Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

SUSTAINABLE CATASTROPHIC CYBER-RISK MANAGEMENT IN IOT SOCIETIES

Ranjan Pal*
Ziyuan Huang*
Xinlong Yin
Mingyan Liu

Electrical and Computer Engineering
University of Michigan
1301 Beal Ave
Ann Arbor, MI 48109, USA

Sergey Lototsky

Department of Mathematics
University of Southern California
3620 S. Vermont Ave
Los Angeles, CA 90089-2532, USA

Jon Crowcroft

Computer Laboratory
University of Cambridge
15 JJ Thompson Ave
Cambridge, CB3 0FD, UK

ABSTRACT

IoT-driven smart cities are popular service-networked ecosystems, whose proper functioning is hugely based on digitally secure and reliable supply chain relationships. However, the naivety in the current security efforts by concerned parties to protect IoT devices, pose tough challenges to scalable and expanding cyber-risk management markets for IoT societies, post a systemic cyber-catastrophe. As firms increasingly turn to cyber-insurance for reliable risk management, and insurers turn to reinsurance for their own risk management, questions arise as to how modern-day cyber risks aggregate and accumulate, and whether reinsurance is a feasible model for reliable catastrophic risk management and transfer in smart cities. In this introductory effort, we analyze (a) whether traditional cyber-risk spreading is a sustainable risk management practice and (b) under what conditions, *for the quite conservative scenario when proportions of i.i.d. catastrophic cyber-risks of a significant heavy-tailed nature are aggregated by a cyber-risk manager.*

1 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 (Coburn et al. 2018)) 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 (Gilchrist 2017) 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 (Coburn et al. 2018) for more examples). Due to 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 shortcited in detail by the authors in Biener et.al. (Biener et al. 2015), and include (i) indemnification of loss events, (ii) helping corporations estimate cost of cyber-risk, and (iii) improve cyber-security in the IoT age. 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 (Anderson and Moore 2009), (c) higher board level concerns in organizations post notable cyber-breach incidents (e.g., Sony, Target, WannaCry) and their negative effect on stock prices (Shetty et al. 2018; Gatzlaff and McCullough 2010), and (d) the growing perception of cyber-risk in the digital society (Pooser et al. 2018).

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) (Coburn et al. 2018), whereas the cumulative global public and private sector spendings on cyber-security amount only to USD 174 billion (Wang 2019). In addition, the total yearly market for cyber-insurance services - the most popular form of commercial third party commercial cyber-risk management offerings, approximates to a paltry USD 6 billion globally (Wang 2019), 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) (Wang 2019), (b) unavailability of quality plus quantity data on cyber-risks and demand side cyber-hygiene behavior, that contribute to policy pricing nuances (Romanosky et al. 2019; Franke 2017; Wang 2019), and (c) the empirical evidence of certain cyber-risk distributions being heavy-tailed and tail-dependent (Biener et al. 2015; Xu et al. 2018; Maillart and Sornette 2010), 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.

The idea of spreading aggregate cyber-risk among multiple risk managers (e.g., cyber-insurers) is gaining traction (Coburn et al. 2018; Kessler 2014) 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 (Kessler 2014)) may not apply to the former. Consequently, without a formal analysis, aggregate cyber-risk managers may not have the confidence to scale their service markets. *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.*

Research Contributions - As the primary contributions in this paper, we prove (and validate through simulations) 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, whereas spreading i.i.d. heavy-tailed cyber-risks that are *not catastrophic* is. 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. We also show 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.

2 PRELIMINARIES

In this section, we provide the necessary mathematical background for (a) families of distribution convolutions, and (b) basic *majorization theory* (Marshall et al. 1974), both of which we use for analysis. In the interest of space, we assume the reader of this paper to have a basic knowledge of the widely popular Value-at-Risk (VaR) measure (Embrechts et al. 2002) and its mathematical properties, and also of statistically stable distributions (Zolotarev 1986), both of which form the building blocks to (a) and (b) above. The reader is referred to the APPENDIX (Pal et al. 2020) for their basic background.

2.1 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. In this paper, we will limit ourselves to a specific but popular family of heavy tailed distributions whose tails decline parametrically as a polynomial function of some $\alpha > 0$. Within this context, a random variable X is said to have a *heavy-tailed* distribution if $0 < c \leq \frac{P(|X|>x)}{x^{-\alpha}} \leq C < \infty$ 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 $\mathrm{E}[|X|^p]$ for $p < \alpha$, and infinite statistical moments for $p \geq \alpha$.

Families Related to Convolution of Symmetric Stable Distributions - For $0 \leq r < 2$, we denote by $\overline{\mathscr{CS}}(r)$ 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, $\overline{\mathscr{CS}}(r)$ consists of cyber-risk distributions of r.v.'s X for which, with some $k \geq 1, X = Y_1 + ... + Y_k$, 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 $0 \leq r \leq 2$, we denote by $\underline{\mathscr{CS}}(r)$ 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, $\underline{\mathscr{CS}}(r)$ consists of cyber-risk distributions of r.v.'s X for which, with some $k \geq 1, X = Y_1 + ... + Y_k$, 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 $\overline{\mathscr{CS}}(r)$ and $\underline{\mathscr{CS}}(r)$ 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 $\overline{\mathscr{CS}}(r)$ 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 $\underline{\mathscr{CS}}(r)$ 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 $0 < r_1 < r_2 \leq 2$, the following inclusions hold: $\underline{\mathscr{CS}}(r_2) \subset \underline{\mathscr{CS}}(r_1)$ and $\underline{\mathscr{CS}}(r_1) \subset \overline{\mathscr{CS}}(r_2)$. Cauchy distributions $S_1(\sigma,0,0)$ are at the dividing boundary between the classes $\underline{\mathscr{CS}}(1)$ and $\overline{\mathscr{CS}}(1)$ 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 $\underline{\mathscr{CS}}(r)$ and $\underline{\mathscr{CS}}(r)$. More precisely, the Cauchy distributions $S_1(\sigma,0,0)$ are the only ones that belong to all the classes $\underline{\mathscr{CS}}(r)$ with $r > 1$ and all the classes $\overline{\mathscr{CS}}(r)$ with $r < 1$. Stable distributions $S_r(\sigma,0,0)$ are the only ones that belong to all the classes $\underline{\mathscr{CS}}(r')$ with $r' > r$ and all the classes $\overline{\mathscr{CS}}(r')$ with $r' < r$. The properties of stable distributions discussed herein imply that the p-th absolute moments $\mathbb{E}[|X|^p]$ of a r.v. $X \sim \overline{\mathscr{CS}}(r)$, $r \in (0,2)$, are finite if $p \leq r$. However, all the r.v.'s $X \sim \underline{\mathscr{CS}}(r), r \in (0,2]$ have infinite moments of order r: $\mathbb{E}[|X|^r] = \infty$. In particular, the distributions of r.v.'s X from the class $\underline{\mathscr{CS}}(1)$ are extremely heavy-tailed (representing catastrophic cyber-risks) in the sense that their first moments are infinite: $\mathbb{E}[|X|] = \infty$.

2.2 Basics of Majorization Theory

A vector[1] 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 $\sum_{i=1}^k w_{[i]} \leq \sum_{i=1}^k v_{[i]}$, $k = 1, \cdots, n-1$, and $\sum_{i=1}^n a_{[i]} = \sum_{i=1}^n b_{[i]}$, where $w_{[1]} \geq \cdots \geq w_{[n]}$ and $v_{[1]} \geq \cdots \geq v_{[n]}$ 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 (Marshall et al. 1974)). For instance, it is easy to see that the following holds:

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

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

$$\left(\frac{1}{n+1}, \cdots, \frac{1}{n+1}, \frac{1}{n+1}\right) \prec \left(\frac{1}{n}, \cdots, \frac{1}{n}, 0\right) \ . \tag{2}$$

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) (Boyd and Vandenberghe 2004) if $(w \prec v) \implies (\phi(w) \leq \phi(v))$ (resp. $(w \succ v) \implies (\phi(w) \geq \phi(v))$, $\forall w, v \in \mathbf{R}_+^n$. 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:

$$\phi(w_1, \cdots, w_n) = \phi(w_{\pi(1)}, \cdots, w_{\pi}(n)), \tag{3}$$

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 (Marshall et al. 1979)).

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

$$Z_{\mathbf{w}} = \sum_{i=1}^n w_i X_i \ . \tag{4}$$

Denote by $\mathscr{I}_n = \{w = (w_1, ..., w_n) : w_i \geq 0, i = 1, \cdots, n, \sum_{i=1}^n w_i = 1\}$ the simplex of all vectors where weights sum to 1. Define two special vectors $\underline{w} = (\frac{1}{n}, \frac{1}{n}, ... \frac{1}{n}) \in \mathscr{I}_n$ and $\overline{w} = (1, 0, ..., 0) \in \mathscr{I}_n$. 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 $\mathscr{I}_n$, while the less diverse $\underline{w}$ has more varied weights (equally spread over n risks). Similarly, Eqn (2) suggests that $(\frac{1}{n+1}, \cdots, \frac{1}{n+1}) \in \mathscr{I}_{n+1}$ has more varied weights than $(\frac{1}{n}, \cdots, \frac{1}{n}, 0) \in \mathscr{I}_{n+1}$ 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 $n \geq 2$:

$$VaR_q(Z_{\underline{w}}) = \frac{1}{\sqrt{n}} VaR_q(X_1) = \frac{1}{\sqrt{n}} VaR_q(Z_{\overline{w}}) < VaR_q(Z_{\overline{w}}) \ . \tag{5}$$

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.

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

3 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. Specifically, in this paper we will often consider the average of n 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 $w = [w_1, \cdots, w_n]$: $Z_w = \sum_{i=1}^{n} w_i X_i$.

3.1 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* and *Cauchy* distributions (Forbes et al. 2011), as a representative of the latter that are *statistically stable* (Zolotarev 1986); 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 $\mathscr{L}(\mu, \sigma)$, 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 $X_1, \cdots, X_n \sim \mathscr{L}(\mu, \sigma)$, we have $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathscr{L}(\mu, n\sigma)$. 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. As another example, consider the Cauchy distribution denoted by $\mathscr{G}(\mu, \sigma)$, with location parameter μ and scale σ, and the pdf - cdf combination given by $\phi(x) = \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{(x-\mu)^2}{\sigma^2}\right)}$, $F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-\mu}{\sigma}\right)$.

Again, standard results suggest that for IID $X_1, \cdots, X_n \sim \mathscr{G}(\mu, \sigma)$, we have $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathscr{G}(\mu, \sigma)$, 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.

This suggests that risk aggregation in this case is undesirable - specifically, 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.

3.2 Aggregating IID Catastrophic and Non-catastrophic Cyber-risks

We first consider aggregating IID risks X_i from the family $\underline{\mathscr{CS}}(1)$, 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. We have the following result, the proof of which is in the APPENDIX (Pal et al. 2020).

Theorem 1 For IID r.v's $X_i \sim \underline{\mathscr{CS}}(1), i = 1, \cdots, n$, $q \in (0,1)$, and n-vector of weights $w, v \in \mathbf{R}^n_+$,

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, $VaR_q(Z_{\bar{w}}) < VaR_q(Z_w) < VaR_q(Z_{\underline{\mathrm{W}}})$, $\forall w \in \mathscr{I}_n$ such that $w \neq \underline{\mathrm{w}}$ 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 - *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 $w = (w_1, ..., w_n) \in \mathscr{I}_n$, $Z_w = \sum_{i=1}^n w_i X_i =_d X_1$. 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 $w \in \mathscr{I}_n$. 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 $X_i \sim S_1(\sigma, 0, 0)$.

We now consider aggregating IID risks X_i from the family $\overline{\mathscr{CSLC}}$, 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. We now have a result on VaR performance post cyber-risk aggregation, the proof of which is in APPENDIX (Pal et al. 2020).

Theorem 2 *For IID r.v's $X_i \sim \overline{\mathscr{CSLC}}, i = 1, \cdots, n$, $q \in (0,1)$, and n-vector of weights $w, v \in \mathbf{R}^n_+$,*

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, $VaR_q(Z_{\underline{\mathrm{W}}}) < VaR_q(Z_w) < VaR_q(Z_{\overline{w}})$, $\forall w \in \mathscr{I}_n$ 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 1 is a surprising new facet that we unravel in this paper via theory.

4 FEASIBILITY ANALYSIS AFTER CURTAILING CYBER-RISK TAILS

Most if not all cyber-(re)insurance firms are risk-averse in cyber-settings. Consequently, they would enforce coverage limits on incumbent heavy-tailed catastrophic cyber-risks. The conservative question we ask in this section is: *is it statistically feasible (incentive compatible) for an aggregate risk manager to cover aggregate curtailed (non) i.i.d. cyber-risks that have heavy tails*? We emphasize again that the term

'statistical feasibility' is synonymous with the commercial (un)viability of a cyber-risk manager to cover aggregate risks with increasing spreads. with respect to the nature of aggregate cyber-risk spread. Here, we define[2] the a-truncated version of a random variable X_i as:

$$Y_i(a) = \begin{cases} X_i, & \text{if } |X_i| \leq a, \\ -a, & \text{if } X < a, \\ a & \text{if } X > a\ . \end{cases} \tag{6}$$

Similarly, we note by $Y_w(a)$ the a-truncated version of the r.v. $X_w = \sum_{i=1}^{n} w_i X_i$ given a weight vector w. We note that a negative feasibility outcome of conservative case of i.i.d. cyber-risks provides a clear qualitative indication that the feasibility outcome might be negative for non-identical and dependent cyber-risks.

4.1 Tail behavior when aggregating individually curtailed risks

We now analyze what happens when aggregating multiple heavy-tailed risks each of which has been curtailed. 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 1 for curtailed catastrophic cyber-risks, in this regard.

Theorem 3 *Let $n \geq 2$ and let $w \in \mathscr{I}_n$ be a weight vector with $w_{[1]} \neq 1$. Let $X_i, i = 1, \cdots, n$ be IID r.v.'s $\sim \underline{\mathscr{CL}}(r)$ 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 1). For any $z > 0$, and all*

$$a > \left(\frac{\mathbb{E}[|X_1|^r](n-1)}{2G(w,z)} \right)^{\frac{1}{r}}, \tag{7}$$

the following inequality holds:

$$P(Y_w(a) > z) > P(Y_1(a) > z). \tag{8}$$

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 1 in the case of bounded cyber-risks. More specifically, cyber-risk aggregation coverage continues to be disadvantageous 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 1 and 3 indicate that VaR is not sub-additive and, thus, its coherency is always violated in the class of extremely heavy-tailed cyber-risks with infinite first moments. More specifically, Theorem 3 implies that VaR may also be non-coherent in the world of cyber-risks with bounded distributional support.

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

$$G(\tilde{w}_n, z) = P\left(\frac{X_1 + X_2}{2} > z \right) - P(X_1 > z). \tag{9}$$

This means that the length of the distributional support reflecting statistical incentive non-compatibility to aggregate cyber-risk coverage in Theorem 3 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 $w_{[1]} < 1 - \varepsilon$, (and, thus, $w_i < 1\varepsilon$ for all i), where $0 < \varepsilon < \frac{1}{2}$. As follows from the proof of Theorem 3, for all such portfolios w, the theorem

[2]This definition of truncation moves probability masses to the edges of the distribution.

holds for $a > \left(\frac{\mathbb{E}[|X_1|^r](n-1)}{2\tilde{G}(w,z)}\right)^{\frac{1}{r}}$, where $\tilde{G}(\varepsilon,z) = P((1-\varepsilon)X_1 + \varepsilon X_2 > z)) < G(w,z)$. This follows since any vector w with $w_{[1]} < 1-\varepsilon$ is majorized by the vector $(1-\varepsilon,\varepsilon,0,..,0)$.

Implication 3 - From the proof of Theorem 3, 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

$$F_n(z) = P\left(\frac{\sum_{i=1}^n X_i}{n} > z\right) - P(X_1 > z) \tag{10}$$

Note that, by Theorem 1, $F_n(z) > H(z) = G(\tilde{w}_n,z)$ 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 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 infeasibility or lack of incentive compatibility feature 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 $VaR_q[Y_w(a)]$ and $VaR_q[Y_1(a)]$, 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, the length of the distributional support a depends on the value of the disaster level z - denoting the degree of heavy-tailedness). The crucial qualitative difference of the results in Theorem 2 for bounded/curtailed cyber-risk distributions and their implications for the value at risk, from those given by Theorem 1 and Theorem 3 for unbounded risks, where the inequalities hold for all $z > 0$ and all $q \in (0,1)$.

Implication 5 (The special case of non-identical distributions) - Analogues of Theorem 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, this implies that complete analogues of the results in the present section for bounded versions of symmetric risks from the classes $\underline{\mathscr{CS}}(r)$ 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 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 $a > \left(\frac{E|X_1|^r(n-1)}{G(w,z)}\right)^{1/r}$.

4.2 When Not to Spread Curtailed Cyber-Risks?

In Theorem 3, 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 Theorem 3, 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 $\sigma > 0$. From (Zolotarev 1986), we have that, for $X \in S_\alpha(\sigma,\beta,0), r < \alpha < 1$

$$E|X - \text{med}(X)|^r \leq 2^{2+r/\alpha}\sigma^r\Gamma\left(1-\frac{r}{\alpha}\right)\Gamma(r)\sin\left(\frac{\pi}{2}r\right) \tag{11}$$

where 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 (Zolotarev 1986), if $\alpha \in (0,1)$, then, using the notation $Q_{\alpha,\beta,\sigma}(x)$ for $P(X > x)$

$$Q_{\alpha,\beta,\sigma}(x) = \frac{1}{\alpha\pi}\sum_{k=1}^{\infty}(-1)^{k-1}\frac{\Gamma(k\alpha+1)}{k\Gamma(k+1)}\sin\left(\frac{k\pi\alpha(1+\beta)}{2}\right)\frac{\sigma^{k\alpha}}{x^{k\alpha}} \tag{12}$$

$x > 0$ 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

$$F_n(z) = Q_{\alpha,\beta,\sigma}\left(zn^{1-1/\alpha}\right) - Q_{\alpha,\beta,\sigma}(z), \tag{13}$$

where $Q_{\alpha,\beta,\sigma}$ is defined in (12).

If we wish to introduce a time dimension, we can define the T-scaling operator: $\Lambda_T : x \mapsto Tx$. The well-known $^{\text{the}}$ $T^{1/2}$ 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 (Mandelbrot et al. 1997)), 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 $|X_i(T)| \le aT^{1/\alpha}$,$X_i^a(T) = aT^{1/\alpha}$ 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 $(13)-(15)$ 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 (11) is (from (Zolotarev 1986)

$$E|X|^r \le 2\sigma^r\Gamma\left(1-\frac{r}{\alpha}\right)\Gamma(r)\sin\left(\frac{\pi}{2}r\right) \tag{14}$$

Furthermore, the asymptotic expansion (12) implies the following bounds for the tail of $Q_{\alpha,0,\sigma}$

$$\frac{1}{\alpha\pi}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)\frac{\sigma^\alpha}{x^\alpha} - \frac{1}{\alpha\pi}\frac{\Gamma(2\alpha+1)}{4}\sin(\pi\alpha)\frac{\sigma^{2\alpha}}{x^{2\alpha}} < Q_{\alpha,0,\sigma}(x) < \frac{1}{\alpha\pi}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)\frac{\sigma^\alpha}{x^\alpha} \tag{15}$$

Using (15) for $G(w,z)$, we get

$$G(w,z) > \frac{1}{\alpha\pi}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)\frac{\sigma^\alpha}{z^\alpha}\left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha - 1\right) - \frac{1}{\alpha\pi}\frac{\Gamma(2\alpha+1)}{4}\sin(\pi\alpha)\frac{\sigma^{2\alpha}\left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha\right]^2}{z^{2\alpha}} \tag{16}$$

Using bounds (7),(14) and (16) we get that Theorem 3 holds with the following easy to compute estimate for the length of the distribution support:

$$\tilde{a} = \frac{z^{\alpha/r(\alpha\pi)^{1/r}\sigma^{(r-\alpha)/r}\left(\Gamma\left(1-\frac{r}{\alpha}\right)\Gamma(r)\sin\left(\frac{\pi}{2}r\right)\right)^{1/r}}(n-1)^{1/r}}{\left[\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)\left(\left(w^{(1)}\right)^\alpha + \left(w^{(2))\alpha} - 1\right) - \frac{\Gamma(2\alpha+1)}{4}\sin(\pi\alpha)\frac{\sigma^\alpha\left(\left(w^{(1)}\right)^\alpha + \left(w^{(2))}\right)^2\right.}{z^2}\right]^{1/r}} \tag{17}$$

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 (14) and the right-hand-side inequality in (15) implies the following bound for the median $\text{med}(X)$ of a r.v. $X \sim S_\alpha(\sigma, \beta, 0)$

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

This and (11) imply that

$$E|X|^r \leq 2^{r/\alpha}\sigma^r\left(\frac{1}{\alpha\pi}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha(1+\beta)}{2}\right)\right)^{r/\alpha} + 2^{2+r/\alpha}\sigma^r\Gamma\left(1-\frac{r}{\alpha}\right)\Gamma(r)\sin\left(\frac{\pi}{2}r\right) \tag{18}$$

Similar to (16), we obtain that, in the general case of skewed stable distributions,

$$\begin{aligned} G(w,z) > \tfrac{1}{\alpha\pi}\Gamma(\alpha+1)\sin\left(\tfrac{\pi\alpha(1+\beta)}{2}\right)\tfrac{\sigma^\alpha}{z^\alpha}\left(\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha - 1\right) - \\ \tfrac{1}{\alpha\pi}\tfrac{\Gamma(2\alpha+1)}{4}\sin(\pi\alpha(1+\beta))\tfrac{\sigma^{2\alpha}\left[\left(w^{(1)}\right)^\alpha + \left(w^{(2)}\right)^\alpha\right]^2}{z^{2\alpha}} \end{aligned} \tag{19}$$

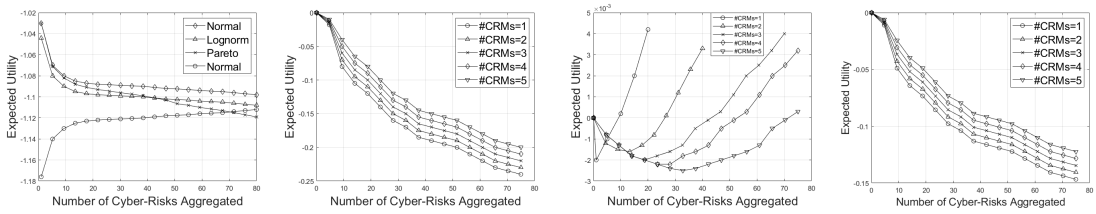
Using bounds (18) and (19), we obtain that in the case of general skewed stable risks $X_i \sim S_\alpha(\sigma, \beta, 0)$, Theorem 3 holds with the following easy to compute estimate for the length of the distribution support:

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

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

5 TRACE-DRIVEN SIMULATION

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 (Biener et al. 2015) and (Eling and Wirfs 2019). 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, and not necessarily stable, 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 (Nešlehová et al. 2006). *We thus can confirm that cyber risks are indeed very heavy tailed and the expectation and variance do not exist.* 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: $V(x) = X \,\text{if}\, X < k$, and $k \,\text{if}\, X \geq k$, 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 Morgentern (VNM) utility function. β is degree of risk-aversion of the cyber-insurer.



(a) 1a (b) 1b (c) 1c (d) 1d

Figure 1: Simulation Output.

Figure 1a shows the EU-theoretic performance based on a power utility function $u(x)$ for aggregating i.i.d. As expected (from theory), for normally distributed i.i.d. cyber-risks, 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. 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 1b. *Since the expected utility decreases monotonically, not providing any (pooled with multiple CRMs) coverage management would be optimal and the aggregate coverage market would fail completely.* Figure 1c(d). shows that for cyber-risk with a Pareto Index of 1 (0.81) 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 (decreasing). The strange U-shape denotes that the benefit from aggregation first decreases before it eventually increases again, only for borderline-heavy tailed cyber-risks.

6 SUMMARY AND FUTURE WORK

In this paper, we provided a rigorous general theory to elicit conditions on i.i.d. 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. We conducted a real-data driven numerical simulation study to validate claims made in theory, where we relaxed the assumption regarding the stable structure of cyber-risk distributions, as is usual in practice. As part of future work, we intend to extend our analysis to non-i.i.d., i.e., tail-dependent heavy-tailed cyber-risks.

ACKNOWLEDGMENTS

R. Pal and M. Liu have been supported by the NSF under grants CNS-1616575, CNS-1939006, and ARO W911NF1810208.

REFERENCES

Anderson, R., and T. Moore. 2009. “Information Security: Where Computer Science, Economics and Psychology Meet”. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367(1898):2717–2727.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. “Coherent Measures of Risk”. *Mathematical Finance* 9(3):203–228.

Biener, C., M. Eling, and J. H. Wirfs. 2015. “Insurability of Cyber Risk: An Empirical Analysis”. *The Geneva Papers on Risk and Insurance-Issues and Practice* 40(1):131–158.

Bouchaud, J.-P., and M. Potters. 2003. *Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management*. Cambridge: Cambridge university press.

Boyd, S., and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge: Cambridge university press.

Coburn, A., E. Leverett, and G. Woo. 2018. *Solving Cyber Risk: Protecting Your Company and Society*. Hoboken, New Jersey: Wiley.

Eling, M., and J. Wirfs. 2019. "What are the Actual Costs of Cyber Risk Events?". *European Journal of Operational Research* 272(3):1109–1119.

Embrechts, P., A. McNeil, and D. Straumann. 2002. "Correlation and Dependence in Risk Management: Properties and Pitfalls". *Risk Management: Value at Risk and Beyond* 1:176–223.

Forbes, C., M. Evans, N. Hastings, and B. Peacock. 2011. *Statistical Distributions*. New York: John Wiley & Sons.

Franke, U. 2017. "The Cyber Insurance Market in Sweden". *Computers & Security* 68:130–144.

Gatzlaff, K. M., and K. A. McCullough. 2010. "The Effect of Data Breaches on Shareholder Wealth". *Risk Management and Insurance Review* 13(1):61–83.

Gilchrist, A. 2017. *IoT Security Issues*. Boston, Massachusetts: Walter de Gruyter GmbH & Co KG.

Kessler, D. 2014. "Why (Re) Insurance is not Systemic". *Journal of Risk and Insurance* 81(3):477–488.

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

Mandelbrot, B., A. Fisher, and L. Calvet. 1997. "A Multifractal Model of Asset Returns". Cowles Foundation Discussion Papers 1164, Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut.

Marshall, A. W., I. Olkin et al. 1974. "Majorization in Multivariate Distributions". *The Annals of Statistics* 2(6):1189–1200.

Marshall, A. W., I. Olkin, and B. C. Arnold. 1979. *Inequalities: Theory of Majorization and its Applications*, Volume 143. New York: Springer.

Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin. 2006. "Infinite-mean Models and the LDA for Operational Risk". *The Journal of Operational Risk* 1(1):3–25.

Pal, R., Z. Huang, X. Yin, M. Liu, S. Lototsky, and J. Crowcroft. 2020. "Sustainable Catastrophic Cyber-risk Management in IOT Societies Appendix". https://drive.google.com/open?id=1GZCOLqQAeBHYK23jqx70-Fu9Xk8DaHjC, accessed 15th June.

Pooser, D. M., M. J. Browne, and O. Arkhangelska. 2018. "Growth in the Perception of Cyber Risk: Evidence from US P&C Insurers". *The Geneva Papers on Risk and Insurance-Issues and Practice* 43(2):208–223.

Romanosky, S., L. Ablon, A. Kuehn, and T. Jones. 2019. "Content Analysis of Cyber Insurance Policies: How do Carriers Price Cyber Risk?". *Journal of Cybersecurity* 5(1). tyz002.

Ross, S. M. 2014. *Introduction to Probability Models*. New York: Academic press.

Shetty, S., M. McShane, L. Zhang, J. P. Kesan, C. A. Kamhoua, K. Kwiat, and L. L. Njilla. 2018. "Reducing Informational Disadvantages to Improve Cyber Risk Management". *The Geneva Papers on Risk and Insurance-Issues and Practice* 43(2):224–238.

Uchaikin, V. V., and V. M. Zolotarev. 2011. *Chance and Stability: Stable Distributions and their Applications*. Berlin: Walter de Gruyter.

Wang, S. S. 2019. "Integrated Framework for Information Security Investment and Cyber Insurance". *Pacific-Basin Finance Journal* 57:101173.

Xu, M., K. M. Schweitzer, R. M. Bateman, and S. Xu. 2018. "Modeling and Predicting Cyber Hacking Breaches". *Institute of Electrical and Electronics Engineers Transactions on Information Forensics and Security* 13(11):2856–2871.

Zolotarev, V. M. 1986. *One-dimensional Stable Distributions*, Volume 65. Providence: American Mathematical Soc.

AUTHOR BIOGRAPHIES

RANJAN PAL is a faculty member of ECE at University of Michigan. His primary research interest is in engineering robust cyber-security and information privacy solutions using decision and the applied mathematical sciences. He serves as an Associate Editor of IEEE Networking Letters, and ACM Transactions on MIS. His email address is palr@umich.edu.

ZIYUAN HUANG is an undergraduate student in ECE at University of Michigan. His research interests include financial mathematics and cybersecurity. He is a student member of the IEEE and the ACM. His email address is ziyuanh@umich.edu.

XINLONG YIN is an undergraduate student in ECE at the University of Michigan. His research interests include cybersecurity-related machine learning and statistics. He is a student member of the IEEE and the ACM. His email address is connory@umich.edu.

MINGYAN LIU is an entrepreneur and the Peter and Evelyn and Fuss Chair Professor of Electrical and Computer Engineering at University of Michigan. Her research interests include incentive design for cybersecurity, privacy and experimental data science related to cyber security. She is a Fellow of the IEEE and a member of the ACM. Her email address is mingyan@umich.edu.

SERGEY LOTOSKY is a Professor of Mathematics at University of Southern California. His research interests include probability theory, stochastic processes, and stochastic partial differential equations. He is an Associate Editor of Stochastic and Partial Differential Equations, and the SIAM Journal on Mathematical Analysis. His email address is lototsky@usc.edu.

JON CROWCROFT is the Marconi Professor of Communications Systems in the Computer Laboratory at the University of Cambridge. His research interests include opportunistic communications and privacy preserving analytics. He is a Fellow of the Royal Society, the ACM, the British Computer Society, the Royal Academy of Engineering, and the IEEE. His email address is jon.crowcroft@cl.cam.ac.uk.