might not always be the case in practice, and (b) we assume that cyber-risk managers are boundedly rational in estimating the extent of cyber-risk. In a nutshell, we first show using real world data that individual cyber-losses can indeed exhibit a heavy-tailed statistical nature. We then investigate the VaR/EU trends with increasing number of heavy-tailed cyber-risks to be aggregated, for both rational, and boundedly rational cyber-risk manager behavior.

A. Experimental Setting

We consider 1553 cyber losses between 1995 and 2014 extracted from the *SAS OpRisk* database. For detailed description of the data, we refer the reader to [3] and [33]. To model the bounded rationality of cyber-risk managers in gauging the extent of cyber-risk, we use prospect theory introduced by Kahneman and Tversky [34] to model their behavior. We first perform several goodness-of-fit tests for several widely used distributions to characterize the true nature of the cyber-loss distribution. Namely, we use the *normal*, *log-normal*, *general Pareto*, and *peak-over-threshold* (POT) distributions for the purpose of comparison. Based on the goodness-of-fit-statistics (using Log-Likelihood, AIC, BIC, Kolmogorov-Smirnoff, and Anderson-Darling tests), we find that the generalized Pareto distribution and the POT approach fit the data best. The estimated *Pareto Index* (the exponent in a power law distribution) characterizing a heavy-tailed distribution for the generalized Pareto distribution is 0.62 and for the POT approach it is 0.81, using analysis adopted from [35]. *We thus can confirm that cyber risks are indeed very heavy tailed and the expectation and variance do not exist.* Empirically, illustrating the tail dependencies on cyber-loss is more difficult because of the lack of data (exhibiting tail dependencies on loss data) and analyses. For this reason, different potential dependency structures, *generated via statistical copulas*, will be considered in our empirical part.

If a cyber-risk manager (e.g., an insurer) takes on a random risk X, a function of n - the number of cyber-risks it accepts to aggregate, the effective outcome (before opting for cyber re-insurance services) for the insurer once X is realized is:

$$V(x) = \begin{cases} X \text{ if } X < k, \\ k \text{ if } X \geq k, \end{cases} \tag{6}$$

where k is the limit of the amount of cyber-risk it can accept - true of practice. In the special case when there is no limited liability, i.e., when $k = \infty$, we have $V(X) = X$ for all X. If $k < \infty$, u is defined only on $[0, k]$, and without loss of generality $u(k) = 0$. Here, we assume the utility function of a **perfectly rational** and risk-averse cyber-insurer to be generally of the following form:

$$u(x) = (V(x))^\beta,\ \beta \in (0,1),$$

which is the power utility function, and for x being a risk variable, is a Von-Neumann Morgenstern (VNM) utility function. β is degree of risk-aversion of the cyber-insurer. However, for the purpose of this section, we will assume a **boundedly rational** cyber-risk manager, whose behavior-driven parameters (in contrast to the perfectly rational setting) is given by

$$V(x) = E_w(x) = \int w(p(x))V(x)dx;$$

$$V(x) = \begin{cases} X^\beta \text{ if } 0 < X \geq k, \\ -\lambda(-x)^\beta \text{ if} X > k, \end{cases} \tag{7}$$

and the probability weighing function is specified as

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}, \tag{8}$$

where λ is the prospect-theoretic loss aversion coefficient of the cyber-risk manager, $p(x)$ is the cyber-risk distribution function, and β - the manager's VNM-theoretic coefficient of risk aversion. We set the parameters β = 0.88, λ = 2.25, and γ = 0.61, as in [34]. To capture dependency between cyber-risks, we use the Gaussian and Clayton statistical copulas, as suggested in [36][37][38].

B. Experimental Results

For a prospect-theoretic setting, we first investigate the effect of the number of i.i.d. cyber-risks (of various types) to be aggregated, on the value at risk (VaR) for a cyber-risk manager, as the first moments may not exist to compute E(X).

We observe from Figure 2 that $VaR_{0.995}(X)$ monotonically decreases for normal and log-normal individual cyber-risk distributions (fitted using our data set) - though the VaR for log-normal risks decreases, at a slower rate. On the other hand, $VaR_{0.995}(X)$ (denoted as VaR from now on throughout the

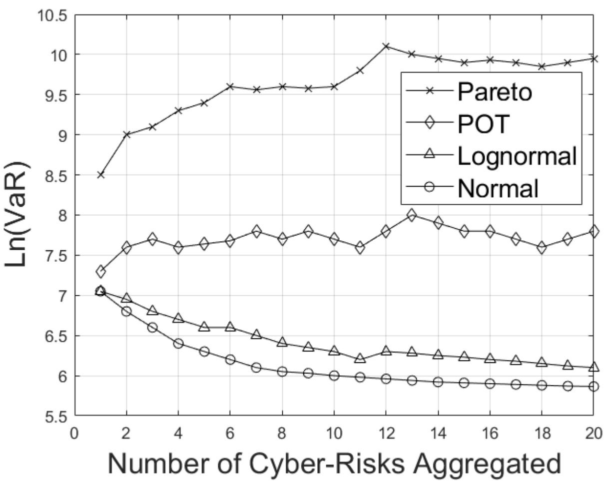


Fig. 2: Cyber-Risk Aggregation Performance (on the VaR Metric) for **i.i.d.** Risks of Pareto, POT, Log-Normal, and Normal Distributions.

section) increases (not monotonically) for Pareto and POT individual cyber-risk distributions, as is expected in theory. *However, the non-monotonicity indicates (also in accordance to our theory) that for heavy-tailed cyber-risks, there exists a certain number of i.i.d risks, aggregating which does not increase VaR.* To focus on our empirical data set, we use statistical bootstrapping to simulate the VaR for varying number of aggregated cyber-risks. In this regard, we draw directly from our original sample instead of the different distributions assumed above. The sample is drawn with replacement (thus, i.i.d.) and is of equal size as the original data set (m=1553 observations). Moreover, we calculate the confidence interval by repeating the bootstrapping itself.

Figure 3a shows the bootstrapped VaR and its confidence interval. We observe that the bootstrapped VaR (induced by the empirical loss distribution) always lies above the log-normal VaR and the aggregation benefit is much less prevalent than assumed. As a consequence, *in accordance with theory, not to aggregate heavy-tailed risks at all would be optimal from a cyber-risk management perspective.* Since the dependency affects the aggregation results, we also simulate the VaR for different dependency structures, but for non-heavy tailed cyber-risks. Figure 3b plots the VaR again as a function of the number of cyber-risks to be aggregated, for identical (but not independent) distributed cyber-risks and different copulas.

We observe that the VaR is decreasing for both the Gaussian (an instance of a symmetrical copula) and Clayton (an instance of a non-symmetrical copula) copulas with increasing number of cyber-risks to be aggregated - *implying that aggregating non i.i.d. non heavy-tailed cyber-risks is sustainable for the risk manager.* However, stronger dependency between the cyber-risks would cause extreme losses to become more likely and the consequent relative increase to VaR.

We now focus on an expected utility (EU) setting induced on limited liability where applicable, to compare cyber-risk aggregation performance with the prospect-theoretic setting. Figures 4a and 4b show the EU-theoretic performance based on a power utility function $u(x)$ for aggregating i.i.d. and non i.i.d. cyber-risks respectively. As expected, for normally distributed i.i.d. cyber-risks (Figure 4a), we attain an increase in expected utility with increase in the number of cyber-risks aggregated. However, this is not true for a heavy tailed distribution such as the Pareto or the log-normal distributions. However, in the special case of non-i.i.d. normal cyber-risks, risk aggregation increases expected utility.

We also study the role of pool of homogeneous cyber-risk managers (CRMs) that share[1] aggregate cyber-risk, on the EU of a single manager in that pool. We consider various instances of individual cyber-risks with Pareto index α that either is 1 (characterizing heavy-tail nature of cyber-risk), or lie below 1 (characterizing extremely heavy-tailed cyber-risks). Figure 5 shows that for risk with a Pareto Index of 1 and limited liability of k = 60, the expected utility of a single manager for different aggregation and cyber-risk pooling sizes (#CRMs), is U-shaped. The U-shape denotes that the benefit from aggregation first decreases before it eventually increases again (similar trend to that in Figure 1).

Using a Pareto index of 0.62 (as estimated from the data, and indicating an extreme heavy-tailed distribution) changes, *ceteris paribus*, the result completely, as shown in Figure 6a. *Since the expected utility decreases monotonically not providing any (pooled) coverage management such as insurance would be optimal and the aggregate coverage market would fail completely.*

A numerical analysis shows that the U-shape can only be observed if the Pareto tail index is in the range of (0.8, 1.2). While the situation in Figure 5 leaves room for regulatory intervention, the model in Figure 6a does not. Figure 6b shows the same analysis for the POT model (with k = 60) that combines the log-normal distribution for the body with the Pareto distribution for the tail. Similar to the Pareto model in Figure 6a, the expected utility monotonically decays for all pool sizes (#CRMs) as the cyber-risk aggregation sizes increases. *Therefore, it is not beneficial for cyber-risk managers (CRMs) to supply any (pooled) aggregate cyber-coverage, and the subsequent coverage market fails.*

VI. Related Work

In this section, we solely focus on research related to cyber-risk aggregation. We partition this section in two parts: (i) the heavy-tailed and tail-dependent nature of cyber-risk, and (ii) feasibility insights regarding the profitable coverage of aggregate heavy-tailed cyber-risk.

A. On the Heavy-Tailed and Dependent Nature of Cyber-Risk

There are quite a few instances in practice where cyber-risks have shown heavy-tailed impact. In [17], Maillart and Sornette analyzed a *Datalossdb* 2017 dataset consisting of 956 personal identity loss incidents that occurred in the United States between year 2000 and 2008. They found that the personal identity losses per incident, denoted by X, can be modeled by a heavy tail distribution $P(X > n) \sim n^{\alpha}$ where α = 0.7 +/- 0.1, and more importantly this result holds for a variety of organizations: business, education, government, or

[1] We do not explicitly consider the strategic aspects of sharing in this work.

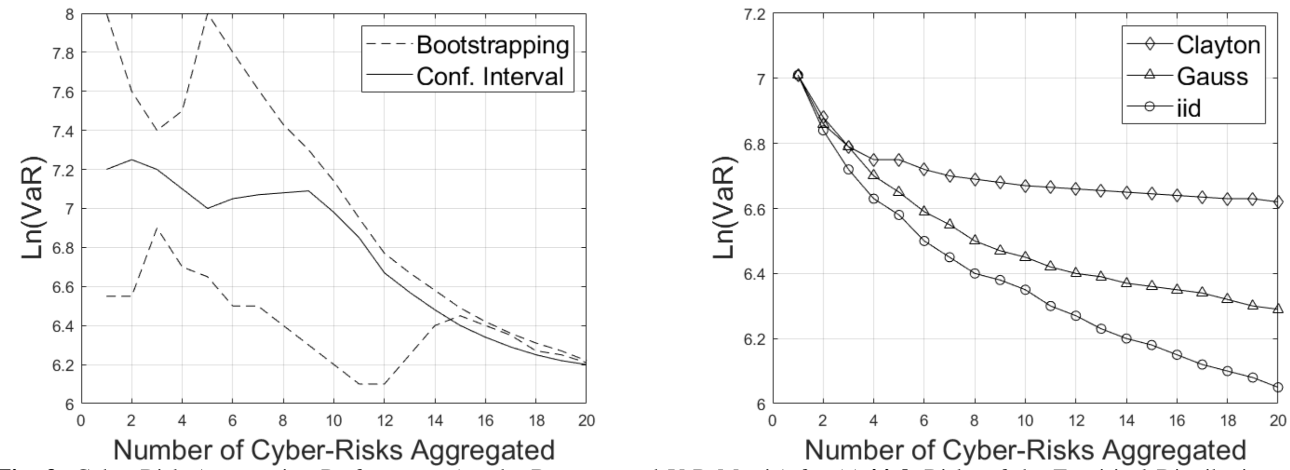


Fig. 3: Cyber-Risk Aggregation Performance (on the Bootstrapped VaR Metric) for (a) **i.i.d.** Risks of the Empirical Distributions, and (b) **non i.i.d.** Risks of the Log-Normal Distribution with Gauss and Clayton Copulas.

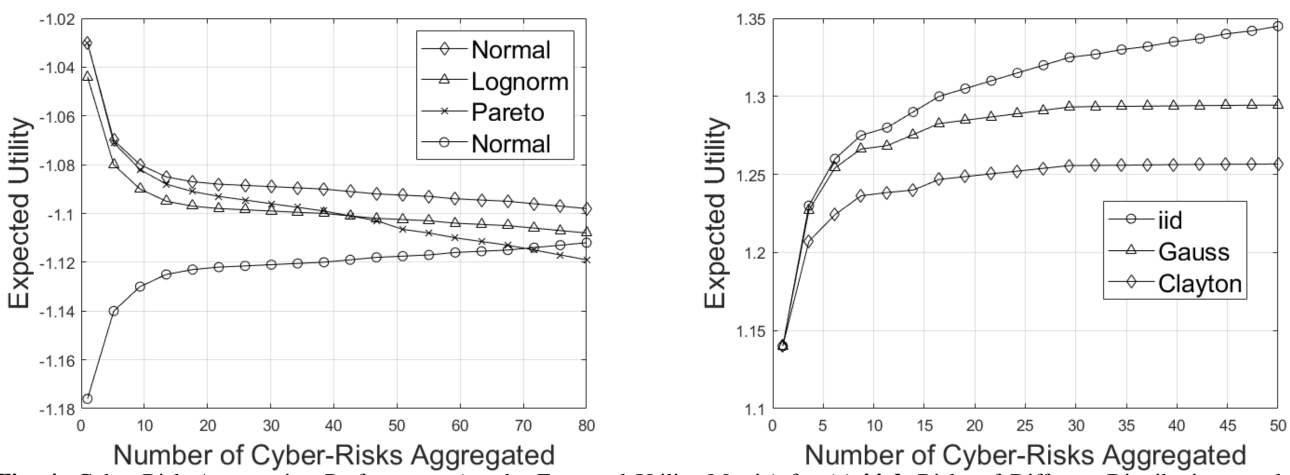


Fig. 4: Cyber-Risk Aggregation Performance (on the Expected Utility Metric) for (a) **i.i.d.** Risks of Different Distributions, and (b) **non i.i.d.** Normal Risk Distributions with Gauss and Clayton Copulas.

a medical institution. Because the probability density function of the identity losses per incident is static, the situation of identity loss is stable from the point of view of the breach size. Edwards et al. [39] analyzed a Privacy Rights Clearinghouse database of 2017 consisting of 2,253 breach incidents that span over a decade from 2005 to 2015. These breach incidents include two categories: negligent breaches (i.e., incidents caused by lost, discarded, stolen devices, or other reasons) and malicious breaching (i.e., incidents caused by hacking, insider and other reasons). They showed that the breach size can be modeled by log-normal or log-skewnormal distribution that are heavy-tailed distributions, and the breach frequency can be modeled by the negative binomial distribution. In [40], Wheatley et.al., merged and analyzed cyber-breach incidents from the Datalossdb and the Privacy Rights Clearinghouse database spanning over a decade (2000 to 2015). They used the Extreme Value Theory (EVT) [38] to study the maximum breach size, and further modeled the large breach sizes by a doubly truncated heavy-tailed Pareto distribution. There are also studies establishing the dependence among cyber risks. Notable among them are [41][7][42][43][44][16][45][46].

Shortcomings - Existing research in cyber-security has been successful in elucidating the heavy-tailed and tail-dependent nature of cyber-risk; however, *is yet to propose formally proven directions to allow a profit-minded cyber-risk manager to judge whether a collection of such risks is suitable to aggregate, under various degrees of heavy-tailedness.* This decision making problem will increasingly arise in the IoT age where major cyber-risks affecting smart societies will give rise to a systemic effects that cyber-risk managers have to deal with. It is a common perception from empirical studies and insurance literature that i.i.d. cyber-risks, even though heavy-tailed, are suitable for aggregation. In this paper, we showed quite the contrast for i.i.d. catastrophic heavy-tailed risks.

B. Covering Aggregate Cyber-Risk in IoT Societies

In a recent work, a group of researchers [27] have studied the problem of whether (a) the underlying network of service organizations in society relying on IT/IoT technologies, and (b) the statistical nature of cyber-risk distributions, positively or negatively affect aggregate cyber-risk managers in expanding their business. The authors surprisingly show that both,

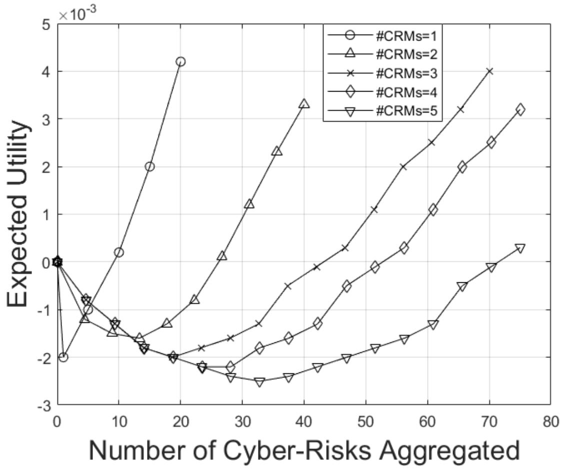


Fig. 5: Curtailed Cyber-Risk Aggregation Performance (on the Expected Utility Metric) for i.i.d. Pareto Risk Distributions, with varying number of CRMs. Here, k = 60, β = 0.0315, α = 1.

the underlying network, as well as i.i.d. and non i.i.d. non-heavy tailed cyber-risk distributions does not have a major role to play (does not imply independence) in encouraging or discouraging aggregate cyber-risk managers to expand or contract their coverage business.

Shortcomings - The cited work, though tackling the problem of judging the role of the network and the nature of cyber-risk distributions on the future of cyber-risk aggregation business, does not model catastrophic and tail-dependent heavy-tailed cyber-risks that may be a possibility in modern IoT-driven societies. However, as a major positive, their result in the work does provide confidence to aggregate cyber-risk managers to boost their cyber-loss coverage business for non-heavy tailed cyber-risks in a networked interdependent setting - something the digital society is in need of.

VII. Discussion and Summary

In this section, we first provide a brief review of the current state of insurance-driven CRM (an indicator of the degree of cyber-risk control) in small and medium IT-driven businesses that represent the majority of IT businesses in operation, and gauge the likelihood of cyber-risk distributions that may be sourced at these businesses. More importantly SMBs are highly service networked among themselves, and this network can pose significant cyber-risk aggregation challenges for CRM solution providers [27]. Our review is based on recent *Advisen* and *CyberScout* reports - indsutry leaders in CRM and cyber-security solutions. Finally, we summarize the paper.

A. Discussion

Small and medium-sized businesses are an important driver of the economy and should be empowered with progressive insurance policies that include cyber risk protection services, incident response and insurance coverages to provide the financial support needed to keep the doors open after an attack. As of 2020, insurers and cybersecurity services firms are innovating around the clock to create risk mitigation policies and procedures that can provide peace of mind to SMB leaders. However, despite a rise in cyberattacks against small and mid-size businesses, about 69% of SMB respondents to a recent survey by CyberScout said they did not carry cyber insurance coverage and worryingly many don't even have the appropriate security safeguards in place - clearly indicating a lack of seriousness by SMBs to improve their cyber-hygiene. Moreover, in the age of COVID, business owners are under a lot of pressure from the economic disruptions caused by the pandemic, and finding it even more challenging now to find the time to prioritize cyber-security. CyberScout found that 16% of the respondents had experienced a ransomware event and 40% said they would not know who to contact if they did fall victim to ransomware. SMBs also may not be aware enough of the ransomware risk – data breach ranks as the highest concern for 30% of respondents, but ransomware is tops for only 10%. And only 22% have a backup plan in place. Over half (51%) of survey respondents had no formal cyber-security training program, but 76% said they felt confident about their company's security infrastructure. However, the results revealed some possible gaps. A quarter of respondents said they send out "best practices" emails to employees, 22% reported performing "live fire" trainings and 20 percent also performed vulnerability testing. Annual trainings were the only measure taken by 18% of the respondents. Due to the pandemic, just over half (53%) reported having employees work remotely, but only 34% required the use of a VPN connection and only 17% took any steps to create or remind employees of remote work security protocols. In fact, 14% said they had no specific cyber measures for remote working. *Clearly, even in 2020, the state of cyber-security strength in SMBs is far from desired, and there is a significant likelihood of each being a source of heavy-tailed, i.e., catastrophic, cyber-risks in the event of major cyber-attacks.*

B. Paper Summary

In this paper, we provided a rigorous general theory to elicit conditions on (tail-dependent) heavy-tailed cyber-risk distributions under which a risk management firm will find it (un)profitable to provide aggregate cyber-risk coverage for IoT-driven smart societies. As our primary novel contributions, we proved that (a) spreading *catastrophic* heavy-tailed cyber-risks that are identical and independently distributed (i.i.d.), i.e., not tail-dependent, *is not* an effective practice for aggregate cyber-risk managers, whereas spreading *non-catastrophic* i.i.d. heavy-tailed cyber-risks is, and (b) spreading *catastrophic and tail-dependent* heavy-tailed cyber-risks *is not* an effective practice for aggregate cyber-risk managers. *A summary of cyber-risk management effectiveness results for various i.i.d./non-i.i.d. distributions is shown in Figure 7.* We conducted a real-data driven numerical study to validate claims made in theory - in the process we relaxed certain assumptions (made in theory) on the mathematical structure of cyber-risk distributions, and assumed that cyber-risk managers are boundedly rational rather than perfectly rational in the interpreting the extent of cyber-risk, as is usual in practice.

Acknowledgments

This work has been supported by the NSF under grants CNS-1616575, CNS-1939006, and ARO W911NF1810208.

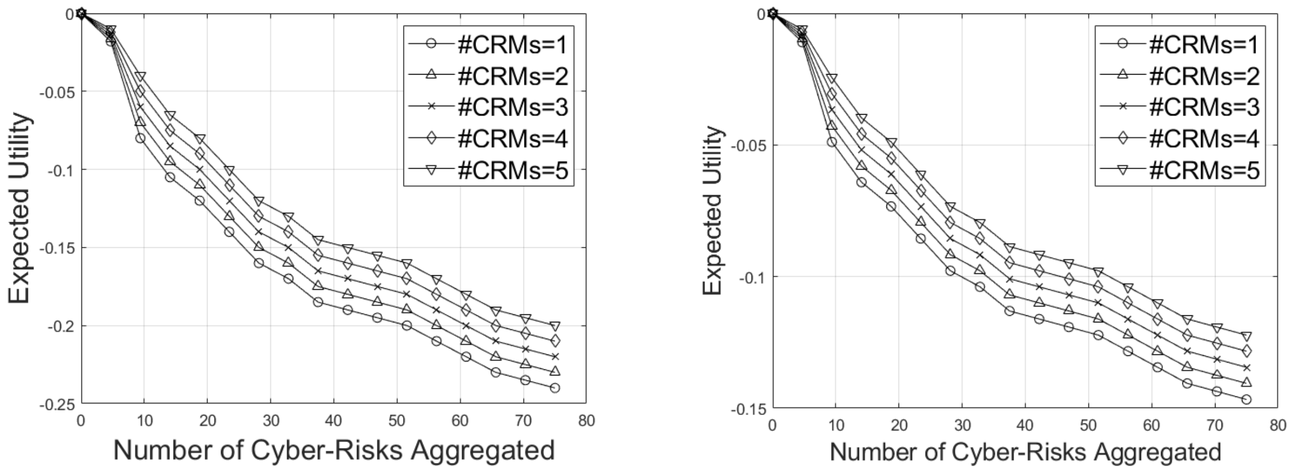


Fig. 6: Curtailed Cyber-Risk Aggregation Performance (on the Expected Utility Metric) for i.i.d. Pareto Risk Distributions, with varying number of CRMs. Here, k = 60, β = 0.0315, α = (a) 0.62 (b) 0.81.

HT - Heavy Tail
CHT - Curtailed Heavy Tail
α- Pareto Index

		moderately heavy-tailed		very heavy-tailed	
Dependence \ Distribution	Light Tail	HT (α>1)	CHT (α>1)	HT (α<1)	CHT (α<1)
IID	✓	✓	✓	✗	✗
non-IID	✓	✗	✗	✗	✗

Fig. 7: Summary of The Effectiveness (Yes (Tick)/No (Cross)) of Aggregate (Large Enough n) Cyber-Risk Management for Light and Heavy-Tailed IID/non-IID Distributions.

REFERENCES

[1] A. Coburn, E. Leverett, and G. Woo, *Solving Cyber Risk: Protecting Your Company and Society*. Wiley, 2018.

[2] A. Gilchrist, *IoT security issues*. Walter de Gruyter GmbH & Co KG, 2017.

[3] C. Biener, M. Eling, and J. H. Wirfs, "Insurability of cyber risk: An empirical analysis," *The Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 40, no. 1, pp. 131–158, 2015.

[4] R. Pal and L. Golubchik, "Analyzing self-defense investments in internet security under cyber-insurance coverage," in *2010 IEEE 30th International Conference on Distributed Computing Systems*, pp. 339–347, IEEE, 2010.

[5] R. Pal, L. Golubchik, K. Psounis, and P. Hui, "Will cyber-insurance improve network security? a market analysis," in *INFOCOM, 2014 Proceedings IEEE*, pp. 235–243, IEEE, 2014.

[6] R. Pal, L. Golubchik, K. Psounis, and P. Hui, "Improving cyber-security via profitable insurance markets," *ACM SIGMETRICS Performance Evaluation Review*, vol. 45, no. 4, pp. 7–15, 2018.

[7] R. Pal, L. Golubchik, K. Psounis, and T. Bandyopadhyay, "On robust estimates of correlated risk in cyber-insured it firms: A first look at optimal ai-based estimates under "small" data," *ACM Transactions on Management Information Systems (TMIS)*, vol. 10, no. 3, pp. 1–18, 2019.

[8] R. Anderson and T. Moore, "Information security: where computer science, economics and psychology meet," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1898, pp. 2717–2727, 2009.

[9] S. Shetty, M. McShane, L. Zhang, J. P. Kesan, C. A. Kamhoua, K. Kwiat, and L. L. Njilla, "Reducing informational disadvantages to improve cyber risk management," *The Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 43, no. 2, pp. 224–238, 2018.

[10] K. M. Gatzlaff and K. A. McCullough, "The effect of data breaches on shareholder wealth," *Risk Management and Insurance Review*, vol. 13, no. 1, pp. 61–83, 2010.

[11] D. M. Pooser, M. J. Browne, and O. Arkhangelska, "Growth in the perception of cyber risk: evidence from us p&c insurers," *The Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 43, no. 2, pp. 208–223, 2018.

[12] S. S. Wang, "Integrated framework for information security investment and cyber insurance," *Pacific-Basin Finance Journal*, vol. 57, p. 101173, 2019.

[13] S. Romanosky, L. Ablon, A. Kuehn, and T. Jones, "Content analysis of cyber insurance policies: how do carriers price cyber risk?," *Journal of Cybersecurity*, vol. 5, no. 1, p. tyz002, 2019.

[14] U. Franke, "The cyber insurance market in sweden," *Computers & Security*, vol. 68, pp. 130–144, 2017.

[15] M. Eling and W. Schnell, "Extreme cyber risks and the nondiversification trap," 2020.

[16] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu, "Modeling and predicting cyber hacking breaches," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2856–2871, 2018.

[17] T. Maillart and D. Sornette, "Heavy-tailed distribution of cyber-risks," *The European Physical Journal B*, vol. 75, no. 3, pp. 357–364, 2010.

[18] D. Kessler, "Why (re) insurance is not systemic," *Journal of Risk and Insurance*, vol. 81, no. 3, pp. 477–488, 2014.

[19] M. LI, "Scor paper," 2018.

[20] J. W. Welburn and A. Strong, *Systemic cyber risk and aggregate impacts*. RAND, 2019.

[21] S. Analytics, "Global connected and IoT device forecast update." https://www.strategyanalytics.com/access-services/devices/connected-home/consumer-electronics/reports/report-detail/global-connected-and-iot-device-forecast-update.

[22] Ericsson, "Internet of Things forecast." https://www.ericsson.com/en/mobility-report/internet-of-things-forecast.

[23] I. Markit, "The internet of things: a movement, not a market," *Englewood, CO: IHS Markit. Accessed December*, vol. 28, p. 2018, 2017.

[24] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*. John Wiley & Sons, 2011.

[25] V. M. Zolotarev, *One-dimensional stable distributions*, vol. 65. American Mathematical Soc., 1986.

[26] R. Pal, Z. Huang, X. Yin, M. Liu, S. Lototsky, and J. Crowcroft, "Sustainable catastrophic cyber-risk management in iot societies," in *To Appear in 2010 Winter Simulation Conference*, IEEE/INFORMS, 2020.

[27] R. Pal, K. Psounis, A. Kumar, J. Crowcroft, P. Hui, J. Kelly, A. Chatterjee, L. Golubchik, and S. Tarkoma, "When are cyber blackouts in modern service networks likely? a network oblivious theory on cyber (re) insurance feasibility," *ACM Transactions on Management Information Systems (TMIS)*, vol. 11, no. 4, 2020.

[28] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Management Science*, vol. 45, no. 2, pp. 208–224, 1999.

[29] P. J. Danaher and M. S. Smith, "Modeling multivariate distributions using copulas: applications in marketing," *Marketing Science*, vol. 30, no. 1, pp. 4–21, 2011.

[30] D. D. Mari and S. Kotz, *Correlation and dependence*. World Scientific, 2001.

[31] R. Nelsen, J. Quesada-Molina, and a. J. Rodriguez-Lallena, "Bivariate copulas with cubic sections," *Journal of Nonparametric Statistics*, vol. 7, no. 3, pp. 205–220, 1997.

[32] M. Ibragimov, R. Ibragimov, and J. Walden, *Heavy-tailed distributions and robustness in economics and finance*, vol. 214. Springer, 2015.

[33] M. Eling and J. Wirfs, "What are the actual costs of cyber risk events?," *European Journal of Operational Research*, vol. 272, no. 3, pp. 1109–1119, 2019.

[34] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[35] J. Nešlehová, P. Embrechts, and V. Chavez-Demoulin, "Infinite mean models and the lda for operational risk," *Journal of Operational Risk*, vol. 1, no. 1, pp. 3–25, 2006.

[36] P. Embrechts, A. McNeil, and D. Straumann, "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, vol. 1, pp. 176–223, 2002.

[37] P. Embrechts, J. Nešlehová, and M. V. Wüthrich, "Additivity properties for value-at-risk under archimedean dependence and heavy-tailedness," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 164–169, 2009.

[38] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. Berlin, Germany: Springer-Verlag, 2013.

[39] B. Edwards, S. Hofmeyr, and S. Forrest, "Hype and heavy tails: A closer look at data breaches," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 3–14, 2016.

[40] S. Wheatley, T. Maillart, and D. Sornette, "The extreme risk of personal data breaches and the erosion of privacy," *The European Physical Journal B*, vol. 89, no. 1, pp. 1–12, 2016.

[41] H. Herath and T. Herath, "Copula-based actuarial model for pricing cyber-insurance policies," *Insurance markets and companies: analyses and actuarial computations*, vol. 2, no. 1, pp. 7–20, 2011.

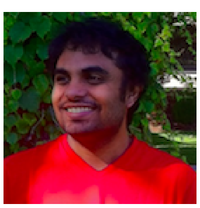
[42] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and S. K. Sadhukhan, "Cyber-risk decision models: To insure it or not?," *Decision Support Systems*, vol. 56, pp. 11–26, 2013.

[43] R. Böhme and G. Kataria, "Models and measures for correlation in cyber-insurance.," in *WEIS*, 2006.

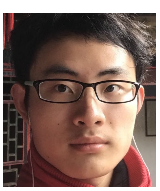
[44] M. Xu, L. Hua, and S. Xu, "A vine copula model for predicting the effectiveness of cyber defense early-warning," *Technometrics*, vol. 59, no. 4, pp. 508–520, 2017.

[45] M. Xu and L. Hua, "Cybersecurity insurance: Modeling and pricing," *North American Actuarial Journal*, vol. 23, no. 2, pp. 220–249, 2019.

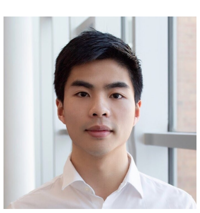
[46] C. Peng, M. Xu, S. Xu, and T. Hu, "Modeling multivariate cybersecurity risks," *Journal of Applied Statistics*, vol. 45, no. 15, pp. 2718–2740, 2018.



Ranjan Pal is a junior faculty member of ECE at University of Michigan Ann Arbor. His research interest lies in engineering robust cyber-security and information privacy solutions using decision and the applied mathematical sciences. He received his PhD in Computer Science from USC's Viterbi School of Engineering, and was a postdoctoral fellow in the CS department, and the Center for Mathematical Sciences at the University of Cambridge. Ranjan is an Associate Editor of IEEE Networking Letters, ACM Transactions on MIS, and a member of IEEE.



Ziyuan Huang is working towards the bachelor's degree in electrical and computer engineering at the University of Michigan Ann Arbor. His main research interests include financial mathematics and cyber-security. He is a student member of the IEEE and the ACM.



Xinlong Yin is working towards the bachelor's degree in electrical and computer engineering at the University of Michigan Ann Arbor. His main research interests include applications of machine learning and statistics to cyber-security. He is a student member of the IEEE and the ACM.

Sergey Lototsky is a Professor of Mathematics at University of Southern California (USC). He received his PhD in Applied Mathematics from USC, was a CLE Moore Instructor at MIT, and a Sloan Research Fellow. His research interests include probability theory and stochastic processes, stochastic partial differential equations, statistical inference for stochastic processes, and mathematical finance. He is an Associate Editor of Stochastics and Partial Differential Equations: Analysis and Computations, and the SIAM Journal on Mathematical Analysis.

Swades De is a Professor of Electrical Engineering at Indian Institute of Technology Delhi. His current directions broadband wireless access and IoT communications. He received the Ph.D. degree from the State University of New York at Buffalo, NY, USA, in 2004. He is a Senior Member of the IEEE, a Fellow of the Indian National Academy of Engineering, and a Fellow of the National Academy of Sciences, India. Swades is an editor of IEEE Communications Letters, IEEE Networking Letters, and IEEE Transactions on Vehicular Technology.

Sasu Tarkoma is a Professor of Computer Science with the University of Helsinki and the Head of the Department of Computer Science. He is also affiliated with the Helsinki Institute for Information Technology. His research interests are Internet technology, distributed systems, data analytics, and mobile and ubiquitous computing. He was a recipient of several best paper awards and mentions, for example, IEEE PerCom, ACM CCR, and ACM OSR. He is an Editorial Board Member of the Computer Networks journal, and a senior member of the IEEE.

Mingyan Liu is an entrepreneur and the Peter and Evelyn and Fuss Chair Professor of Electrical and Computer Engineering at University of Michigan Ann Arbor. She received her PhD in Electrical Engineering from University of Maryland College Park. One of her major current research interests lie in incentive design and experimental data science for cybersecurity and privacy. She was a cofounder of the cybersecurity scoring startup Quadmetrics in 2014 that got acquired by FICO in 2016. She is a Fellow of the IEEE and a member of the ACM.

Jon Crowcroft is the Marconi Professor of Communications Systems in the Computer Laboratory at the University of Cambridge. His current active research areas are Opportunistic Communications, Social Networks, Privacy Preserving Analytics, and techniques and algorithms to scale infrastructure-free mobile systems. Since 2016, he has been Programme Chair at the Alan Turing Institute. He is a Fellow the Royal Society, a Fellow of the ACM, a Fellow of the British Computer Society, a Fellow of the Royal Academy of Engineering and a Fellow of the IEEE.

Nishanth Sastry is a Professor in Computer Science at University of Surrey, UK. He holds a PhD in Computer Science from the University of Cambridge. His current research interests include computer and social networks, computational social science, and data analytics aspects of these two areas. He has been a visiting researcher at the Alan Turing Institute and MIT. His research has been granted nine US patents for work done at IBM. Nishanth is a member of the ACM and a Senior Member of the IEEE.

I. Appendix A - Preliminaries

In this section, we provide (i) the necessary mathematical background, including the definition of Value-at-Risk (VaR) - a commonly used risk measure, the class of heavy-tailed distributions that are mathematically stable, and the basics of *majorization theory* [1] which is essential to our analysis, and (ii) the proofs of the various theorems in the main paper.

A. VaR and Its Properties

Given a risk tolerance q, $0 \le q \le 1$, and a random variable X denoting the severity of losses in our context, the Value-at-Risk (VaR) of X at level q (or the $(1-q)$-quantile) is denoted by $VaR_q(X)$ and defined as:

$$VaR_q(X) = \inf\{z \in \mathbf{R} : P(X > z) \le q\},$$

where $\mathbf{R}$ denotes the real line. This quantity denotes an amount (the VaR), such that the likelihood of losing more than this amount is no more than some tolerance q (e.g., 1%). For this reason the literature is generally interested in the regime $q \le 1/2$.

A function $\mathcal{F} : \mathcal{X} \to \mathbf{R}$, where $\mathcal{X}$ is a linear space of r.v.'s defined on a probability space, is called a coherent risk measure [2][3] if it satisfies the following axioms:

- (Monotonicity) $\mathcal{F}(X) \ge \mathcal{F}(Y)$ if $X \ge Y$ a.s., for all $X, Y \in \mathcal{X}$.
- (Translation Invariance) $\mathcal{F}(X + a) = \mathcal{F}(X) + a$ for all $X \in \mathcal{X}$ and any $a \in \mathbf{R}$.
- (Positive Homogeneity) $\mathcal{F}(\lambda X) = \lambda \mathcal{F}(X)$ for all $X \in \mathcal{X}$ and any $\lambda \ge 0$.
- (Subadditivity) $\mathcal{F}(X + Y) \le \mathcal{F}(X) + \mathcal{F}(Y)$ for all $X, Y \in \mathcal{X}$.

Here we assume that $\mathcal{X}$ contains all degenerate r.v.'s $X = a \in \mathbf{R}$. It is easy to verify that $VaR_q(X)$ always satisfies monotonicity, translation invariance, and positive homogeneity, but generally fails to satisfy subadditivity, see counterexamples in [2][3]. For this reason the VaR measure is generally considered non-coherent[1].

B. Basics of Heavy-Tailed Stable Distributions

We will limit ourselves to a specific but popular family of heavy tailed distributions whose tails decline parametrically as a polynomial function of some $\alpha > 0$. Specifically, within this context, a random variable X is said to have a *heavy-tailed* distribution if $0 < c \le \frac{P(|X|>x)}{x^{-\alpha}} \le C < \infty$ for large x, for constants c and C. This is also denoted as $P(|X| > x) \asymp x^{-\alpha}$. Such distributions have finite statistical moments $\mathrm{E}[|X|^p]$ if the order $p < \alpha$, and infinite statistical moments for $p \ge \alpha$.

A distribution is said to be *stable* if a linear combination of two independent random variables with this distribution has the same distribution, up to location and scale parameters [6][4][5]. The *Normal*, *Cauchy* and the *Levy* distributions are the only stable distributions for which closed form expressions exist, and consequently are often used in analysis for their tractability. Of these three, Cauchy and Levy are heavy-tailed; they are also stable for $\alpha \in (0, 2)$. Therefore, we shall focus on these distributions in this paper. Another attractive property with respect to heavy-tailed stable distributions is the applicability of the central limit theorem for such (non) IID random variables with undefined variance. This generalization (due to Gnedenko and Kolmogorov [4]) states that the sum of a number of random variables with symmetric distributions with infinite variances and having power-law tails (Paretian tails), will tend to a stable distribution as the number of summands increase.

Specifically, we will denote by $S_\alpha(\sigma, \beta, \mu)$, $0 < \alpha \le 2$, the distribution of a stable, heavy-tailed r.v. X. Here α is also referred to as the characteristic exponent (or index of stability, which characterizes the heaviness or the rate of decay of the tail); $\sigma > 0$ is the scale parameter, which is a generalization of the concept of standard deviation (it coincides with the standard deviation in the special case of Gaussian distributions ($\alpha = 2$)); $\beta \in [-1, 1]$ is the symmetry index that characterizes the skewness of the distribution – a stable distributions with $\beta = 0$ are symmetric about the location parameter μ. In order to uniquely determine the distribution $S_\alpha(\sigma, \beta, \mu)$ of a random variable X, we would need its characteristic function (c.f.) that always exists [7], and is defined as:

$$\mathrm{E}[e^{i\theta X}] = \begin{cases} e^{i\mu\theta - \sigma^\alpha |\theta|^\alpha (1 - i\beta \mathrm{sign}(\theta) \tan(\frac{\pi\alpha}{2}))}, \text{ if } \alpha \ne 1, \\ e^{i\mu\theta - \sigma|\theta|^\alpha (1 + (\frac{2}{\pi} i\beta \mathrm{sign}(\theta) \ln|\theta|)}, \text{ if } \alpha = 1, \end{cases} \tag{1}$$

where $\theta \in \mathbf{R}$, $i^2 = -1$ and $\mathrm{sign}(\theta)$ is the sign of x defined by $\mathrm{sign}(\theta) = 1$ if $\theta > 0$; $\mathrm{sign}(0) = 0$; and $\mathrm{sign}(\theta) = -1$ otherwise.

In what follows, we write $X \sim S_\alpha(\sigma, \beta, \mu)$, if the random variable X has the stable distribution $S_\alpha(\sigma, \beta, \mu)$. Throughout this paper, we will also limited ourselves to the case of $\mu = 0$ without loss of generality.

Now consider IID stable r.v's $X_i \sim S_\alpha(\sigma, \beta, 0)$, such that $\beta \ne 0$ for $\alpha = 1$. These distributions are called strictly stable. If $X_i \sim S_\alpha(\sigma, \beta, 0), \alpha \in (0, 1) \cup (1, 2]$ are strictly stable r.v.'s, for all $a_i \ge 0$, $i = 1, ..., n$, we have

$$\frac{\sum_{i=1}^n a_i X_i}{(\sum_{i=1}^n a_i^\alpha)^{\frac{1}{\alpha}}} \sim S_\alpha(\sigma, \beta, 0). \tag{2}$$

[1] It is shown in [4] and [5] that mathematical constructions can be designed under small q and with $\alpha < 1$ that makes VaR coherent; however, such constructions are rare in practice. The also commonly used alternative risk measure, conditional VaR (the average of worst losses - also called expected shortfall or Average VaR), or CVaR, is coherent by the above definition. However, CVaR is the *average* of the worst losses of a (cyber-risk) portfolio (i.e., for $q \in (0, 1]$, $\mathrm{CVaR}_q(Y) = 1/q \int_{1-q}^1 VaR_{1-\tau}(Y) d\tau$), and subsequently requires existence of the statistical first moments of the loss distribution, which may not be true of catastrophic cyber-risks. For this reason we will limit ourselves to the VaR measure in this letter.

C. Families of Distribution Convolutions

A fundamental operation for a cyber-risk aggregator is the convolution (aggregation) of individual risk distributions. In this section, we define and mention some salient features (where applicable) of various classes/families of distribution convolutions.

Families Related to Convolution of Symmetric Stable Distributions - For $0 \le r < 2$, we denote by $\overline{\mathcal{CS}}(r)$ the class of cyber-risk distributions which are convolutions of individually symmetric stable cyber-risk distributions $S_\alpha(\sigma, 0, 0)$ with indices of stability $\alpha \in [r, 2)$ and $\sigma > 0$. That is, $\overline{\mathcal{CS}}(r)$ consists of cyber-risk distributions of r.v.'s X for which, with some $k \ge 1, X = Y_1 + ... + Y_k$, where $Y_i, i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in [r, 2), \sigma_i > 0, i = 1, ..., k$.

For $0 \le r \le 2$, we denote by $\underline{\mathcal{CS}}(r)$ the class of cyber-risk distributions which are convolutions of individually symmetric and stable cyber-risk distributions $S_{\alpha_i}(\sigma_i, 0, 0)$ with indices of stability $\alpha_i \in (0, r)$ and $\sigma_i > 0$. That is, $\underline{\mathcal{CS}}(r)$ consists of cyber-risk distributions of r.v.'s X for which, with some $k \ge 1, X = Y_1 + ... + Y_k$, where $Y_i, i = 1, ..., k$, are independent r.v.'s such that $Y_i \sim S_{\alpha_i}(\sigma_i, 0, 0)$, $\alpha_i \in (0, r), \sigma_i > 0, i = 1, ..., k$.

Salient Features of Convolution Families - The classes $\overline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$ are mathematically *closed* under convolutions - *a powerful property contributing to tractable analysis of cyber-risks in these families.* A linear combination of independent stable r.v.'s with the same characteristic exponent α also has a stable distribution with the same α. However, in general, this does not hold in the case of convolutions of stable distributions with different indices of stability. Therefore, the class $\overline{\mathcal{CS}}(r)$ of convolutions of symmetric stable distributions with different indices of stability $\alpha \in (r, 2]$ is wider than the class of all symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (r, 2]$ and $\sigma > 0$. Similarly, the class $\underline{\mathcal{CS}}(r)$ is wider than the class of all symmetric stable distributions $S_\alpha(\sigma, 0, 0)$ with $\alpha \in (0, r)$ and $\sigma > 0$.

By definition, for $0 < r_1 < r_2 \le 2$, the following inclusions hold: $\underline{\mathcal{CS}}(r_2) \subset \underline{\mathcal{CS}}(r_1)$ and $\underline{\mathcal{CS}}(r_1) \subset \underline{\mathcal{CS}}(r_2)$. Cauchy distributions $S_1(\sigma, 0, 0)$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}(1)$ and $\overline{\mathcal{CS}}(1)$ Similarly, for $r \in (0, 2)$, stable distributions $S_r(\sigma, 0, 0)$ with the characteristic exponent $\alpha = r$ are at the dividing boundary between the classes $\underline{\mathcal{CS}}(r)$ and $\underline{\mathcal{CS}}(r)$. More precisely, the Cauchy distributions $S_1(\sigma, 0, 0)$ are the only ones that belong to all the classes $\underline{\mathcal{CS}}(r)$ with $r > 1$ and all the classes $\overline{\mathcal{CS}}(r)$ with $r < 1$. Stable distributions $S_r(\sigma, 0, 0)$ are the only ones that belong to all the classes $\underline{\mathcal{CS}}(r')$ with $r' > r$ and all the classes $\overline{\mathcal{CS}}(r')$ with $r' < r$. The properties of stable distributions discussed herein imply that the p-th absolute moments $E[|X|^p]$ of a r.v. X $\sim \overline{\mathcal{CS}}(r)$, $r \in (0, 2)$, are finite if $p \le r$. However, all the r.v.'s $X \sim \underline{\mathcal{CS}}(r), r \in (0, 2]$ have infinite moments of order r: $E[|X|^r] = \infty$. In particular, the distributions of r.v.'s X from the class $\underline{\mathcal{CS}}(1)$ are extremely heavy-tailed (representing catastrophic cyber-risks) in the sense that their first moments are infinite: $E[|X|] = \infty$.

D. Basics of Majorization Theory

A vector[2] with n components $w \in \mathbf{R}_+^n$ is said to be majorized by a vector $v \in \mathbf{R}^n$, written as $w \prec v$, if $\sum_{i=1}^k w_{[i]} \le \sum_{i=1}^k v_{[i]}$, $k = 1, \cdots, n-1$, and $\sum_{i=1}^n a_{[i]} = \sum_{i=1}^n b_{[i]}$, where $w_{[1]} \ge \cdots \ge w_{[n]}$ and $v_{[1]} \ge \cdots \ge v_{[n]}$ denote the elements of w and v in decreasing order, respectively. The relation $w \prec v$ implies that the components of w are *less* diverse than those of v (see [1]). For instance, it is easy to see that the following holds:

$$\left(\sum_{i=1}^n \frac{w_i}{n}, \cdots, \sum_{i=1}^n \frac{w_i}{n} \right) \prec (w_1, \cdots, w_n) \prec \left(\sum_{i=1}^n w_i, 0, ...0 \right), \quad \forall w \in \mathbf{R}_+^n \ . \tag{3}$$

In particular, we have the following for two vectors in $\mathbf{R}_+^{n+1}$, $n \ge 1$:

$$\left(\frac{1}{n+1}, \cdots, \frac{1}{n+1}, \frac{1}{n+1} \right) \prec \left(\frac{1}{n}, \cdots, \frac{1}{n}, 0 \right) \ . \tag{4}$$

It is also immediate that if $w \prec v$, then the same is true for their respective permutations: $(w_{\pi(1)}, \cdots, w_{\pi(n)}) \prec (v_{\pi(1)}, \cdots, v_{\pi(n)})$ for all permutations π of the set $\{1., .., n\}$.

A function $\phi : \mathbf{R}_+^n \to \mathbf{R}$ is called *Schur-convex* (resp. Schur-concave) [8] if $(w \prec v) \implies (\phi(w) \le \phi(v))$ (resp. $(w \succ v) \implies (\phi(w) \ge \phi(v))$, $\forall w, v \in \mathbf{R}_+^n$. If the inequalities are strict whenever $a \prec b$ and a is not a permutation of b, then ϕ is said to be strictly Schur-convex (resp. strictly Schur-concave). Evidently, if $\phi : \mathbf{R}_+^n \to \mathbf{R}$ is Schur-convex or Schur-concave, then $\forall w \in \mathbf{R}_+^n$, we have:

$$\phi(w_1, \cdots, w_n) = \phi(w_{\pi(1)}, \cdots, w_\pi(n)), \tag{5}$$

where π is any permutation of the set $\{1, ..., n\}$. Examples of strictly Schur-convex functions $\phi : \mathbf{R}_+^n \to \mathbf{R}$ are given by $\phi_\alpha(w_1, \cdots, w_n) = \sum_{i=1}^n w_i^\alpha$ for $\alpha > 1$. The functions $\phi_\alpha(w_1, \cdots, w_n)$ are strictly Schur-concave for $\alpha < 1$ (see Proposition 3.C.1.a in [9]).

Consider a portfolio of cyber-risks $X_1, ..., X_n$ with weights $w = (w_1, ..., w_n) \in \mathbf{R}_n^+$ denoting the fraction of each risk the portfolio is exposed to, i.e., the fraction of each risk an insurer is responsible for covering. The aggregate risk is denoted by

$$Z_{\mathbf{w}} = \sum_{i=1}^n w_i X_i \ . \tag{6}$$

[2]In this letter, we denote a vector $(v_{[i]} ... v_{[n]})$ with n components by v.

Denote by $\mathcal{I}_n = \{w = (w_1, ..., w_n) : w_i \geq 0, i = 1, \cdots, n, \sum_{i=1}^n w_i = 1\}$ the simplex of all vectors where weights sum to 1. Define two special vectors $\underline{w} = (\frac{1}{n}, \frac{1}{n}, ...\frac{1}{n}) \in \mathcal{I}_n$ and $\overline{w} = (1, 0, ..., 0) \in \mathcal{I}_n$. Given the same set of risks, the theory of majorization suggests that $\underline{w} \prec \overline{w}$, and a portfolio based on the latter weights is more diverse. This notion of diversity is in a way the opposite of what one might consider to be the variability among the weights: the more diverse $\overline{w}$ has the least varied weights (consisting of a single risk) within $\mathcal{I}_n$, while the less diverse $\underline{w}$ has more varied weights (equally spread over n risks). Similarly, Eqn (4) suggests that $(\frac{1}{n+1}, \cdots, \frac{1}{n+1}) \in \mathcal{I}_{n+1}$ has more varied weights than $(\frac{1}{n}, \cdots, \frac{1}{n}, 0) \in \mathcal{I}_{n+1}$ since it contains an additional non-redundant cyber-risk X_{n+1}, but the former is actually less diverse using majorization.

A simple example demonstrating the conventional wisdom that portfolio variation is preferable is given by the case with normally distributed risks. Let $X_1, ..., X_n \sim S_2(\sigma, 0, 0)$ be i.i.d. symmetric normal r.v.'s. Then, for a portfolio of equal weights $\underline{w} = (\frac{1}{n}, ..., \frac{1}{n})$ we have $Z_{\underline{w}} \sim S_2(\frac{\sigma}{\sqrt{n}}, 0, 0) \sim \frac{1}{\sqrt{n}} X_1$. By positive homogeneity of the VaR, we have for $n \geq 2$:

$$VaR_q(Z_{\underline{w}}) = \frac{1}{\sqrt{n}} VaR_q(X_1) = \frac{1}{\sqrt{n}} VaR_q(Z_{\overline{w}}) < VaR_q(Z_{\overline{w}}) \ . \tag{7}$$

That is, the most varied portfolio with equal weights $\underline{w}$ has lower value at risk than that of the least varied portfolio concentrating on a single risk.

E. Introduction to Copulas and Dependence

In this section provide a basic introduction to copulas and a specific form of copula that accommodates distributions with heavy-tailed marginals.

1) Basics of Copulas: Copulas are joint distributions with uniform marginals. They are useful because given the marginal distributions, they represent the dependence in the joint distribution. Specifically, let $H(x_1, \ldots, x_n)$ and $h(x_1, \ldots, x_n)$ denote the joint distribution and density, respectively, of n random variables $(X_1, \ldots, X_n)$ and suppose that the marginal density and cdf of X_j are $f_j(x_j)$ and $F_j(x_j)$ respectively, $j = 1, \ldots, n$. Then, an n -dimensional copula of $(X_1, \ldots, X_n)$ is a function $C : [0,1]^n \to$ [0,1] such that (a) $C(u_1, \ldots, u_n)$ is increasing in each $u_i, i = 1, \ldots, n$ (b) $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_n) = 0, i = 1, \ldots, n$ (c)$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i, i = 1, \ldots, n$ (d) for any $a_j \leq b_j, j = 1, \ldots, n$

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C(u_{1i_1}, \ldots, u_{ni_n}) \geq 0$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j = 1, \ldots, n$ (e) $H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$, or, for absolutely continuous copulas with density

$$c(u_1, \ldots, u_n), h(x_1, \ldots, x_n) = c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} f(x_i)$$

It is well known that c is uniquely determined if F_j is monotone. The probability integral trans- forms $u_j = F_j(x_j), j = 1, 2$, are the uniform random variables that form the marginals of c. So, equivalently C can be defined as a joint cdf of n random variables, each of which is uniform on $[0,1]$. The fact that we can model F_j separately from modelling the dependence between F_j 's is what makes copulas natural in the analysis of dependent power-law marginals.

A well known property of the copula function is that it is bounded by the Frechet-Hoeffding bounds, which correspond to extreme positive and extreme negative dependence. For a bivariate copula, let x_1 be a fixed increasing function of x_2, then the copula of (x_1, x_2) can be written as $\min(u_1, u_2)$ and this is the upper bound for bivariate copulas. Now let x_1 be a fixed decreasing function of x_2; then the copula of (x_1, x_2) can be written as $\max(u_1 + u_2 - 1, 0)$. So the two extreme cases when diversification does not have any effect (comonotonicity) and when it is always beneficial (countermonotonicity) regardless of the heavy-tailedness are nested within the copula framework. [10] and [11] provide excellent introductions to copulas.

If we return to the two-risk example above, we are interested in how the aggregate loss proba- bility for a diversified portfolio compares to that of a single risk. That is, we are interested in the behavior of

$$\begin{aligned} P\left(\frac{X_1 + X_2}{2} > x\right) &= \iint_{\frac{z_1+z_2}{2} > x} f(z_1; \alpha) f(z_2; \alpha) c(F(z_1; \alpha), F(z_2; \alpha); \gamma) \, dz_1 dz_2 \\ &= E\left\{c(F(\xi_1; \alpha), F(\xi_2; \alpha); \gamma) I\left[\frac{\xi_1 + \xi_2}{2} > x\right]\right\} \end{aligned}$$

where $c(u_1, u_2; \gamma)$ is a copula density parameterized by $\gamma, f(\cdot; \alpha)'$ s are power-law marginal densities, $I[\cdot]$ is the indicator function and ξ_j 's are independent copies of X_j 's. There is no general way to express this in terms of $P(X_1 > x)$ and whether diversification decreases or increases VaR depends on the copula family as well as on the interaction between α and γ. However, there exist classes of copulas for which we can make explicit comparisons.

2) Power-type copulas: We now discuss a class of copula families which will be used in the paper. The class contains copulas that are multiplicative or additive in powers of the margins, or can be approximated using such copulas. We call this class power-type. It is similar but more general than the power copula family and than the polynomial copula family which we discuss below.

The most common family in this class is the Eyraud-Farlie-Gumbel-Morgenstern (EFGM) cop- ula family and its generalizations. The bivariate EFGM copula family can be written as follows

$$C(u_1, u_2) = u_1 u_2 \left[1 + \gamma (1 - u_1)(1 - u_2)\right] \tag{8}$$

where $\gamma \in [-1, 1]$, and its density has the form $c(u_1, u_2) = 1 + g(u_1, u_2)$, where $g(u_1, u_2)$ is an expansion by linear functions $1 - 2u_j, j = 1, 2$. This is a non-comprehensive copula in the sense that it has a limited range of dependence it can accommodate. For example, Kendall's τ of an EFGM copula is restricted to $\left[-\frac{2}{9}, \frac{2}{9}\right]$

The multivariate version of the EFGM copula introduced by [12] has the following form:

$$C(u_1, u_2, \ldots, u_n) = u_1 u_2 \ldots u_n \left[1 + \sum_{c=2}^{n} \sum_{1 \leq i_1 < i_2 < \ldots < i_c \leq n} \gamma_{i_1, i_2, \ldots, i_c} (1 - u_{i_1})(1 - u_{i_2}) \ldots (1 - u_{i_c})\right]$$

where $-\infty < \gamma_{i_1, i_2, \ldots, i_c} < \infty$ are such that $\sum_{c=2}^{n} \sum_{1 \leq i_1 < i_2 < \ldots < i_c \leq n} \gamma_{i_1, i_2, \ldots, i_c} \delta_{i_1} \cdots \delta_{i_c} \geq -1$ for all $\delta_i \in [-1, 1], i = 1, \ldots, n$. This copula family can be viewed as a special case of a wider family of n -dimensional power copulas introduced by [13]. The power copula family can be written as follows

$$C(u_1, \ldots, u_n) = u_1 u_2 \cdots u_n \left[1 + \sum_{c=2}^{n} \sum_{1 \leq i_1 < i_2 < \ldots < i_c \leq n} \gamma_{i_1, i_2, \ldots, i_c} \left(u_{i_1}^{l} - u_{i_1}^{l+1}\right)\left(u_{i_2}^{l} - u_{i_2}^{l+1}\right) \ldots \left(u_{i_c}^{l} - u_{i_c}^{l+1}\right)\right]$$

where $\gamma_{i_1, i_2, \ldots, i_c} \in (-\infty, \infty)$ are such that

$$\sum_{c=2}^{n} \sum_{1 \leq i_1 < i_2 < \ldots < i_c \leq n} \left|\gamma_{i_1, i_2, \ldots, i_c}\right| \leq 1$$

This corresponds to using nonlinear rather than linear functions in the expansion of the copula density function.

Another relevant copula family, of which the EFGM copula in (8) is a special case, is known as a polynomial copula family (see, e.g., [14], p. 74). An order $m (m \geq 4)$ polynomial copula can be written as follows:

$$C(u, v) = uv \left[1 + \sum_{k \geq 1, q \geq 1}^{k+q \leq m-2} \gamma_{kq} \left(u^k - 1\right)\left(v^q - 1\right)\right]$$

$$\text{where } \gamma_{kq} = \frac{\theta_{kq}}{(k+1)(q+1)} \text{ and } 0 \leq \min\left(\sum_{k \geq 1, q \geq 1}^{k+q \leq m-2} q\gamma_{kq}, \sum_{k \geq 1, q \geq 1}^{k+q \leq m-2} k\gamma_{kq}\right) \leq 1$$

One example of this copula family is [15] copula with cubic section, which is written as follows

$$C(u, v) = uv + 2\gamma uv(1 - u)(1 - v)(1 + u + v - 2uv) \tag{9}$$

where $\gamma \in \left[0, \frac{1}{4}\right]$ Several other copula families can be written as approximations of the EFGM copula. For example, it is well known that the EFGM copula is a first-order approximation to the Ali-Mikhail- Haq (AMH) copula family. The AMH copula can be written as follows:

$$C(u_1, \ldots, u_n) = (1 - \gamma)\left[\prod_{i=1}^{n}\left(\frac{1 - \gamma}{u_i} + \gamma\right) - \gamma\right]^{-1}$$

where $\gamma \in [-1, 1]$ A less known result is that the Plackett and the Frank copula families are first order Taylor approximations of the EFGM copula at independence (see, e.g., [11], p.100,133). The $n-$ variate Frank copula, which is comprehensive, radially symmetric and Archimedian, can be written as follows

$$C(u_1, \ldots, u_n) = \log_\gamma \left[1 + \frac{\prod_{i=1}^{n}\left(\gamma^{u_i} - 1\right)}{(\gamma - 1)^{n-1}}\right]$$

where $\gamma \geq 0$ The n -variate Plackett copula, which is also comprehensive, is rarely discussed in the literature unless $n = 2$, in which case it has the following form:

$$C(u_1, u_2) = \frac{1}{2(\gamma - 1)}\left[1 + (\gamma - 1)(u_1 + u_2) - \sqrt{\left[1 + (\theta - 1)(u_1 + u_2)\right]^2 - 4\gamma(\gamma - 1)u_1 u_2}\right]$$

where $1 \neq \gamma > 0$. However, a way to generalize to $n > 2$ is presented by [16]. It is also worth mentioning that for all the three copula families, there exist improved second-order approximations (see, e.g., [11], p.83).

An interesting set of approximation results are given by [15], [17] and [18]. [15] provide a generalization of the bivariate EFGM copula using cubic terms as in (9) and show that it can be used to approximate some well-known families of copulas, both symmetric and not, such as the copulas of [19] and [20], as well as the Sarmanov copula. They also show that copulas in (9) are second- degree Maclaurin approximations to members of the Frank and Plackett copula families.

[17] studies the power series class of copulas, obtained as weighted geometric means of the EFGM and AMH copulas, and shows that the Gumbel-Barnett and Cuadras-Auge copulas can be expressed as first-order approximations to that class. [18] provide approximations of the tail-dependent Clayton-Oakes copula, which also have the form of a power- type generalization of the EFGM copula.

II. Appendix B - Proofs

Proof of Theorem 3.1: Since v is not a permutation of w, we have $\sum_{i=1}^n v_i \neq 0$ and $\sum_{i=1}^n w_i \neq 0$. This together with the fact that $X_1, \dots, X_n$ are IID according to $S_\alpha(\sigma,0,0), \alpha \in (0,1]$, means that, using Eqn (2), we have

$$Z_v = \sum_{i=1}^n v_i X_i \ =^d \ (\sum_{i=1}^n v_i^\alpha)^{1/\alpha} X_1. \tag{10}$$

Using positive homogeneity of VaR, we thus obtain

$$VaR_q(Z_v) = \left(\sum_{i=1}^n v_i^\alpha\right)^{1/\alpha} VaR_q(X_1). \tag{11}$$

Proposition 3.C.1.a in [9] implies that the function $h(v_1, \dots, v_n) = \sum_{i=1}^n v_i^\alpha$ is strictly Schur-concave in $v \in \mathbf{R}_+^\mathbf{n}$ if $\alpha < 1$.

Therefore, we have $\sum_{i=1}^n w_i^\alpha \ < \ \sum_{i=1}^n v_i^\alpha$, if $\alpha < 1$. This, together with the Schur-concavity definition given in Section I-D implies that for $\alpha < 1$,

$$VaR_q(Z_v) = \left(\sum_{i=1}^n v_i^\alpha\right)^{1/\alpha} VaR_q(X_1) < \left(\sum_{i=1}^n w_i^\alpha\right)^{1/\alpha} VaR_q(X_1) = VaR_q(Z_w), \text{if } 0 < \alpha < 1 \tag{12}$$

Thus we prove the Theorem 2.1. ∎

Proof of Theorem 3.2: Let $\alpha \in (0,2]$, $\sigma > 0$, and let $v = (v_1, \dots, v_n) \in \mathbf{R}_+^n$ and $\omega = (\omega_1, \dots, \omega_n) \in \mathbf{R}_+^n$ be two vectors of portfolio weights such that $(v_1, \dots, v_n) \prec (\omega_1, \dots, \omega_n)$ and $(v_1, \dots, v_n)$ is not a permutation of $(\omega_1, \dots, \omega_n)$ (clearly, $\sum_{i=1}^n v_i \neq 0$ and $\sum_{i=1}^n \omega_i \neq 0$). Let $X_1, \dots, X_n$ be i.i.d. risks such that $X_i \sim S_\alpha(\sigma,0,0)$, $i = 1, \dots, n$. It follows that if $c = (c_1, \dots, c_n) \in \mathbf{R}_+^\mathbf{n}$, $\sum_{i=1}^n c_i \neq 0$, then $Z_c = \sum_{i=1}^n c_i X_i \ =^d \ (\sum_{i=1}^n c_i^\alpha)^{1/\alpha} X_1$. Using positive homogeneity of the value at risk (property a3 in Section 3), we thus obtain that, for all $q \in (0, 1/2)$,

$$VaR_q(Z_c) = VaR_q(X_1) \left(\sum_{i=1}^n c_i^\alpha\right)^{1/\alpha}. \tag{13}$$

Proposition 3.C.1.a in Marshall & Olkin (1979) implies that the function $h(c_1, \dots, c_n) = \sum_{i=1}^n c_i^\alpha$ is strictly Schur-convex in $(c_1, \dots, c_n) \in \mathbf{R}_+^\mathbf{n}$ if $\alpha > 1$ and is strictly Schur-concave in $(c_1, \dots, c_n) \in \mathbf{R}_+^\mathbf{n}$ if $\alpha < 1$. Therefore, we have $\sum_{i=1}^n v_i^\alpha \ < \ \sum_{i=1}^n \omega_i^\alpha$, if $\alpha > 1$. This implies that, for all $q \in (0, 1/2)$,

$$VaR_q(Z_v) < VaR_q(Z_\omega) \tag{14}$$

if $\alpha > 1$. This completes the proof of parts (i) of Theorem 3.2 in the case of i.i.d. stable risks $X_i \sim S_\alpha(\sigma,0,0), i = 1, \dots, n$.

Let now $X_1, \dots, X_n$ be i.i.d. risks such that $X_i \sim \overline{\mathcal{CSLC}}$, $i = 1, \dots, n$. By definition, $X_i = \gamma Y_{i0} + \sum_{j=1}^k Y_{ij}$, $i = 1, \dots, n$, where $\gamma \in \{0,1\}$, $k \geq 0$, $Y_{i0} \sim \mathcal{LC}$, $i = 1, \dots, n$, and $(Y_{1j}, \dots, Y_{nj})$, $j = 0, 1, \dots, k$, are independent vectors with i.i.d. components such that $Y_{ij} \sim S_{\alpha j}(\sigma_j, 0, 0)$, $\alpha \in (1,2]$, $\sigma_j > 0$, $i = 1, \dots, n, j = 1, \dots, k$. From results in Proschan (1965) for tail probabilities of log-concavely distributed r.v.'s it follows that, for all $q \in (0, 1/2)$ and all $j = 0, 1, \dots, k$, $VaR_q(\sum_{i=1}^n v_i Y_{ij}) < VaR_q(\sum_{i=1}^n \omega_i Y_{ij})$. The densities of the r.v.'s Y_{i0}, $i = 0, 1, \dots, n$, are symmetric and unimodal. In addition, the density of the r.v.'s Y_{ij}, $i = 1, \dots, n$, $j = 1, \dots, k$, are symmetric and unimodal. We thus conclude that the densities of the r.v.'s $\sum_{i=1}^n v_i Y_{ij}$ and $\sum_{i=1}^n \omega_i Y_{ij}$, $j = 0, 1, \dots, k$, are symmetric and unimodal as well. We thus obtain

$$\begin{aligned} VaR_q(Z_c) = VaR_q\left(\sum_{i=1}^n v_i X_i)\right) = VaR_q\left(\gamma \sum_{i=1}^n v_i Y_{i0} + \sum_{j=1}^k \sum_{i=1}^n v_i Y_{ij}\right) < \\ VaR_q\left(\gamma \sum_{i=1}^n \omega_i Y_{i0} + \sum_{j=0}^k \sum_{i=1}^n \omega_i Y_{ij}\right) = VaR_q(\sum_{i=1}^n \omega_i X_i) = VaR_q(Z_\omega) \end{aligned} \tag{15}$$

This completes the proof of part (i) of Theorem 3.2. The bounds in parts (ii) of Theorem 3.2 follow from their parts (i) and majorization comparisons. Thus we prove the Theorem 3.2. ∎

Proof of Theorem 4.1: We have

$$\begin{aligned}
P(X_w > z) &\leq P(Y_w(a) \geq X_w > z) + P(X_w > z, X_w > Y_w(a)) \\
&\leq P(Y_w(a) > z) + P(X_w > Y_w(a)) \\
&= P(Y_w(a) > z) + P(X_1 > a \cup X_2 > a \cup \cdots X_n > a) \\
&\leq P(Y_w(a) > z) + \sum_{i=1}^{n} P(X_i > a) \\
&= P(Y_w(a) > z) + nP(X_1 > a)\ ,
\end{aligned} \tag{16}$$

where the last inequality is due to the union bound. We get for all $z > 0$

$$P(X_w > z) > P(X_{\bar{w}} > z) + G(w,z) = P(X_1 > z) + G(w,z) = P(X_1 > a) + P(Y_1(a) > z) + G(w,z). \tag{17}$$

Relations (16) and (17) together imply that

$$P(Y_w(a) > z) - P(Y_1(a) > z) > G(w,z) - (n-1)P(X_1 > a). \tag{18}$$

Since $E[|X_1|^r] < \infty$, by the Chebyshev's inequality, we get

$$P(X_1 > a) = \frac{1}{2}P(|X_1| > a) \leq \frac{E[|X_1|^r}{2a^r}. \tag{19}$$

Estimates (17) and (18) gives

$$P(Y_w > z) - P(Y_1(a) > z) > G(w,z) - \frac{(n-1)E[|X_1|^r]}{2a^r}. \tag{20}$$

Under conditions of the theorem, the RHS of (20) is positive. Consequently,

$$P(Y_w(a) > z) > P(Y_1(a) > z), \tag{21}$$

completing the proof. ■.

Proof of Theorem 5.1: We start with the case $n = 2$. Due to independence between ξ_1 and ξ_2, we have that

$$P\left(\frac{\xi_1(\beta_1) + \xi_2(\beta_2)}{2} > z\right) = \beta_1\beta_2 \int_{\frac{s+t}{2} > z} s^{-\beta_1 - 1} t^{-\beta_2 - 1} ds dt \tag{22}$$

Now for non-independent (X_1, X_2) under the EFGM copula, we can write using (22)

$$\begin{aligned}
P\left(\frac{X_1 + X_2}{2} > z\right) &= \int_{\frac{s+t}{2} > z} f_1(s) f_2(t) \left[1 + \gamma\left(1 - 2F_1(s)\right)\left(1 - 2F_2(t)\right)\right] ds dt \\
&= P\left(\frac{\xi_1(\alpha) + \xi_2(\alpha)}{2} > z\right) + \gamma \int_{\frac{s+t}{2} > z} f_1(s) f_2(t) \left(1 - 2F_1(s)\right)\left(1 - 2F_2(t)\right) ds dt \\
&= P\left(\frac{\xi_1(\alpha) + \xi_2(\alpha)}{2} > z\right) + \gamma E\left(1 - 2F_1(\xi)\right)\left(1 - 2F_2(\eta)\right) I\left(\frac{\xi_1 + \xi_2}{2} > z\right)
\end{aligned}$$

where $I(\cdot)$ denotes the indicator function.

Now consider the last term:

$$\begin{aligned}
\int_{\frac{s+t}{2} > z} f_1(s) f_2(t) \left(1 - 2F_1(s)\right)\left(1 - 2F_2(t)\right) ds dt &= \int_{\frac{s+t}{2} > z} \alpha^2 s^{-\alpha-1} t^{-\alpha-1} \left(2s^{-\alpha} - 1\right)\left(2t^{-\alpha} - 1\right) ds dt \\
&= 4\alpha^2 \int_{\frac{s+t}{2} > z} s^{-2\alpha-1} t^{-2\alpha-1} ds dt \\
&- 2\alpha^2 \int_{\frac{s+t}{2} > z}^{2-2\alpha-1} t^{-\alpha-1} ds dt \\
&- 2\alpha^2 \int_{\frac{s+t}{2} > z}^{2} s^{-\alpha-1} t^{-2\alpha-1} ds dt \\
&+ \alpha^2 \int_{\frac{s+t}{2} > z} s^{-\alpha-1} t^{-\alpha-1} ds dt \\
&= 4\alpha^2 I_1 - 2\alpha^2 I_2 - 2\alpha^2 \mathcal{I}_3 + \alpha^2 \mathcal{I}_4
\end{aligned}$$

where $\mathcal{I}_1 = P\left(\frac{\xi_1(2\alpha) + \xi_2(2\alpha)}{2} > z\right), \mathcal{I}_2 = P\left(\frac{\xi_1(2\alpha) + \xi_2(\alpha)}{2} > z\right), \mathcal{I}_3 = P\left(\frac{\xi_1(\alpha) + \xi_2(2\alpha)}{2} > z\right)$ and $\mathcal{I}_4 = P\left(\frac{\xi_1(\alpha) + \xi_2(\alpha)}{2} > z\right)$

Thus we obtain

$$\begin{aligned}P\left(\frac{X+Y}{2}>z\right)=&\left(1+\gamma\alpha^2\right)P\left(\frac{\xi_1(\alpha)+\xi_2(\alpha)}{2}>z\right)\\&-2\gamma\alpha^2P\left(\frac{\xi_1(\alpha)+\xi_2(2\alpha)}{2}>z\right)\\&-2\gamma\alpha^2P\left(\frac{\xi_1(2\alpha)+\xi_2(\alpha)}{2}>z\right)\\&+4\gamma\alpha^2P\left(\frac{\xi_1(2\alpha)+\xi_2(2\alpha)}{2}>z\right)\end{aligned}$$

It is a well-known result in the power law literature (see, among others, Corollary 1.3 .2 in Embrechts et al. 1997) that, asymptotically as $z\to\infty$

$$P\left(\frac{\xi_1(\beta)+\xi_2(\beta)}{2}>z\right)\sim 2P\left(\xi_1(\beta)>2z\right)\sim 2^{1-\beta}z^{-\beta} \tag{23}$$

for all $\beta>0$. In addition, if $\beta_1<\beta_2$, then

$$P\left(\frac{\xi_1\left(\beta_1\right)+\xi_2\left(\beta_2\right)}{2}>z\right)\sim P\left(\xi_1\left(\beta_1\right)>2z\right)\sim 2^{-\beta_1}z^{-\beta_1} \tag{24}$$

It follows from (23)-(24) that, as $z\to\infty$

$$\begin{aligned}P\left(\frac{X+Y}{2}>z\right)&\sim\left(1+\gamma\alpha^2\right)2^{1-\alpha}z^{-\alpha}-2\gamma\alpha^2 2^{1-\alpha}z^{-\alpha}+4\gamma\alpha^2 2^{1-2\alpha}z^{-2\alpha}\\&\sim\left(1-\gamma\alpha^2\right)2^{1-\alpha}z^{-\alpha}\\&\sim P\left(\frac{\xi_1(\alpha)+\xi_2(\alpha)}{2}>z\right)\end{aligned}$$

We now provide a generalization for any n. Let $X_1,X_2,\ldots,X_n$ have a multidimensional EFGM copula

$$C\left(u_1,u_2,\ldots,u_n\right)=u_1u_2\ldots u_n\left[1+\sum_{c=2}^{n}\sum_{1\le i_1<i_2<\ldots<i_c\le n}\gamma_{i_1,i_2,\ldots,i_c}\left(1-u_{i_1}\right)\left(1-u_{i_2}\right)\ldots\left(1-u_{i_c}\right)\right] \tag{25}$$

where $\gamma_{i_1,i_2,\ldots,i_c}$ are real constants satisfying certain inequalities that guarantee that (25) represents a proper copula.

Let $X_1,X_2,\ldots,X_n$ have power law distributions with the same parameter $\alpha>0$. It follows from (25) that the joint cdf of $X_1,X_2,\ldots,X_n$ has the form

$$\begin{aligned}F\left(x_1,x_2,\ldots,x_n\right)=&F_1\left(x_1\right)F_2\left(x_2\right)\ldots F_n\left(x_n\right)\\&\times\left[1+\sum_{c=2}^{n}\sum_{1\le i_1<i_2<\ldots<i_c\le n}\gamma_{i_1,i_2,\ldots,i_c}\left(1-F_{i_1}\left(x_{i_1}\right)\right)\left(1-F_{i_2}\left(x_{i_2}\right)\right)\ldots\left(1-F_{i_c}\left(x_{i_c}\right)\right)\right]\end{aligned}$$

Let, $\xi_1\left(\beta_1\right),\xi_2\left(\beta_2\right),\ldots,\xi_n\left(\beta_n\right)$ denote the independent random variables with power law distri- butions with tail indices $\beta_1,\beta_2,\ldots,\beta_n$, respectively. That is,

$$P\left(\xi_i\left(\beta_i\right)>x\right)=x^{-\beta_i}$$

$x\ge 1, i=1,2,\ldots,n,$. In particular, $\xi_1(\alpha),\xi_2(\alpha),\ldots,\xi_n(\alpha)$ are independent copies of $X_1,X_2,\ldots,X_n$ Then, it follows that

$$\begin{aligned}P\left(\sum_{i=1}^{n}X_i>zn\right)=&P\left(\sum_{i=1}^{n}\xi_i(\alpha)>zn\right)\\&+\sum_{c=2}^{n}\sum_{i\le i_1<i_2<\ldots<i_c\le n}\gamma_{i_1,i_2,\ldots,i_c}\\&\times E\left[\left(1-2F_{i_1}\left(\xi_{i_1}(\alpha)\right)\right)\left(1-2F_{i_2}\left(\xi_{i_2}(\alpha)\right)\right)\ldots\left(1-2F_{i_c}\left(\xi_{i_c}(\alpha)\right)\right)I\left(\sum_{i=1}^{n}\xi_i(\alpha)>zn\right)\right]\end{aligned}$$

Thus, since the random variables $\xi_1(\alpha),\xi(\alpha),\ldots,\xi_n(\alpha)$ are i.i.d.

$$\begin{aligned}P\left(\sum_{i=1}^{n}X_i>zn\right)=&P\left(\sum_{i=1}^{n}\xi_i(\alpha)>zn\right)\\&+\left(\sum_{c=2}^{n}\sum_{1\le i_1<i_2<\ldots<i_c\le n}\gamma_{i_1,i_2,\ldots,i_c}\right)\\&\times E\left[\left(1-2F_1\left(\xi_1(\alpha)\right)\right)\left(1-2F_2\left(\xi_2(\alpha)\right)\right)\ldots\left(1-2F_c\left(\xi_c(\alpha)\right)\right)I\left(\sum_{i=1}^{n}\xi_i(\alpha)>zn\right)\right]\end{aligned}$$

Now consider the last term

$$\begin{aligned}E\left[\left(1-2F_1\left(\xi_1(\alpha)\right)\right)\left(1-2F_2\left(\xi_2(\alpha)\right)\right)\ldots\left(1-2F_c\left(\xi_c(\alpha)\right)\right)I\left(\sum_{i=1}^n \xi_i(\alpha)>zn\right)\right]\\ =\sum_{s=0}^c \sum_{1\le j_1<j_2<\ldots<j_s\le c}(-1)^{c-s}\int_{\sum_{i=1}^n x_i>zn}\prod_{k\in\{j_1,j_2,\ldots,j_s\}}(2\alpha)x_k^{-2\alpha-1}\\ \times\sum_{k\in\{1,2,\ldots,n\}\backslash\{j_1,j_2,\ldots,j_s\}}\alpha x_k^{-\alpha-1}dx_1dx_2\ldots dx_n\\ =\sum_{s=0}^c \sum_{1\le j_1<j_2<\ldots<j_s\le c}(-1)^{c-s}P\left(\sum_{k\in\{j_1,j_2,\ldots,j_s\}}\xi_k(2\alpha)+\sum_{k\in\{1,2,\ldots,n\}\backslash\{j_1,j_2,\ldots,j_s\}}\xi_k(\alpha)>z\right)\end{aligned} \tag{26}$$

where $1\le j_1<j_2<\ldots<j_s\le c, s=0,1,\ldots,c, c=2,\ldots,n, (s,c)\ne(n,n)$ (and , thus , $(j_1,j_2,\ldots,j_c)$ is different from $(1,2,\ldots,n)$) Consequently, for large z, we obtain

$$P\left(\sum_{k\in\{j_1,j_2,\ldots,j_s\}}\xi_k(2\alpha)+\sum_{k\in\{1,2,\ldots,n\}\backslash\{j_1,j_2,\ldots,j_s\}}\xi_k(\alpha)>z\right)\sim P\left(\sum_{k\in\{1,2,\ldots,n\}\backslash\{j_1,j_2,\ldots,j_s\}}\xi_k(\alpha)>zn\right)\cdot(22)$$

In addition, by Corollary 1.3 .2 of Embrechts et al. (1997), we have, for large $z>0$

$$P\left(\sum_{k\in\{1,2,\ldots,n\}\backslash\{j_1,j_2,\ldots,j_s\}}\xi_k(\alpha)>z\right)\sim(n-s)P\left(\xi_1(\alpha)>zn\right)\sim\frac{n-s}{z^\alpha n^\alpha}$$

So, for for $s=c=n, (j_1,j_2,\ldots,j_n)=(1,2,\ldots,n)$

$$P\left(\sum_{k=1}^n \xi_k(2\alpha)>zn\right)\sim nP\left(\xi_1(2\alpha)>zn\right)\sim\frac{n}{z^{2\alpha}n^{2\alpha}} \tag{27}$$

From (26)-(27) it follows that, with $1\le j_1<j_2<\ldots<j_s\le c, s=0,1,\ldots,c, c=2,\ldots,n$

$$(s,c)\ne(n,n)$$

$$\begin{aligned}E\left[\left(1-2F_1\left(\xi_1(\alpha)\right)\right)\left(1-2F_2\left(\xi_2(\alpha)\right)\right)\ldots\left(1-2F_c\left(\xi_c(\alpha)\right)\right)I\left(\sum_{i=1}^n \xi_i(\alpha)>zn\right)\right]\sim\\ \sum_{s=0}^c \sum_{1\le j_1<j_2<\ldots<j_s\le c}(-1)^{c-s}\frac{n-s}{z^\alpha n^\alpha}=\left(\sum_{s=0}^c(-1)^{c-s}C_c^s\right)z^{-\alpha}n^{1-\alpha}-\left(\sum_{s=0}^c(-1)^{c-s}sC_c^s\right)z^{-\alpha}n^{-\alpha}\end{aligned}$$

where $C_c^s=c!/(s!(c-s)!)$ denotes binomial coefficients. Now, by the well-known identity for binomial coefficients,

$$\sum_{s=0}^c(-1)^{c-s}C_c^s=\sum_{s=0}^c(-1)^sC_c^s=0$$

It thus follows that $P\left(\sum_{i=1}^n X_i>zn\right)\sim P\left(\sum_{i=1}^n \xi_i(\alpha)>zn\right)$. Thus we have proved Theorem 5.1. ■

Proof of Theorem 5.2: The density is a polynomial of a lower order, which we write in the following generic form:

$$c\left(u_1,\ldots,u_n\right)=\sum_{k_1,\ldots,k_n=0,1,\ldots}\phi_{k_1,k_2,\ldots,k_n}\cdot u_1^{k_1}\cdot u_2^{k_2}\cdot\ldots\cdot u_n^{k_n}$$

Then,

$$\begin{aligned}P\left(\sum_{i=1}^n X_i>zn\right)=&E\left[\sum_{k_i\in\{0,1,\ldots\}}\phi_{k_1,k_2,\ldots,k_n}\right.\\ &\left.\times F_1^{k_1}\left(\xi_1(\alpha)\right)F_2^{k_2}\left(\xi_2(\alpha)\right)\ldots F_n^{k_n}\left(\xi_n(\alpha)\right)I\left(\sum_{i=1}^n\xi_i(\alpha)>zn\right)\right]\\ =&P\left(\sum_{i=1}^n\xi_i(\alpha)>zn\right)+E\left[\sum_{k_i\in\{0,1,\ldots\}\backslash\{k_i=0\forall i\}}\phi_{k_1,k_2,\ldots,k_n}\right.\\ &\left.\times F_1^{k_1}\left(\xi_1(\alpha)\right)F_2^{k_2}\left(\xi_2(\alpha)\right)\ldots F_n^{k_n}\left(\xi_n(\alpha)\right)I\left(\sum_{i=1}^n\xi_i(\alpha)>zn\right)\right]\end{aligned}$$

Now consider the last term

$$\begin{aligned}\sum\left[\sum_{k_i\in\{0,1,\ldots\}\backslash\{k_i=0\forall i\}}\phi_{k_1,k_2,\ldots,k_n}F_1^{k_1}\left(\xi_1(\alpha)\right)F_2^{k_2}\left(\xi_2(\alpha)\right)\ldots F_n^{k_n}\left(\xi_n(\alpha)\right)I\left(\sum_{i=1}^n\xi_i(\alpha)>zn\right)\right]\\ =\sum_{\sum_{i=1}^n s_i>nz_{k_i\in\{0,1,\ldots\}}}\psi_{k_1,k_2,\ldots,k_n}s_1^{-\alpha(k_1+1)}s_2^{-\alpha(k_2+1)}\ldots s_n^{-\alpha(k_n+1)}ds_1\ldots ds_n\\ =\sum_{k_i\in\{0,1,\ldots\}}\psi_{k_1,k_2,\ldots,k_n}\int_{\sum_{i=1}^n s_i>nz}s_1^{-\alpha(k_1+1)}s_2^{-\alpha(k_2+1)}\ldots s_n^{-\alpha(k_n+1)}ds_1\ldots ds_n\\ \sum_{k_i\in\{0,1,\ldots\}}\psi_{k_1,k_2,\ldots,k_n}P\left(\frac{\xi_1(\alpha(k_1+1))+\ldots+\xi_n(\alpha(k_n+1))}{n}>z\right)\end{aligned}$$

where the new coefficients ψ 's are different from ϕ 's because we have expressed $\left(1-s_i^{-\alpha}\right)^{k_i}$ in terms of powers of s_i^α. Now, using the same arguments as for (23)-(24)

$$\begin{aligned}P\left(\frac{\xi_1(\alpha)+\ldots+\xi_n(\alpha)}{n}>z\right)\sim nP\left(\xi_1(\alpha)>nz\right)\sim n^{1-\alpha}z^{-\alpha}\\ P\left(\frac{\xi_1(\alpha(k_1+1))+\ldots+\xi_n(\alpha(k_n+1))}{n}>z\right)\sim P\left(\xi_1(\alpha)>nz\right)\sim n^{-\alpha}z^{-\alpha}\end{aligned}$$

for all $k_i \geq 0$. It thus follows that $P\left(\sum_{i=1}^{n} X_i > zn\right) \sim P\left(\sum_{i=1}^{n} \xi_i(\alpha) > zn\right)$

Thus we have proved Theorem 5.2. ■

III. Appendix C - When Not to Spread Curtailed Cyber-Risks?

We proposed conditions under which it is statistically incentive compatible for a (re)-insurer to spread catastrophic cyber-risks having heavy tails. In this section, we further study the implications of it, by analyzing under which conditions it will *not be optimal* to spread risks. To calculate bounds from (14), we need bounds on $E|X|^r, G(\omega, z)$, and for uniformly diversified portfolios, on $F_n(z)$.

We assume i.i.d. risks $X_1, X_2, \ldots, X_n$ in $S_\alpha(\sigma, \beta, 0)$ with $\alpha \in (r, 1), \beta \in [-1, 1]$ and $\sigma > 0$.[30] From [4], we have that, for $X \in S_\alpha(\sigma, \beta, 0), r < \alpha < 1$

$$E|X - \operatorname{med}(X)|^r \leq 2^{2+r/\alpha} \sigma^r \Gamma\left(1 - \frac{r}{\alpha}\right) \Gamma(r) \sin\left(\frac{\pi}{2} r\right) \tag{28}$$

where $\operatorname{med}(X)$ denotes the median of X and $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ is the Gamma function. Furthermore, according to [4], if $\alpha \in (0, 1)$, then, using the notation $Q_{\alpha,\beta,\sigma}(x)$ for $P(X > x)$

$$Q_{\alpha,\beta,\sigma}(x) = \frac{1}{\alpha\pi} \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Gamma(k\alpha + 1)}{k\Gamma(k+1)} \sin\left(\frac{k\pi\alpha(1+\beta)}{2}\right) \frac{\sigma^{k\alpha}}{x^{k\alpha}} \tag{29}$$

$x > 0$ We also use the fact that

$$P\left(w^{(1)} X_1 + w^{(2)} X_2 > z\right) = P\left(X_1 > \frac{z}{\left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha\right]^{1/\alpha}}\right)$$

and more generally (for arbitrary nonnegative vectors summing to one, w)

$$P\left(\sum_{i=1}^{n} w_i X_i > z\right) = Q_{\alpha,\beta,\sigma}\left(z / \|\dot{w}\|_\alpha\right)$$

where $\|w\|_\alpha = \left(\sum_{i=1}^{n} (w_i)^\alpha\right)^{1/\alpha}$. Specifically, $1/\left\|\tilde{w}_n'\right\|_\alpha = n^{1-1/\alpha}$. Therefore, we have: $G(w,z) = Q_{\alpha,\beta,\sigma}\left(\frac{z}{\left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha\right]^{1/\alpha}}\right) - Q_{\alpha,\beta,\sigma}(z)$ and

$$F_n(z) = Q_{\alpha,\beta,\sigma}\left(z n^{1-1/\alpha}\right) - Q_{\alpha,\beta,\sigma}(z), \tag{30}$$

where $Q_{\alpha,\beta,\sigma}$ is defined in (25).

If we wish to introduce a time dimension, we can define the T-scaling operator: $\Lambda_T : x \mapsto Tx$. The well-known the $T^{1/2}$ rule for Brownian processes, W, implies that $W \circ \Lambda_T \stackrel{\mathrm{d}}{=} T^{1/2} \times W$. For processes in $S_\alpha(\sigma, 0, 0)$ this generalizes to the $T^{1/\alpha}$ " rule [21]), i.e., for $X : \mathbf{R}_+ \to \mathbf{R}$, a stable stochastic process with $X(1) \sim S_\alpha(\sigma, 0, 0)$, we have $X \circ \Lambda_T \triangleq T^{1/\alpha} \times X$. Thus, for such processes properties scale-up faster over time than for Brownian processes. With this $T^{1/\alpha}$ scaling in mind, for $X_1, \ldots, X_n$ stable processes $X_i : \mathbf{R}_+ \to \mathbf{R}$ and $X_i(t) \in S_\alpha\left(t^{1/\alpha}\sigma, 0, 0\right)$, we can define the truncated processes $X_i^a(T) = X_i(T)$, if $|X_i(T)| \leq aT^{1/\alpha}, X_i^a(T) = aT^{1/\alpha}$ if $X_i > aT^{1/\alpha}$ and $X_i^a(T) = -aT^{1/\alpha}$ if $X_i < -aT^{1/\alpha}$. With these definitions, it is clear that σ changes to $(T_2/T_1)^{1/\alpha} \sigma$ in equations $(24) - (26)$ when going from time-scale T_1, to time-scale T_2

We first study the symmetric case, i.e., the case when $\beta = 0$. For simplicity, we begin with the case when there are two assets, $n = 2$, and study how a depends on $w^{(1)}$ (and $w^{(2)} = 1 - w^{(1)}$). In this case, the analogue of equation (24) is (from [4]

$$E|X|^r \leq 2\sigma^r \Gamma\left(1 - \frac{r}{\alpha}\right) \Gamma(r) \sin\left(\frac{\pi}{2} r\right) \tag{31}$$

Furthermore, the asymptotic expansion (25) implies the following bounds for the tail of $Q_{\alpha,0,\sigma}$

$$\frac{1}{\alpha\pi} \Gamma(\alpha+1) \sin\left(\frac{\pi\alpha}{2}\right) \frac{\sigma^\alpha}{x^\alpha} - \frac{1}{\alpha\pi} \frac{\Gamma(2\alpha+1)}{4} \sin(\pi\alpha) \frac{\sigma^{2\alpha}}{x^{2\alpha}} < Q_{\alpha,0,\sigma}(x) < \frac{1}{\alpha\pi} \Gamma(\alpha+1) \sin\left(\frac{\pi\alpha}{2}\right) \frac{\sigma^\alpha}{x^\alpha} \tag{32}$$

Using (28) for $G(w,z)$, we get

$$\begin{aligned} G(w,z) > {} & \tfrac{1}{\alpha\pi} \Gamma(\alpha+1) \sin\left(\tfrac{\pi\alpha}{2}\right) \tfrac{\sigma^\alpha}{z^\alpha} \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha - 1\right) - \\ & \tfrac{1}{\alpha\pi} \tfrac{\Gamma(2\alpha+1)}{4} \sin(\pi\alpha) \tfrac{\sigma^{2\alpha}\left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha\right]^2}{z^{2\alpha}} \end{aligned} \tag{33}$$

Using bounds (4),(27) and (29) we get that the theorem holds with the following easy to compute estimate for the length of the distribution support:

$$\tilde{a} = \frac{z^{\alpha/r} (\alpha\pi)^{1/r} \sigma^{(r-\alpha)/r} \left(\Gamma\left(1 - \frac{r}{\alpha}\right) \Gamma(r) \sin\left(\frac{\pi}{2} r\right)\right)^{1/r} (n-1)^{1/r}}{\left[\Gamma(\alpha+1) \sin\left(\frac{\pi\alpha}{2}\right) \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)})\alpha\right) - 1\right) - \frac{\Gamma(2\alpha+1)}{4} \sin(\pi\alpha) \frac{\sigma^\alpha \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)})\right)^2\right.}{z^2}\right]^{1/r}} \tag{34}$$

Thus, $\tilde{a}$ as a function of $w^{(1)}$ provides a sufficient condition for cyber-risk spreading into (w_1, w_2) not being preferred to not spreading cyber-risk among other insurers.

Finally, we generalize to the case $\beta \neq 0$. Equation (27) and the right-hand-side inequality in (28) implies the following bound for the median $\operatorname{med}(X)$ of a r.v. $X \sim S_\alpha(\sigma, \beta, 0)$

$$|\operatorname{med}(X)| \leq 2^{1/\alpha} \sigma \left(\frac{1}{\alpha \pi} \Gamma(\alpha+1) \sin\left(\frac{\pi \alpha (1+\beta)}{2} \right) \right)^{1/\alpha}$$

This and (24) imply that

$$E|X|^r \leq 2^{r/\alpha} \sigma^r \left(\frac{1}{\alpha \pi} \Gamma(\alpha+1) \sin\left(\frac{\pi \alpha (1+\beta)}{2} \right) \right)^{r/\alpha} + 2^{2+r/\alpha} \sigma^r \Gamma\left(1 - \frac{r}{\alpha}\right) \Gamma(r) \sin\left(\frac{\pi}{2} r \right) \tag{35}$$

Similar to (29), we obtain that, in the general case of skewed stable distributions,

$$\begin{aligned} G(w,z) > \tfrac{1}{\alpha\pi} \Gamma(\alpha+1) \sin\left(\tfrac{\pi\alpha(1+\beta)}{2} \right) \tfrac{\sigma^\alpha}{z^\alpha} \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha - 1 \right) - \\ \tfrac{1}{\alpha\pi} \tfrac{\Gamma(2\alpha+1)}{4} \sin(\pi\alpha(1+\beta)) \tfrac{\sigma^{2\alpha} \left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha \right]^2}{z^{2\alpha}} \end{aligned} \tag{36}$$

Using bounds (31) and (32), we obtain that in the case of general skewed stable risks $X_i \sim S_\alpha(\sigma, \beta, 0)$, the theorem holds with the following easy to compute estimate for the length of the distribution support:

$$\tilde{a} > \frac{2^{1/r_z \alpha / r (\alpha\pi)^{1/r} \sigma^{(r-\alpha)/r}} \left(\left(\tfrac{1}{\alpha\pi} \Gamma(\alpha+1) \sin\left(\tfrac{\pi\alpha(1+\beta)}{2} \right) \right)^{r/\alpha} + 4\Gamma\left(1 - \tfrac{r}{\alpha}\right) \Gamma(r) \sin\left(\tfrac{\pi(1+\beta)}{2} r \right) \right)^{1/r} (n-1)^{1/r}}{\left[\Gamma(\alpha+1) \sin\left(\tfrac{\pi\alpha(1+\beta)}{2} \right) \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha - 1 \right) - \tfrac{\Gamma(2\alpha+1)}{4} \sin(\pi\alpha(1+\beta)) \tfrac{\sigma^\alpha \left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha \right)^2}{z^\alpha} \right]^{1/r}} \tag{37}$$

The same type of analysis as for the case with $\beta = 0$ could now be carried out for general β 's.

REFERENCES

[1] A. W. Marshall, I. Olkin, *et al.*, "Majorization in multivariate distributions," *The Annals of Statistics*, vol. 2, no. 6, pp. 1189–1200, 1974.
[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.
[3] P. Embrechts, A. McNeil, and D. Straumann, "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, vol. 1, pp. 176–223, 2002.
[4] V. M. Zolotarev, *One-dimensional stable distributions*, vol. 65. American Mathematical Soc., 1986.
[5] J.-P. Bouchaud and M. Potters, *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge university press, 2003.
[6] V. V. Uchaikin and V. M. Zolotarev, *Chance and stability: stable distributions and their applications*. Walter de Gruyter, 2011.
[7] S. M. Ross, *Introduction to probability models*. Academic press, 2014.
[8] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[9] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, vol. 143. Springer, 1979.
[10] H. Joe, *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.
[11] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.
[12] S. Cambanis, "Some properties and generalizations of multivariate eyraud-gumbel-morgenstern distributions," *Journal of Multivariate Analysis*, vol. 7, no. 4, pp. 551–559, 1977.
[13] R. Ibragimov, "Copula-based characterizations for higher order markov processes," *Econometric Theory*, vol. 25, no. 3, pp. 819–846, 2009.
[14] D. D. Mari and S. Kotz, *Correlation and dependence*. World Scientific, 2001.
[15] R. Nelsen, J. Quesada-Molina, and a. J. Rodriguez-Lallena, "Bivariate copulas with cubic sections," *Journal of Nonparametric Statistics*, vol. 7, no. 3, pp. 205–220, 1997.
[16] G. Molenberghs and E. Lesaffre, "Marginal modeling of correlated ordinal data using a multivariate plackett distribution," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 633–644, 1994.
[17] C. M. Cuadras, "Constructing copula functions with weighted geometric means," *Journal of Statistical Planning and Inference*, vol. 139, no. 11, pp. 3766–3772, 2009.
[18] C. M. Cuadras and W. Diaz, "Another generalization of the bivariate fgm distribution with two-dimensional extensions," *Acta et Commentationes Universitatis Tartuensis de Mathematica*, vol. 16, no. 1, pp. 3–12, 2012.
[19] G. Kimeldori and A. Sampson, "Uniform representations of bivariate distributions," *Communications in Statistics–Theory and Methods*, vol. 4, no. 7, pp. 617–627, 1975.
[20] G. D. Lin, "Relationships between two extensions of farlie-gumbel-morgenstern distribution," *Annals of the Institute of Statistical Mathematics*, vol. 39, no. 1, pp. 129–140, 1987.
[21] B. B. Mandelbrot, A. J. Fisher, and L. E. Calvet, "A multifractal model of asset returns," 1997.