Essays on Risk Modeling and Aggregation with a Focus on Cyber Risk

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Finance

submitted by

Kwangmin Jung

from

South Korea

Approved on the application of

Prof. Dr. Martin Eling

and

Prof. Dr. Enrico De Giorgi

Dissertation no. 4953

Difo-Druck GmbH, Untersiemau 2020

The University of St.Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, October 22, 2019

The President:

Prof. Dr. Thomas Bieger

Table of Contents

List of Figures

List of Tables

Summary

This dissertation consists of four essays exploring risk modeling and aggregation, with a particular focus on cyber risk in the insurance context. It aims to provide appropriate answers to research questions that arise from practical challenges on both the supply and demand sides of the cyber-insurance market. By answering these questions, this dissertation constructs data-driven risk modeling and aggregation methods and investigates an economic framework to represent potential insureds' behavior observed in the current market. In particular, risk modeling and aggregation methods are further applied in the study on the improvement of the regulatory model for risk capital.

The first two essays, "Copula approaches for modeling cross-sectional dependence of data breach losses" and "Probable maximum cyber loss: Empirical estimation and reinsurance design with public intervention," investigate supply side challenges in the current cyber-insurance market by constructing risk modeling and aggregation methods and an alternative approach to estimating extreme losses. These essays attempt to provide a statistical process of data breach losses, their dependence structure in risk pooling, the evaluation of extreme loss events beyond a widely used Pareto model, and a possible collaboration between the public sector and the private sector in the risk transfer scheme.

A challenge on the demand side of the cyber-insurance market represented by a behavioral bias is examined in the third essay, "Decision-making on cyber risk management: Interaction between market insurance and risk control measures under prospect theory", which models an economic decision on risk transfer and control measures under loss aversion assumption in the presence of interdependent risk. The finding of this essay supports anecdotal evidence of the behavioral bias against cyber risks, showing that an agent tends to not implement additional measures for cyber risk management.

The fourth essay, "Risk aggregation in non-life insurance: Standard models vs. internal models", provides an approach (internal model) for more accurately estimating risk capital for insurers based on risk modeling and aggregation methods studied in the first essay. Using two datasets from Korea and Germany and assessing three regulatory schemes (Korean Risk-based Capital, Solvency II and Swiss Solvency Test), the essay concludes that the standard models lead to more than 50% higher capital requirements on average than those of internal models, which can imply a potential distortion of the competition when both approaches are available in a single market.

Zusammenfassung

Diese Dissertation besteht aus vier Arbeiten zum Thema Risikomodellierung und -aggregation mit Fokus auf Cyberrisiken in einem Versicherungskontext. Das Ziel ist es Forschungsfragen zu beantworten, die sich in Zusammenhang mit Herausforderungen bei der Nachfrage- und Angebotsseite im Cyberversicherungsmarkt stellen. Die Forschungsfragen werden bearbeitet indem datenbasierte Risiko- und Aggregationsmodelle konstruiert werden und das Nachfrageverhalten im bestehenden Markt durch ökonomische Modelle repräsentiert wird. Im speziellen werden Risikomodellierungs- und Aggregationsmethoden verwendet um Verbesserungsvorschläge für die regulatorischen Kapitalmodelle herzuleiten.

Die ersten beiden Arbeiten, „Copula approaches for modeling cross-sectional dependence of data breach losses“ und „Probable maximum cyber loss: Empirical estimation and reinsurance design with public intervention“, untersuchen die Herausforderungen, die sich im aktuellen Cyberversicherungsmarkt auf der Nachfrageseite ergeben, indem Risiko-, Aggregations- und alternative Extremwertmodelle konstruiert werden. Die Modelle beschreiben den statistischen Prozess von Datenverlusten, Abhängigkeiten im Risikopools, extreme Verluste, die über die häufig verwendete Paretoverteilung hinausgehen, und mögliche Kooperationen zwischen öffentlichem und privatem Sektor bei der Risikotragung.

Die Herausforderung auf der Nachfrageseite des Cyberversicherungsmarktes, beschrieben als Verhaltensbias, wird in der dritten Arbeit, „Decision-making on cyber risk management: Interaction between market insurance and risk control measures under prospect theory“, untersucht. Dabei werden ökonomische Entscheidungen betreffend Risikotransfer und Risikokennzahlen unter der Annahme von Verlustaversion und statistischer Unabhängigkeit modelliert. Die Ergebnisse dieser Analyse stützen tatsächlich die anekdotischen Beobachtungen eines Verhaltensbias für Cyberrisiken und entsprechen würden die Akteure keine zusätzlichen Risikokennzahlen im Rahmen des Cyber-Risikomanagements implementieren.

Die vierte Arbeit, „Risk aggregation in non-life insurance: Standard models vs. internal models“, erweitert die Risikomodellierungs- und Aggregationsmethoden aus der ersten Arbeit um ein adäquateres (internes) Kapitalmodell. Basierend auf zwei Datensätze aus Korea und Deutschland sowie drei verschiedenen regulatorischen Modellen (Koreanisches risikobasiertes Kapitalmodel, Solvency II und Swiss Solvency Test), schlussfolgert die Arbeit, dass das Standardmodell durchschnittlich über 50% mehr Kapital verlangt als das interne Modell. Dies könnte zu Wettbewerbsverzerrungen führen, falls beide Modelle gleichzeitig angewendet werden.

Acknowledgement

First of all, I would like to thank my advisor, Martin Eling, for tremendous support to improve the quality of this dissertation and guidance on the attitude towards research. His enthusiasm for academic life has inspired me and indeed made me perceive improvements in my way of research. I strongly believe that this inspiration would help me keep passion for my work and achievement in the future. Furthermore, I would like to thank my co-advisor, Enrico De Giorgi, for his great advice and consideration on my research.

In addition to my advisors, I am grateful to my colleagues at I.VW. for their great supports, discussions, motivations and encouragement. I believe that this dissertation would not have been done without their supports at the institute.

I also thank my parents and family in Korea for their heartful supports on my journey since I left Korea. They have been always encouraging me in my studies despite being about 8,500 kilometers away, the encouragement that has been making me be able to constantly motivate myself.

Lastly, special thanks go to my wife, Minji Jung, and daughter, Hanni Jung, whose supports have been the main ground of my life and endurance over the years. It is obvious that my achievement so far is absolutely credited to them by keeping stability in life, making me smile on a daily basis and inspiring me with confidence. It is hard for me to imagine what my time in Europe would have been without them.

St. Gallen, December 2019, Kwangmin Jung

Introduction

Risk landscapes currently facing non-life insurance businesses in relation to information technology are rapidly evolving. Classical, widely used methodologies in both academia and industry might not appropriately consider data circumstances and fast-changing risk types. This dissertation deals with statistical techniques that can be used to address new risk landscapes by providing data-driven risk modeling and aggregation methods to estimate the size of risk in the future. In particular, with a focus on cyber risk as a new risk type in the digital society, the discussion attempts to resolve current problems arising from both the supply and demand sides of the cyber-insurance market and suggest empirical benchmarks for risk measurement and insurance pricing.

Emerging risks relative to information technology represented by cyber risk in this dissertation demonstrate several challenges for both the supply and demand sides. The supply side usually lacks standard modeling in loss process and risk aggregation associated with new risks, which can result from relatively less accumulated databases and a lack of understanding related to the characteristics of risks. To fill this gap in the cyber risk context, an actuarial model for cyber risk assessment and dependence structure in a cyber-insurance risk pool is constructed with one of the largest databases for data breach losses in the first essay, "Copula approaches for modeling cross-sectional dependence of data breach losses". Specifically, the essay proposes two cross-sectional categorizations of cyber losses for risk aggregation: *cross-industry* (hacking, lost electronic device, unintended disclosure and insider breach) and *cross-breach type* (banking/insurance, governmental entity, medical service, business entity and educational institution). Applying the collective risk model and a high-dimensional dependence model (vine models), the essay concludes the best fit models for loss frequency (negative binomial), loss severity (lognormal) and the dependence structure (R-Vine). It further finds that the risk aggregation under the cross-industry setting generates higher correlated risk for the cyber-portfolio, which can lead cyber-insurers to remain safe from potential simultaneous cyber risks.

Although the first essay provides an answer on the demand for risk modeling and aggregation methodologies driven by a data breach database, it is not sufficient to fully figure out how big the next extreme cyber loss could be. On the supply side of an insurance market, the size of an extreme event likely to occur in the future needs to be evaluated as a benchmark for the cover limit of a coverage. With regard to cyber risk, several studies have explored the estimation of extreme losses with a widely used Pareto-based model. However, some recent loss events have been more extreme than these estimations. The second essay, "Probable maximum cyber loss:

Empirical estimation and reinsurance design with public intervention", proposes an alternative approach to estimating extreme cyber losses using time-series analysis and loss maxima process. It demonstrates that the predicted loss amount likely to occur in the next five years (defined as probable maximum cyber loss in the essay) is almost seven times larger than recent literature's estimate using a widely used Pareto-based model. In particular, the estimates based on the more recent data period show a significant increase compared to those for the older period, with a structural break between pre-2014 and post-2014. Given the estimates of probable maximum cyber loss, the results suggest a three-layer reinsurance portfolio involving the public sector as well as the private sector and provides an empirical benchmark of premiums for the private sector and of average cost per exposure above the estimated cover limit for the public sector.

Despite the lack of understanding relevant to cyber risk and standard modeling, the cyber-insurance market has been growing rapidly in recent years. However, this market growth is less significant than what many experts previously predicted, and the absolute size of the market represented by the premium volume is obviously not comparable with other property and casualty markets. In particular, a behavioral bias on the demand side possibly hindering the explosive growth of the market has been observed in practice and has also been found in some heuristic studies about cyber risk. Such bias shows that although the awareness of cyber risk is increasing, agents are still reluctant to purchase cyber-insurance, especially until they experience a loss event. The third essay, "Decision-making on cyber risk management: Interaction between market insurance and risk control measures under prospect theory", offers a conceptual framework to explain such a demand side anomaly with consideration for loss aversion and interdependent risk as the nature of cyber risk. The essay sets self-protection as the reference point, reflecting the status quo of business parties in the interconnected network environment, and compares it with possible decisions on additional risk management tools—namely, market insurance and self-insurance. It finds that an agent with the reference point of self-protection as an essential effort on cyber risk management is likely to not invest in additional risk management measures, which implies a behavioral bias against cyber risk.

The backbone of this dissertation focusing on risk modeling and aggregation methods is not simply limited to the study of cyber risk, but also considers a possible improvement in insurance regulation. The fourth essay, "Risk aggregation in non-life insurance: Standard models vs. internal models", attempts to make this improvement, particularly for risk aggregation in standard models, by applying the linear correlation assumption between risk factors. By considering non-linear dependence structures for risk factors, it proposes an internal model undertaking specific parameters that regulatory frameworks usually do not count in their standard models. A comprehensive framework with the base-level aggregation between different assets or different lines of business and the top-level aggregation between asset portfolio and insurance portfolio is established using empirical datasets from the Korean and German markets. The internal model provides economic capitals to be compared with those from three regulatory frameworks: the Korean Risk-based Capital (K-RBC), Solvency II and Swiss Solvency Test (SST). The comparison reveals that standard models overestimate the economic capital by an average of 61.2% for the Korean case and 57.8% for the German case, implying that insurers can significantly reduce their risk capital using the proposed internal risk model. The regulated risk parameters in standard models address on average 34.3 percentage points and 29.0 percentage points of the total deviation for the Korean and German cases respectively, whereas, ceteris paribus, the rest of the deviation is led by the correlation parameters in the linear setting for each case.

Ultimately, this dissertation offers potential solutions to current challenges arising from the supply and demand sides of the cyber-insurance market. On the supply side, it answers questions about the lack of standard modeling driven by an empirical dataset for cyber risk and the limitation of the traditional estimation for the cover limit presented in the market (Essay I and II). On the demand side, it provides a framework to better understand a particular behavior against cyber risk observed in the market, especially when that behavior cannot be explained by the classical expected utility theory (Essay III). The dissertation contributes to the literature in the cyber risk context by taking into account the overall cyber-insurance market and penetrating urgent, demanding issues for the next movement. It also contributes to the general insurance context by suggesting an internal model for the estimation of the capital requirement for non-life insurers (Essay IV) using an optimal solution to risk modeling and aggregation issues inherent in existing regulatory frameworks. The findings and models proposed in this dissertation may not be applicable to all relevant fields, but they are believed to offer insights for further research in cyber risk and insurance regulation, which was one of the most driving and significant objectives of this dissertation.

Essay I

Copula approaches for modeling cross-sectional dependence of data breach losses

Abstract

Many experts claim that cyber risks are correlated, but there is not much supporting empirical evidence. We consider 3,327 data breach events from 2005 to 2016 and identify a significant asymmetric dependence of monthly losses in two cross-sectional settings: cross-industry losses in four categories by breach types (hacking, lost electronic device, unintended disclosure and insider breach) and cross-breach type losses in five categories by industries (banking and insurance, government, medical service, retail/other business and educational institution). To identify the method that best fits the dependence structure of the dataset, we implement copula modeling by separating the dependence into pairwise non-zero losses and zero loss arrivals. We model the former by pair copula construction (PCC) allowing for the flexible choice of copula functions, whereas the latter is modeled by Gaussian copula. We illustrate the usefulness of our results in two applications to risk measurement and pricing. Our findings are important for risk managers and actuaries who are designing cyber-insurance policies.

Keywords

Cyber risk, Data breach, Zero-inflated data, Pair copula construction, Vine copula, Risk measurement, Insurance pricing, Diversification effect

---

Martin Eling, Kwangmin Jung (2018).

This paper has been published in *Insurance: Mathematics and Economics*.

1 Introduction

Cyber risks are operational risks to information and technology assets that have consequences for confidentiality, availability, and integrity of information and information systems (Cebula and Young, 2010). Every day, media illustrate the growing economic and social importance of cyber risk (see, e.g., World Economic Forum, 2016). In addition, more businesses than ever are facing cyber risks and incurring considerable corporate losses (Allianz, 2015). Although numerous papers have researched cyber risks, there is a lack of understanding of how to model potential losses from cyber risks and how to price cyber-insurance (Böhme and Schwartz, 2010; Eling and Wirfs, 2019). One aspect often discussed in this context is what the dependence structure between cyber losses might look like. Many experts believe that cyber losses are correlated (see, e.g., Böhme and Schwartz, 2010; Ogut, Raghunathan and Menon, 2011), for example, because all companies are using the same software.

A few papers discuss the correlation of cyber risk, for example on information systems or infected computers (Böhme and Kataria, 2006; Herath and Herath, 2011; Mukhopadhyay et al., 2013; Shah, 2016; Xu and Hua, 2017; Peng et al., 2018)[1]; but there is still a need for a comprehensive empirical study that analyzes the dependence between actual cyber losses, and if such a dependence exists, what it looks like. The reason for the dearth of empirical research on this field might be a lack of data. The availability of data is, however, improving over time, especially with the emergence of the first database on data breaches.[2]

This paper aims at identifying the dependence structure between different cyber losses. For insurers, ex-post analysis on cyber losses is important because estimating the size of risk in a cyber-insurance risk pool is a key task in asset-liability management. The dependence structure in a cyber-insurance risk pool can provide diversification benefits; thus, our modelling helps to correctly identify premiums and capital requirements. We construct a high-dimensional dependence model of cyber losses using different copula methods. For this purpose, we consider 3,327 data breach events from 2005 to 2016 and apply the actuarial toolbox to identify

[1] Appendix A provides an overview of the literature and outlines the contribution of our paper. Only Böhme and Kataria (2006) consider a broader dataset (the number of potential attacks measured by honeypots), but they do not consider loss data and focus on the t-copula to capture potential tail dependencies. The other four papers rely either on simulations or on smaller datasets. Note that our focus is not on the global interconnection of different IT systems or computer networks, but on identifying the potential dependence structure of actual cyber losses, which is important for cyber risk management or to manage cyber-insurance portfolios.

[2] Since 2002, companies in many U.S. states have been legally required to report data breaches to their customers (NCSL, 2019) and with this data breach report databases are becoming increasingly available. Starting in 2018, companies in the European Union will be required to report data breach events (European Union, 2016); this will improve the availability of data. Private Rights Clearinghouse, a nonprofit organization in the U.S. is a good example of a database that has grown since 2005 (PRC, 2016).

the dependence structure between monthly loss events, in terms of both frequency and severity. We are interested in finding the dependence structure that most accurately describes the data, whether that structure is linear or non-linear. Since monthly losses include several zero values indicating no loss event in a certain month, we split the dependence structure into two parts: pairwise non-zero losses and zero loss arrivals. We then fit frequency and severity distributions using different parametric distributions and compound them by convolution. With the results of dependence modeling, we analyze the implications of the models in two applications to risk measurement and insurance pricing by aggregating cyber losses from different risk factors.

This paper contributes to cyber risk research in that we take two categorizations of cyber losses into account in an integrated structure and estimate a more accurate dependence structure of cyber losses in the risk pool. The two cross-sectional categorizations we consider are breach type (hacking, lost electronic device, unintended disclosure and insider attack) and industry (banking and insurance, governmental entity, medical service, retail/merchant and other business and educational institution); we call the former *cross-industry* structure and the latter *cross-breach type* structure.[3] Upon this cross-sectional setting, an up-to-date copula method, the pair copula construction (Aas et al., 2009), is used to build an empirical model for high-dimensional cyber risks. As a result, we find significant asymmetric tail dependence, providing evidence for non-linear dependence between different types of cyber risks. Our results are important for practitioners and regulators working on cyber risk management and for insurance underwriters working on the establishment of cyber-insurance policies.[4] The paper will motivate further research by outlining future research questions on the topic of cyber risk.

The rest of the paper is structured as follows. In Section 2, we describe the theoretical background on the high-dimensional copula method and the methodology of pair copula construction. Then in Section 3 the data is given. The results of the dependence modeling and applications to pricing and risk measurement are presented in Sections 4 and 5 respectively. Finally, the conclusion and possibilities for future research are shown in Section 6.

[3] Cross-industry setting consists of variables categorized by breach types, in which losses occurred across industry level, whereas cross-breach type setting contains variables categorized by industry, in which losses occurred across breach type level. This categorization is important because underwriting a cyber-insurance policy differentiates the sources of risk and industries, which have different risk exposures (Romanosky et al., 2017). The detail on the categorization is shown in the caption of Table 1.

[4] What risk managers are mainly concerned about is the likelihood of the tail risk from different risk factors (McNeil, Frey and Embrechts, 2005, p. 18). For an insurance company, loss aggregation is typically applied to the reserving process to measure the possible total loss amount from different lines of business (Kaas et al., 2008, Chapter 3). Similarly, each cyber risk factor can form an individual line of cyber insurance business with customized policies depending on a specific risk type or industry (Allianz, 2015; Eling and Wirfs, 2016). Thus, there is a need to analyze cyber risk in context of an insurance pool.

2 Theoretical background and methodology

2.1 Research background on cyber loss process

The loss process for cyber risk can be regarded as an operational loss process (Cebula and Young, 2010; Biener et al., 2015). Several methods have been developed to estimate an operational loss process. The loss distribution approach (LDA) has been widely used; it models the frequency and severity of operational risk losses separately (Panjer, 2006, Chapter 1.3). LDA is also frequently used to model underwriting claims in the collective risk model. It is based on the distributional fitting procedure for frequency and severity; the fitted distributions are then compounded by convolution (Wang, 1998; Frachot et al., 2001; McNeil, Frey and Embrechts, 2005, Chapter 10; Panjer, 2006, Chapter 6). A compound loss process estimated for each risk factor j is a full predictive distribution to account for parameter uncertainty and can be described as:

$$\lambda_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} X_{i,t}^{(j)}, \tag{1}$$

where $t = 1,2,\dots$ is discrete time in the monthly unit, $j = 1,\dots,d$ is a breach type (d = 4) or an industry (d = 5), $N_t^{(j)}$ is the monthly count (frequency) process, $X_{i,t}^{(j)}$ is the monthly severity process and $\lambda_t^{(j)}$ is the monthly compound process. We assume that the count process and the severity process are independent (Wang, 1998; Shevchenko, 2010). In addition, we assume that the claim severity process, $X_{i,t}^{(j)}$, is independent and identically distributed (i.i.d.); hence, we do not assume temporal dependency.[5] We also postulate that the aggregate claims process, $\lambda_t^{(j)}$, is a Markov process so that the development of claims at a certain time point does not rely on the development of the aggregate claims up to that time point (Bühlmann, 2007, p. 55).

As an example, a compound Poisson process with Poisson distribution for the frequency and lognormal or other continuous distribution for the severity is frequently used in operational risk

[5] A criticism of this assumption is that there might be temporal dependency in claim amounts (see Araichi, Peretti and Belkacem, 2016, for a general discussion in the context of auto insurance). Indeed, in our cyber context, a malicious attack by hacking (HACK) might trigger temporal dependency in losses on an hourly or perhaps a daily basis (e.g., Wannacry attack in 2017), but not on a monthly scale which is of interest for our modeling. Furthermore, it is not reasonable to assume temporal dependency for the other three risk sources considered in the paper (ELET, DISC and INSD), whose losses typically are firm-specific and occur independently over time; we could imagine temporal dependency within a firm, but not in the aggregate variables considered in this paper. We thus believe that for all four risks considered in this paper, the temporal structure of monthly data can be reasonably assumed away. Also empirically we do not observe non-stationarity and any serial dependence for our variables of interest (e.g., using the augmented Dickey-Fuller test; see Appendix C).

modeling (Panjer, 2006, Chapter 5). Once a compound process has been estimated for each risk factor, we apply a dependence model for different risk factors, which we assume constitute the risk pool of a cyber-insurance provider. One challenge of the empirical study in this paper is that data breach risks on a monthly basis contain a number of zero values, indicating that no loss occurred in a certain month (see column 3 in Table 1). This zero-inflation could generate a misspecification of a dependence model due to discontinuous probability function (Erhardt and Czado, 2012).[6] For this reason, following Erhardt and Czado (2012) and Brechmann, Czado and Paterlini (2014), we model the parametric dependence structure in two separate approaches: dependence in positive loss pairs and dependence in zero value arrival. Hereafter, dependence in positive loss pairs is denoted by non-zero pair dependence and dependence in the zero value arrival by a zero loss dependence structure. The non-zero pair dependence is built upon the equation (1), where monthly loss severity without zero loss is modeled by a parametric continuous distribution and monthly loss frequency is modeled by a parametric discrete distribution in Section 4.1. The zero loss dependence structure is based on a multivariate binary distribution, each margin of which gives 1 to zero value and 0 to non-zero value (Brechmann et al., 2014):[7]

$$v_k := \begin{cases} 1 & zero\ loss \\ 0 & otherwise \end{cases}, \tag{2}$$

where v_k is a binary random variable.

When $v_k(\lambda) = 0$, the random variable has a positive loss and we can derive the cumulative distribution function of v_k, P_{v_k}, with the probability of a positive loss:

$$P_{v_k} := \begin{cases} p_{v_k}(0) & Zero\ loss \\ 1 & otherwise \end{cases}, \tag{3}$$

where p_{v_k} is a probability mass function of v_k and $p_{v_k}(0)$ is the probability of a positive loss. The methodology for dependence modeling in the two approaches is described in Section 2.3.

2.2 Theoretical background on Copula modeling

The copula method is an effective and tractable way to identify complex, non-linear dependences inherent in multivariate distributions.[8] Copula functions can be classified into

---

6 We test the dependence of original zero-inflated data by elliptical copulas, Archimedean copulas, joint Archimedean copulas and rotated Archimedean copulas (90°, 180° and 270°); however, we find that any type of parametric copula function does not fit the zero-inflated dataset.

7 The mathematical description of a zero loss dependence structure is given in detail in Appendix D.

8 Copula modeling is widely used to examine dependence structures and helps to identify non-linear relations between different marginal distributions. The cyber risk literature that employs the copula method still has limitations. For example, Mukhopadhyay et al. (2013) make a normality assumption on each Bayesian network node, thereby using Gaussian copula to integrate all nodes. Furthermore, a simple copula method to identify

different classes. Elliptical copulas are the copula functions of elliptical distributions (e.g., Gaussian, student-t); thus, if a bivariate copula function belongs to the elliptical class, margins will in general belong to the elliptical distribution (Embrechts, Lindskog, and McNeil, 2001). However, elliptical copulas are limited to symmetrical distributions and dependency (Embrechts et al., 2001). As shown in finance and insurance research, there might be a strong asymmetric dependence and tail dependence, for instance, between stock returns or insurance losses that cannot be captured by a symmetric and linear dependence measure.[9] As an alternative, Archimedean copulas incorporate different asymmetric tail dependence structures. However, since they explain the dependence structure by a single parameter only (via the generating function), using simple Archimedean copula to analyze a multivariate dependence structure is restricted in a multivariate case (Embrechts et al., 2001).[10]

To resolve these problems of multivariate dependence modeling, several advanced techniques have been developed (Aas and Berg, 2009). Among them, the pair copula construction (PCC) method, which is also called the vine copula model, is used to reduce the dimension by pairing the variable set (Bedford and Cooke, 2001). In this sense, the high-dimensional copula analysis can be transformed to a bivariate analysis to be more tractable. PCC is flexible in that any type of copula class can be applied to the construction and there is no mathematical complexity when one uses different copula functions in the modeling (Aas and Berg, 2009). These advantages result from the fact that a high-dimensional density function can be factorized by marginal density functions and conditional density functions.[11]

PCC is a tree-based model, where one builds the first tree with given random variables and continues to construct another tree with conditional variables estimated from the previous tree. The conditional variables are generated by copula densities as moving forward to the next trees

the dependence of a high-dimensional structure is theoretically restricted due to lack of explanatory power (Embrechts and Hofert, 2013).

9 For instance, multivariate asset returns or derivatives are not appropriately described by linear correlation measures (Chiou and Tsay, 2008). It also has been shown that the data breach information that we look at in this paper is non-normal and heavy tailed (Edwards, Hofmeyr, and Forrest, 2016) such that linear dependence might not fully illustrate the dependence structure of the data breach risk.

10 Additionally, the single parameter for d-dimensional dependency can induce the permutation-symmetric property in multi-dimensional arguments, thereby resulting in exchangeability of margins (Savu and Trede, 2010). The exchangeability can be described as (in the three-dimensional case):

$$C(u_1, u_2, u_3) = C(u_1, u_3, u_2) = C(u_2, u_1, u_3) = \cdots$$

This property is called permutation-symmetric and the copula distribution is indifferent in d-exchangeable marginal variables; this exchangeability can become problematic when some variables come from the same sector and some from different sectors (Savu and Trede, 2010). Elliptical copulas also correspond to the exchangeability property in the bivariate case, but in a multi-dimensional case, it depends on the variance-covariance matrix of the marginal elliptical distributions (Harder, 2016).

11 Aas et al. (2009) develop this methodology in the inferential way by decomposing a multivariate distribution into bivariate unconditional and conditional distributions based on the mathematical proofs by Joe (1996).

(Aas and Berg, 2009). Three types of PCC have been developed (Kurowicka and Cooke, 2004; Czado, 2010): the drawable vine (D-Vine), the canonical vine (C-Vine) and the regular vine (R-Vine). The D-Vine is a hierarchical structure, the C-Vine is a dependence structure centered by a core risk factor and the R-Vine allows for more flexibility to structure dependency than the D-Vine and the C-Vine do. The D-Vine and the C-Vine can be represented by the R-Vine structure in accordance with the dependency in the first tree; thus, we focus on the R-Vine structure in our empirical modeling. A four-dimensional case for three types is illustrated in Figure 1; the mathematical definitions of the vine models are given in Appendix B.

In Figure 1, the marginal distribution in the first tree is transformed to an uniform distribution ($u_i \in [0,1], i = 1, \dots, d$) and estimated copula functions in the following trees are also accordingly transformed to uniform conditional distributions. Due to high dimensional optimization for the parameter estimation, a two-step approach consisting of marginal estimation and copula estimation is customary (Czado, Jeske and Hofmann, 2013). Joe and Xu (1996) introduce this two-step approach for the marginal estimation and the copula estimation using maximum likelihood method, known as the *inference function for margins (IFM)*.[12] The uniform distribution in our empirical study is estimated non-parametrically using the ranks of the observations, called *pseudo-observations* (Aas and Berg, 2009). The pseudo-observations are used for bivariate conditional copula functions in the sequential estimation (described in the next section) and the optimization of pairwise likelihood function with pseudo-observations is conducted by maximizing the pseudo-likelihood (Aas et al., 2009).[13]

---

12 A simple case of IFM can be described in the following. According to Sklar's theorem, we can describe the d-dimensional probability function for the random vector **X** as:

$$F(\mathbf{X};\ \tau_1, \dots, \tau_d, \boldsymbol{\theta}) = C(F_1(x_1; \tau_1), \dots, F_d(x_d; \tau_d); \boldsymbol{\theta}),$$

where $\tau_i, i = 1, \dots, d$ is a parameter of a marginal function F_i, $\boldsymbol{\theta}$ is a set of dependence parameters by the copula function C. In our case, the random vector X is continuous and we can derive the joint density function of the random variables as:

$$f(\mathbf{X};\ \tau_1, \dots, \tau_d, \boldsymbol{\theta}) = c(F_1(x_1; \tau_1), \dots, F_d(x_d; \tau_d); \boldsymbol{\theta}) \prod_{j=1}^{d} f_j(x_j; \tau_j),$$

where $c(\cdot)$ is a copula density function. Then, the log-likelihood function for the joint density function is:

$$L(\boldsymbol{\theta}, \tau_1, \dots, \tau_d) = \sum_{i=1}^{d} \log f(x_i; \tau_1, \dots, \tau_d, \boldsymbol{\theta}).$$

The parameter estimation by IFM is comprised of separate optimizations for univariate margins and the optimization of the d-dimensional log-likelihood with the dependence parameter. The derivation of the parameters for vine models using IFM has been studied in Haff (2013) and Czado, Jeske and Hofmann (2013).

13 The maximum pseudo-likelihood estimation (MPL) is introduced by Genest, Ghoudi and Rivest (1995) and has been developed in Chen and Fan (2006) for time-series copula modeling and in Aas et al. (2009) for pair copula construction. The estimation using MPL is basically a semi-parametric approach, consisting of non-parametric marginal transformation and parametric estimation for dependence parameters (Genest et al., 1995).

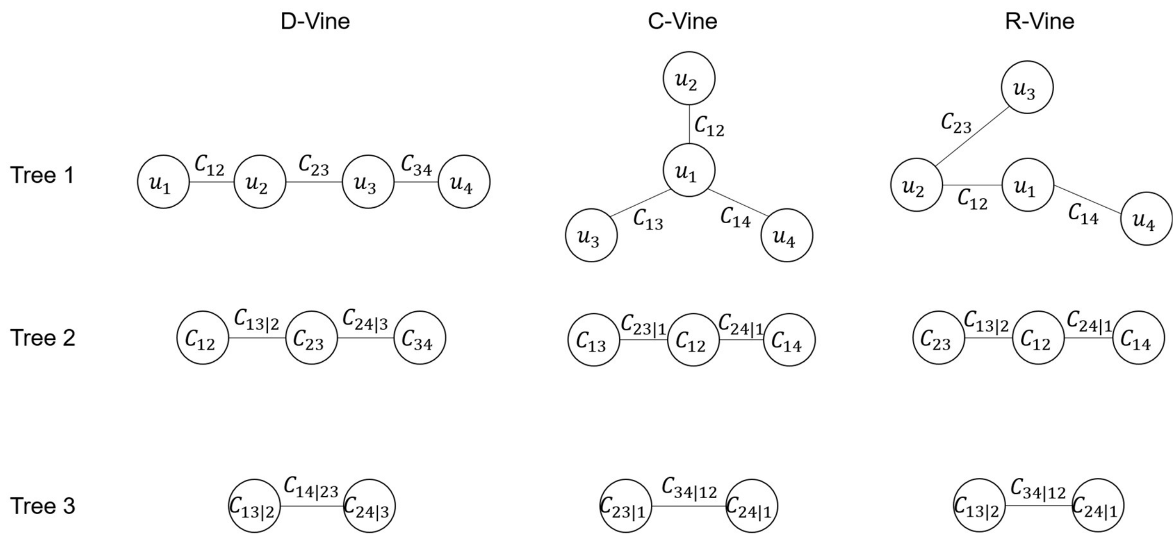


Figure 1. The pair copula structure (four-dimensional case). The D-Vine is a structure showing the dependency in a row and forming a hierarchical tree; hence, the variables are ordered by dependency. The C-Vine is a star-like structure, where a core variable placed in the center connects all other variables. The R-Vine flexibly links the variables by dependency without fixing a certain structure; thus, the R-Vine can also project the D-Vine and the C-Vine structures (see Appendix B for more detail).

2.3 Methodology

In non-zero pair dependence modeling, we apply the estimated compound process for cyber risk into different dependence models. There have been several studies and textbooks on modeling dependence in the operational risk context with copula based on compound distribution method (e.g., Wang, 1998; Frachot et al., 2001; Panjer, 2006, Chapter 8; Embrechts and Puccetti, 2008; Giacometti et al., 2008; Shevchenko, 2010). We estimate the R-Vine, Gaussian, Student-t, Gumbel and Clayton and determine the best fit structure for the cyber loss processes. We consider different copula functions in the R-Vine model: independence, normal, student-t, Clayton, Gumbel, Frank, Joe, survival Archimedean copulas and rotated Archimedean copulas (90° and 270°). We take rotated Archimedean copulas into consideration to model negatively dependent variables (Dissmann et al., 2013). Gaussian and student-t copula models are used in many fields because of their tractability; the Gumbel and Clayton copulas are frequently used in dependence modeling due to the presence of asymmetric tail dependence (Genest and Rivest, 1993; Demanta and McNeil, 2005; Rosenberg and Schuermann, 2006).

The statistical process in the empirical study is designed to identify whether non-linear dependence modeling with the R-Vine is the best fit for the dependence structure of cyber losses or if it can be better described by linear dependence modeling (Gaussian or student-t) or by simple asymmetric tail dependence modeling (Gumbel and Clayton) using uniform margins. Furthermore, we check the statistical test results for the C-Vine and the D-Vine to see whether the R-Vine method formulates the structure identical to the C-Vine or the D-Vine.

We first determine the structure of the first tree and sequentially implement bivariate copula modeling for each adjacent pair and select the best fit copula with the following rules:[14]

Step 1: Select the most appropriate copula candidates by the Vuong-Clarke test (Vuong, 1989; Clarke, 2007; Brechmann and Schepsmeier, 2013):[15]

Vuong test:
$$\tau = \frac{\frac{1}{n}\sum_{i=1}^{n} \log\left[\frac{c_1(u_i|\hat{\theta}_1)}{c_2(u_i|\hat{\theta}_2)}\right]}{\sqrt{\sum_{i=1}^{n}\left(\log\left[\frac{c_1(u_i|\hat{\theta}_1)}{c_2(u_i|\hat{\theta}_2)}\right] - E\left(\log\left[\frac{c_1(u_i|\hat{\theta}_1)}{c_2(u_i|\hat{\theta}_2)}\right]\right)\right)^2}}, \tag{4}$$

Clarke test:
$$\nu = \sum_{i=1}^{n} 1_{(0,\infty)}\left(\log\left[\frac{c_1(u_i|\hat{\theta}_1)}{c_2(u_i|\hat{\theta}_2)}\right]\right), \tag{5}$$

where $u_i \in [0,1]$, $i = 1, \dots, n$, is a uniform margin, n is the number of margins (dimension), $c_j(\cdot)$, $j = 1,2$ is a copula function to be compared, $\hat{\theta}_j$, $j = 1,2$ is the corresponding copula parameter.

Step 2: Check the statistical specification of those candidates by Cramer-von-Mises (CvM) goodness-of-fit test (Genest et al., 2009):

CvM test:
$$S_n = \int_{[0,1]^d} \left[\sqrt{n}\left(C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\right)\right]^2 dC_n(\mathbf{u}), \tag{6}$$

where $\mathbf{u}$ is a vector of uniform margins, $C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} 1_{(U_{i1}\le u_1,\dots,U_{id}\le u_d)}$ is an empirical copula with uniform margins and $C_{\theta_n}(\mathbf{u})$ is a copula function of interest with a parameter θ_n. We exclude the candidates that fail to be accepted by the CvM test.

Step 3: Select the most appropriate copula function by AIC.[16] We choose the function with the minimum AIC for every single pair dependence.

For zero-loss dependence modeling, we use the Gaussian copula. According to Erhardt and Czado (2012) and Brechmann et al. (2014), this dependence modeling on binary margins is not

---

[14] The choice of the fitted parametric copula function in each pair is still an open question (Erhardt and Czado, 2012). Furthermore, the pair copula method with parametric bivariate copula functions might have an inherent model risk arising from the misspecification of marginal copulas; thus, the sequential estimation of pair copulas could be unstable due to the error-prone selection (Scheffer and Weiss, 2017).

[15] In this step, we implement the comparison test proposed by Belgorodski (2010) between different copulas in the bivariate setting, the test that is conducted by using both tests from Vuong (1989) and Clarke (2007). The comparison test allocates "1" to a copula model if it is preferred to another, otherwise "-1" is allocated. No point is assigned if there is no preference between copulas. The final decision on the most appropriate candidates is made by the highest score after all possible comparisons.

[16] The information criterion developed by Akaike (1973) is defined as:
$$AIC := -2Loglik_i(\Theta|\boldsymbol{u}) + 2k,$$
where $Loglik_i(\cdot)$ is the log-likelihood of i-th model, $\Theta = (\theta_1, \dots, \theta_k)$ is a set of parameters, $\boldsymbol{u} = (u_1, \dots, u_d)$ is a d-dimensional set of uniform margins and k is the number of parameters.

standardized; hence, some parametric copulas (Archimedean copulas) and the vine copula method are not appropriate due to the non-existence of a closed form and the non-heterogeneous pairwise dependence. Instead, we follow Brechmann et al. (2014) and model this binary dependence by Gaussian copula with rank correlation parameters as an efficient tool for dependence modeling.

With the estimated structures from non-zero and zero loss dependency, we aggregate monthly losses from both non-zero and zero loss dependence structures as (Brechmann et al., 2014):

$$\lambda_k = v_k \times \lambda_k^0 + (1 - v_k) \times \lambda_k^+ = (1 - v_k) \times \lambda_k^+, \tag{7}$$

where $\lambda_k \geq 0$ $(k = 1, \dots, d)$ is the k-th loss out of the d-dimensional risk, $v_k \sim P_{v_k}$ is the occurrence of zero loss as a binary random variable. We denote zero loss by λ_k^0 and a positive loss by λ_k^+.

The aggregate loss distributions are applied to derive risk measures and insurance premiums in Section 5 and analyze the diversification effect of each dependence model. We compare the estimated dependence structures with the independence structure as a benchmark to see how our dependence modeling affects risk measures, insurance premiums, and diversification results. Figure C1 in Appendix C depicts the methodology of this study.

3 Data

We consider data breaches from January 1, 2005 to December 31, 2016, derived from the Privacy Rights Clearinghouse (PRC). The PRC dataset provides the entity, attack type, and the total number of records breached. PRC collects data breach information from government agencies and verifiable media sources.[17] In this dataset, most damages have been reported with positive breached records, whereas the rest of damages remains with zero record, because these cases either are not publicly acknowledged or are still being investigated (Edwards et al., 2016; Privacy Rights Clearinghouse, 2016). Such a data type with excess zeros is frequently presented in a variety of research areas, for example, insurance claim analyses and ecological studies,

[17] Given the absence of individual cyber-insurance loss data, we use the aggregate data for data breaches available from the PRC dataset. To use this dataset, the reliability of the data needs to be confirmed. Regarding reliability, each loss event has been confirmed at least by one major media source and is thus easily traceable and peer-reviewable. The dataset has already been used in numerous academic papers (e.g., Edwards et al., 2016; Eling and Loperfido, 2017; Rasoulian et al., 2017; Eling and Wirfs, 2019) and is widely accepted in practice. In terms of completeness, one limitation is that the data provider only includes losses that were publicly recognized (Edwards et al., 2016). However, PRC continuously updates the dataset to ensure the best possible completeness and the dataset is the largest public database about breached data information (Edwards et al., 2016).

showing that a dataset consists of a substantial proportion of zero values and extremely skewed distribution of non-zero values (Fletcher, Mackenzie, and Villouta, 2005; Erhardt and Czado, 2012). These zero values could give distorted information on the dependence structure of cyber losses, so we restrict breach data to those with counts (non-zero values). However, zero values appear again when we model the data on a monthly aggregate base, because some months have no breach events. Thus, we consider dependence structures in the presence of zero values.

Descriptive statistics of the monthly loss data are shown in Table 1.[18] Variables are categorized in two cross-sectional settings. The underlying dataset contains 3,327 data breach observations that we group into 144 monthly observations.[19] The monthly dataset is zero-inflated, because in some months no losses occur. It is observed in panel A of Table 1 that all severity distributions of risk factors are highly skewed and leptokurtic. Particularly, hacking risk (HACK) and retail/other business risk (BSE) categories have severer and more frequent losses than other variables do.[20]

Böhme and Kataria (2006) define the correlations of different cyber security attacks in two categories: internal and global correlation, which are consistent, respectively, with cross-industry type and cross-breach type in our case.[21] Based on Table 2 showing rank correlation matrices for the size of the severity per event with statistical significance, we can empirically classify different types of data breach risks with respect to rank dependency.[22] The classification is determined by the test statistics for rank correlations at the 10% critical level.

[18] We choose monthly average data as the standard timeframe that can provide less excessive zeros than, for example, weekly data or bi-weekly data that are not sufficient for meaningful aggregation.

[19] Although there is no clear standard on the sample size in the copula estimation, we do not consider all types of attacks included in the PRC dataset due to lack of data. Specifically, while variables in our model mainly include more than 100 data points, the subcategories of CARD (Payment card fraud), UNKN (Unknown attack) and NGO contain fewer than 20 data points. An analysis with small samples could give rise to a higher probability of assuming a false premise as true (Hogg, McKean, and Craig, 2005). This can be also applied to the case of our dataset that the small sample size might lead to a distorted dependence structure between risks or between industries. Moreover, the industries BSO (Business others) and BSR (Retail/Merchant business) are combined to BSE (Business entities apart from finance and insurance) and the attack types PORT (Portable device) and STAT (Stationary device) combined to ELET (Electronic devices). Once enough data are accumulated, these industries and types of attacks might be analyzed in greater detail.

[20] Plots in Appendix C offer graphical descriptions on the monthly data in both cross-sectional settings. Panels A and B of Figure C2 and C3 display the histograms of frequency, severity and log-severity for cross-sectional variables. Both frequency and severity are right skewed, whereas the distributions of log-severity data seem to be closer to the normal distribution. Figure C4 shows pairwise scatterplots with original monthly losses in panel A and with transformed uniform margins in panel B. Clustering in small losses is observed in panel A, but simultaneous extreme losses are scarce. Zero inflated pairs are more clearly identified in panel B, which are treated separately in the dependence modeling in Section 4.

[21] We determine cross-industry dependency (correlation) as equivalent to internal correlation in Böhme and Kataria (2006) in that we look into the dependence structure between losses from different breach types. Cross-breach type dependency is applied in the identical way.

[22] In this analysis, we use monthly average loss per risk factor, which can describe an expected association between individual losses from different risk factors.

Table 1. Descriptive Statistics of Monthly Loss Data

Panel A: Loss severity *(in the number of breach records)*

Variable	N	N of zeros	Mean	Std. Dev	Skew	Kurtosis	Min	Median	Max
HACK	144	4	5,378,936	24,606,777	9.141	95.518	0	52,316.7	270,131,250
ELET	144	12	185,391	868,203	9.378	97.310	0	24,300.3	9,550,998
DISC	144	12	353,149	2,736,163	10.925	125.022	0	7,336.0	31,835,867
INSD	144	32	305,371	2,457,333	10.643	118.941	0	1,700.0	28,200,000
BSF	144	51	1,990,783	10,801,531	6.548	42.735	0	1,255.5	80,000,000
GOV	144	22	640,856	3,104,316	6.470	44.048	0	14,330.2	25,333,655
MED	144	12	97,605	422,936	8.903	87.057	0	13,229.0	4,500,000
BSE	144	10	16,882,242	95,715,936	8.680	82.987	0	44,039.8	1,000,000,000
EDU	144	21	23,926	37,239	2.927	10.598	0	8,710.0	227,539

Panel B: Loss frequency

Variable	N	N of zeros	Mean	Std. Dev	Skew	Kurtosis	Min	Median	Max
HACK	144	4	5.764	4.591	1.862	4.968	0	4	27
ELET	144	12	7.160	5.752	1.244	1.913	0	6	30
DISC	144	12	4.757	3.343	0.654	-0.080	0	4	14
INSD	144	32	2.319	2.454	2.341	9.145	0	2	17
BSF	144	51	2.007	2.337	1.545	2.882	0	1	12
GOV	144	22	3.167	2.715	1.104	1.281	0	3	14
MED	144	12	7.125	6.671	1.350	1.581	0	5	31
BSE	144	10	3.757	3.299	2.019	5.667	0	3	18
EDU	144	21	3.944	3.148	0.848	0.749	0	3	15

Note (PRC, 2016):
Breach type: HACK = Hacked by outside party or infected by malware; ELET = Lost, discarded or stolen portable devices or Stationary computer loss; DISC = Unintended disclosure, e.g., sensitive information posed in public or mishandled or sent to the wrong party; INSD = Insider breach by employee, contractor or customer
Entity(Industry) type: BSF = Financial and insurance services; GOV = Government and military; MED = Healthcare, medical providers and medical insurance services; BSE = Retail/Merchant and other business parties; EDU = Educational institutions

Our empirical classification in Table 2 is consistent with Böhme and Kataria (2006) in that hacking attacks (including worms and viruses) and insider attacks fall into the high dependence category (internal correlation). This classification could result from the fact that hacking and insider breaches are malicious attacks which are expected to be more correlated than negligent risks such as disclosure risk. For example, if an insider who intends to breach some customer or financial information can access the company's security system to plant a malicious code into the system, an outsider might find it much easier to hack the system as well. Thus, the connection between a malicious insider and a hacker outside of the entity might facilitate a breach event. The correlation between HACK and ELET can be explained in a similar context, since ELET can be regarded as an internal risk. There might be also a case in which a driver by malware or ransomware could trigger an extreme loss event due to highly correlated information systems, but such a case could not be identified in this example. Note that other risks in Böhme and Kataria (2006), for example spyware/phishing and hardware failure, do not fully overlap with the risks in this study apart from insider/hacking attack, as our dataset is based on numerical values of data breach records.

Table 2. Rank Correlation Matrices of Cyber Losses

Cross-industry					*Cross-breach type*					
	HACK	ELET	DISC	INSD		BSF	GOV	MED	BSE	EDU
HACK	1				BSF	1				
ELET	0.119*	1			GOV	0.108*	1			
DISC	0.038	0.070	1		MED	0.091	0.071	1		
INSD	0.268***	-0.006	0.008	1	BSE	-0.067	-0.038	-0.066	1	
					EDU	-0.008	0.028	0.080	0.074	1

Note: The table shows the rank correlation matrices in both cross-sectional settings, which is comparable with the correlation classification in Table 1 of Böhme and Kataria (2006). The correlation measures in the table are calculated upon the monthly average losses, which we regard as the representative of individual loss processes. *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively.

4 Results

4.1 Marginal modeling

In this section, we conduct the distribution fitting for the frequency and severity data. Several studies provide evidence that the distribution of cyber risk frequencies follows a negative binomial distribution (see Edwards et al., 2016; Eling and Loperfido, 2017; Eling and Wirfs, 2019). These studies take into account Poisson and negative binomial as candidates, both of which are widely used in the insurance claim analysis. We consider three additional candidates: zero-inflated Poisson, zero-inflated negative binomal and Geometric, all of which could provide a better fit for our right-skewed and zero-inflated dataset.[23] We evaluate the distributions with the best fit based on the AIC and chi-squared goodness-of-fit test result. In Table 3 and 4, we display fitting results for frequency and severity in the cross-industry setting, whereas the results in the cross-breach type setting are illustrated in Appendix E. Monthly frequencies in Table 3 are best described by the negative binomial distribution, but some are better fitted by the zero-inflated negative binomial. However, we can conclude in line with the literature that the negative binomial distribution well describes the count process of data breach risks.

In the severity fitting (Table 4), we do not consider zero values because, otherwise, zero values are double considered both in freqency and severity. We test several continuous distributions known for right skewed distributions to fit severity: lognormal, skew normal, skew student-t, weibull, gamma, inverse Gaussian, cauchy, burr, generalized Pareto and Peaks-over-Threshold

[23] The zero-inflated Poisson distribution takes the distributional property from Poisson distribution, but more zero values are contained than expected (Zuur et al., 2009). The geometric distribution is a special case of the negative binomial, but different from the negative binomial in that it models the number of trials until the first success (Hogg et al., 2005). Hence, it focuses on the number of failures (the number of no breach events in our dataset). We compare the five discrete distributions by Chi-square goodness-of-fit test results.

(POT) with lognormal in the body and Pareto above 90% threshold.[24] These distributions are widely used in the operational risk modeling and the insurance context[25] and we choose them in accordance with the graphical description on the severity shape of body and tail (see panel B of Figure C2 and C3). The best-fitting distribution is assessed by minimizing the AIC and the Kolmogorov-Smirnov test (K-S test).

Table 3. Goodness-of-fit and Model Comparison for Loss Frequency

Distribution	Log-likelihood	AIC	Chisq-Test	
Panel A: Hacking (HACK)				
Poisson	-466.368	934.747	192.713	***
Zero-inflated Poisson	-460.888	925.777	>10,000	***
Negative Binomial	**-390.279**	**784.558**	**8.609**	
Zero-inflated Neg. Binomial	-390.279	786.558	53.347	***
Geometric	-408.058	818.116	52.927	***
Panel B: Electronic device (ELET)				
Poisson	-571.296	1,144.591	1,614.03	***
Zero-inflated Poisson	-522.838	1,049.675	>10,000	***
Negative Binomial	-430.241	864.482	8.347	
Zero-inflated Neg. Binomial	**-428.399**	**862.799**	**34.408**	
Geometric	-437.078	876.156	45.201	**
Panel C: Disclosure (DISC)				
Poisson	-403.684	809.368	148.609	***
Zero-inflated Poisson	-384.636	773.273	101.345	***
Negative Binomial	-366.322	736.644	4.266	
Zero-inflated Neg. Binomial	**-364.762**	**735.523**	**7.815**	
Geometric	-382.758	767.515	38.043	***
Panel D: Insider (INSD)				
Poisson	-321.857	645.713	108.172	***
Zero-inflated Poisson	-307.818	619.637	>10,000	***
Negative Binomial	**-288.792**	**581.583**	**3.093**	
Zero-inflated Neg. Binomial	-288.792	583.583	52.156	***
Geometric	-292.499	586.999	23.679	

Note: *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit distribution for each loss distribution based on AIC and goodness-of-fit test result.

[24] We test different thresholds for POT from 50% to 99% and determine 90% threshold with the minimum AIC as the optimal value for the dataset. AIC is derived from the minimum negative log-likelihood from the body model and the tail model with the number of parameters for each variable. Scarrott and MacDonald (2012) argue that the tradeoff between bias and variance of the parameter estimates needs to be considered when the optimal threshold is estimated. In particular, they state that a sufficiently high threshold is required to confirm that the asymptotic estimates are reliable and unbiased, thereby making 90% threshold a stable level for the parameter estimates in this case. We use the R package, `evmix`, to carry out POT fitting with continuity constraints (see Scarrott and MacDonald, 2012, for more detail on the continuity constraint at a threshold). The goodness-of-fit test in this distribution is implemented by two sample Smirnov test (Conover, 1971). The estimated shape parameters of the tail distribution (GPD) are for HACK: 0.1763, ELET: 0.2288, DISC: 0.3193, INSD: 0.3586, BSF: 0.1971, GOV: 0.1997, MED: 0.2357, BSE: 0.1977, EDU: 0.1533. We thus find that the distributions of DISC and INSD are more heavy-tailed than those of other variables.

[25] See Frachot, Georges and Roncalli (2001), Moscadelli (2004), Fu and Moncher (2004), Shevchenko (2011) and Frees, Lee and Yang (2016).

Table 4. Goodness-of-fit and Model Comparison for Loss Severity (non-zero values)

Distribution	Log-likelihood	AIC	K-S Test
Panel A: Hacking (HACK)			
Lognormal	**-1,953.437**	**3,910.873**	**0.105**
Skew normal	-2,582.646	5,169.292	0.907 ***
Skew t	-2,507.013	5,022.027	0.986 ***
Weibull	-1,972.419	3,948.838	0.160 ***
Gamma	-2,002.671	4,009.343	0.284 ***
Inverse Gaussian	-2,004.296	4,012.593	0.334 ***
Cauchy	-2,080.374	4,164.748	0.321 ***
Burr	-1,995.684	3,999.367	0.979 ***
GPD	-2,103.127	4,212.254	0.417 ***
POT (lognormal-GPD)	-1,952.288	3,912.576	0.121
Panel B: Electronic Device (ELET)			
Lognormal	**-1,623.860**	**3,251.719**	**0.049**
Skew normal	-1,997.298	3,998.597	0.962 ***
Skew t	-1,877.755	3,763.510	0.985 ***
Weibull	-1,639.617	3,283.235	0.111
Gamma	-1,663.419	3,330.838	0.213 ***
Inverse Gaussian	-1,638.764	3,281.527	0.179 **
Cauchy	-1,694.562	3,393.124	0.277 ***
Burr	-1,624.796	3,257.591	0.059
GPD	-1,680.591	3,367.182	0.351 ***
POT (lognormal-GPD)	-1,623.369	3,254.737	0.138
Panel C: Disclosure (DISC)			
Lognormal	-1,521.159	3,046.317	0.073
Skew normal	-2,148.995	4,301.990	0.977 ***
Skew t	-2,009.115	4,026.229	0.985 ***
Weibull	-1,547.675	3,099.350	0.149 **
Gamma	-1,596.578	3,197.156	0.290 ***
Inverse Gaussian	-1,566.356	3,136.713	0.279 ***
Cauchy	-1,592.378	3,188.756	0.282 ***
Burr	**-1,515.401**	**3,038.802**	**0.034**
GPD	-1,546.286	3,098.573	0.226 ***
POT (lognormal-GPD)	-1,516.954	3,041.908	0.083
Panel D: Insider (INSD)			
Lognormal	**-1,216.649**	**2,437.298**	**0.079**
Skew normal	-1,820.389	3,644.778	0.973 ***
Skew t	-1,697.613	3,403.226	0.982 ***
Weibull	-1,233.527	2,471.054	0.128 **
Gamma	-1,273.039	2,550.078	0.266 ***
Inverse Gaussian	-1,253.774	2,511.548	0.314 ***
Cauchy	-1,310.905	2,625.809	0.327 ***
Burr	-1,215.820	2,439.641	0.045
GPD	-1,317.948	2,641.896	0.464 ***
POT (lognormal-GPD)	-1,215.768	2,439.537	0.143

Note: *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit distribution for each loss distribution based on AIC and goodness-of-fit test result.

Table 4 shows that the lognormal distribution is the best fit for most severity distributions (HACK, ELET and INSD) and DISC is better fitted by burr[26], which implies that cyber loss

[26] Burr distribution is a continuous probability distribution allowing for only non-negative values. This distribution is statistically connected to Pareto distribution and consists of 12 different types as a distribution

severity is long-tailed in general, but extreme losses on the right tail are not necessarily modeled by an infinite-mean model from extreme value theory. The estimation for cross-breach type frequency and severity distributions in Appendix E yields similar results; for instance, for severity the lognormal shows the best fit in three out of five cases, while the burr is the best fit for BSE and Weibull for EDU (see Table E1, E2). In Appendix E, we provide the graphical diagnosis in the fitting outcomes using QQ plots and CDF plots for both cross-sectional cases, where the plots illustrate good fits of the estimated distributions for the risk factors by showing an almost exact match of the theoretical line and the fitting line (see Figure E1, E2).

4.2 Modeling non-zero pair dependence

We now model the dependence structures for pairwise non-zero losses in different dependence settings. We estimate the R-Vine, Gaussian, Student-t, Gumbel and Clayton and determine the best fit structure for the cyber loss processes. Kendall's tau is used to order random variables in the first tree since Kendall's tau can estimate non-parametric correlation independently of the hypothetic distribution and provide a global measure for non-Gaussian families (Dissmann et al., 2013).

Figure 2 illustrates the orders of uniform margins with Kendall's tau at each edge in the R-Vine structure.[27] In the cross-industry setting, ELET connects other risk factors under strong dependency, which draws the identical structure to the C-Vine structure. Similarly, EDU plays a central role in the cross-breach type setting except for indirect connection to MED, providing a more flexible structure than that of D-Vine or C-Vine. It can be assumed that the dependences in the first tree are stronger than those in the other trees because the conditional correlation between variables given a certain variable is smaller than the unconditional correlation in the first tree (see Dissmann et al., 2013, Section 3.1). This property reduces the number of the model parameters by using independence copula in the later trees when the dependence parameters in these trees are close to independence.[28] For this reason, we consider the independence copula in the selection process and find its validity in the later trees in the estimation (see Table 6).

family (Kleiber and Kotz, 2003, Section 2.3). Among them, Burr type XII is most widely used and known; thus, here we use Burr type XII for our severity analysis (see, e.g., Frees and Valdez, 2008; Frees et al., 2016).

[27] The measures of Kendall's tau in Figure 2 are based on the monthly loss sum per risk factor, which better describe the estimated compounding process. For this reason, the rank correlations are different from those in Table 2.

[28] Typically, modeling R-Vine structure is computationally intensive due to a substantial number of possible R-Vine structures (Dissmann et al., 2013).

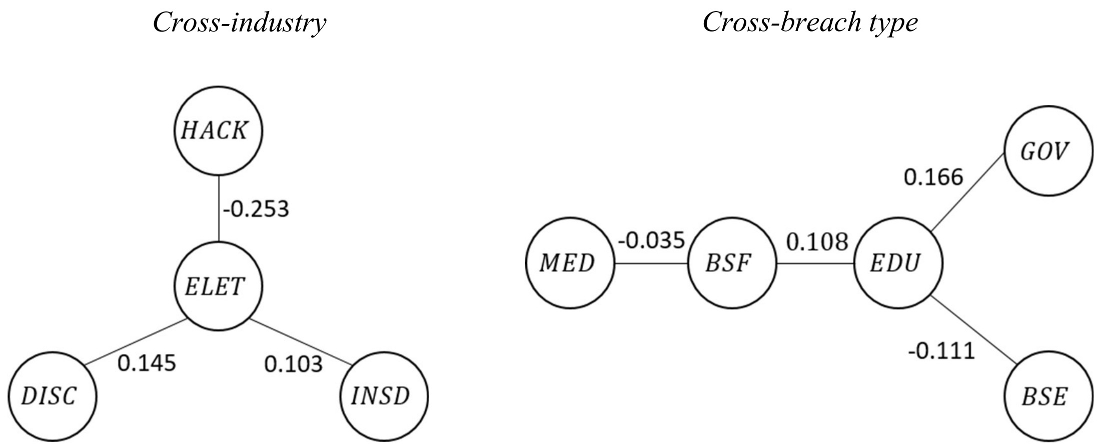


Figure 2. The ordering of variables in the first tree (R-Vine)

Table 5 shows the goodness-of-fit results and the information criteria of pairwise dependence structures. Overall, the pair copula structures (D-Vine, C-Vine and R-Vine) turn out to be the better fit to describe the dependency in both cross-industry risks and cross-breach type risks according to AIC and goodness-of-fit results than elliptical copulas and Archimedean copulas. Among vine models, the R-Vine structure is the most appropriate model for data breach losses and is identical to the C-Vine structure in the cross-industry setting (see Figure 2).

Based on the results from Table 5, the estimation parameters for the best fit pair copula structure (R-Vine) are illustrated in Table 6. In addition, the tree structures in both cross-sectional settings are graphically illustrated in Appendix F (Figure F1). Parameters are obtained by the maximum likelihood estimation based on the copula function fitted for each pair. As above-mentioned, ELET and EDU play key roles in the relationship of the risk factors. In the cross-industry setting, ELET is placed in the center of the dependence being connected to all other risk factors in the first tree; similarly in the cross-breach type setting EDU is directly connected to the other factors except for MED. We observe lower tail dependence between DISC/INSD and ELET (θ_{23} and θ_{42} in panel A of Table 6) and between GOV/BSF and EDU (θ_{52} and θ_{51} in panel B of Table 6). Smaller losses from DISC/INSD and ELET as well as GOV/BSF and EDU are thus dominant in the dependence structure. Moreover, ELET and HACK are negatively dependent described by 270° rotated Clayton copula, indicating a tendency to have a small loss by ELET and a larger loss by HACK at the same time. This also applies to the relationship between EDU and BSE.[29]

[29] Estimated 270° Clayton copula indicates the correlation between small loss of ELET and large loss of HACK. Similarly, estimated 90° Joe demonstrates the correlation between small loss of EDU and large loss of BSE.

Table 5. Comparison of Dependence Models by Different Pair Copula Structures

Model		Cross-industry LogLik	Cross-industry AIC	Cross-industry GoF test		Cross-breach type LogLik	Cross-breach type AIC	Cross-breach type GoF test	
PCC	D-Vine	22.154	-32.308	12.842	**	19.952	-19.904	15.842	**
	C-Vine	27.840	-43.680	10.003		25.654	-37.307	16.365	*
	R-Vine	**27.840**	**-43.680**	**10.003**		**26.274**	**-38.548**	**32.822**	
Elliptical	Gaussian	0.0001	1.999	0.057	**	0.0005	1.999	0.032	**
	Student-t	16.472	-18.943	0.977		15.388	-8.776	1.862	*
Archime-dean	Gumbel	0.048	1.904	0.055	**	0.013	1.974	0.033	**
	Clayton	1.183	-0.365	0.734	**	4.740	-7.481	0.083	**

Note: The table illustrates the results of statistical tests in both cross-sectional settings: Log-likelihood, Akaike Information Criteria and goodness-of-fit test. These measures help compare the model fits and determine which model is the best fit for each dependence structure. The parametric PCCs (C-Vine and R-Vine) are sequentially determined by R-package `VineCopula` and the D-Vine structure is estimated via `CDVine`.[30] *,**,*** indicate that the p-value of GoF test is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit distribution for each loss distribution based on AIC and goodness-of-fit test result.

This common relationship can be explained by the fact that the risk by losing an electronic device containing personal/corporate data can lead to a small loss due to the limit of the device, whereas the risk by hacking can be significantly large due to the interconnected system in the corporation. In the similar context, it can be implied that a business entity operating small sub-units such as a university or school is more likely to be exposed to less risk than other entities possessing a centralized and complex operation in the retail/franchise industry are (Wheatley et al., 2016).

The asymmetric tail dependency on the left tail has implications for an insurance company managing a cyber-insurance portfolio in that data breach losses with high frequency and low severity could mainly constitute the risk pool. That is, small losses of data breach can more frequently occur, whereas the frequency of large losses tends to be rather rare and stable (Wheatley et al., 2016). If extremely large losses frequently occur simultaneously, an insurance company might need more reserves to prepare against those simultaneous claims and these losses might not be insured. In line with the analysis by Biener, Eling and Wirfs (2015) we conclude that cyber risk can be insured in terms of the criteria on maximum possible loss and loss exposure. Such an asymmetric co-movement in losses could not be captured by a linear

[30] Note that the C-Vine and the D-Vine in Table 5 are unique structures by determining the shape of the first tree in accordance with the structure of each vine (see Figure 1). In the implementation of the vine models the package `VineCopula` might convert the C-Vine and the D-Vine to the optimal R-Vine model because the R-Vine can represent the two other models depending on the dependence structure of variables. The C-Vine estimation can be implemented via this R-package using the default option for this model in the sequential estimation, whereas the D-Vine estimation needs to be conducted via the R-package, `CDVine`, to obtain a unique D-Vine structure. `CDVine` is still available in R, but not actively developed anymore so that the package authors ask users to transfer to `VineCopula` package. To model a unique D-Vine structure in the `VineCopula` package, one can employ a function facilitating to convert the D-Vine structure to an R-Vine matrix representation.

dependence modeling, which might lead to an inaccurate estimation on the risk level for the insurance company.

Table 6. Parameter Estimations of Flexible Pair Copula Structure

Copula	Parameter	Lower-tail dependency	Upper-tail dependency
Panel A: Cross-industry			
θ_{21}; $C_{ELET,HACK}$ (270° Rotated Clayton)	-0.60	-	-
θ_{23}; $C_{ELET,DISC}$ (Survival Joe)	1.46	0.391	-
θ_{42}; $C_{ELET,INSD}$ (Survival Joe)	1.27	0.275	-
$\theta_{13\|2}$; $C_{HACK,DISC\|ELET}$ (Frank)	0.92	-	-
$\theta_{41\|2}$; $C_{INSD,HACK\|ELET}$ (Student-t)	0.08 (df=4.04)	0.093	0.093
$\theta_{43\|1,2}$; $C_{INSD,DISC\|HACK,ELET}$ (Independence)	-	-	-
Panel B: Cross-breach type			
θ_{52}; $C_{EDU;GOV}$ (Survival Joe)	1.52	0.424	-
θ_{51}; $C_{EDU,BSF}$ (Survival Joe)	1.44	0.382	-
θ_{54}; $C_{EDU,BSE}$ (90° Rotated Joe)	-1.29	-	-
θ_{13}; $C_{BSF,MED}$ (Student-t)	-0.05 (df=2.69)	0.121	0.121
$\theta_{12\|5}$; $C_{BSF,GOV\|EDU}$ (Independence)	-	-	-
$\theta_{41\|5}$; $C_{BSE,BSF\|EDU}$ (Independence)	-	-	-
$\theta_{53\|1}$; $C_{EDU,MED\|BSF}$ (Independence)	-	-	-
$\theta_{42\|15}$; $C_{BSE,GOV\|BSF,EDU}$ (Independence)	-	-	-
$\theta_{23\|51}$; $C_{GOV,MED\|EDU,BSF}$ (Survival Joe)	1.12	0.146	-
$\theta_{43\|251}$; $C_{BSE,MED\|GOV,EDU,BSF}$ (Survival Joe)	1.12	0.141	-

Note: The parametric PCCs are sequentially determined by R-package `VineCopula`. The numbers as subscripts of the copula parameters indicate in the following:
Cross-industry: 1 = HACK (Hacking), 2 = ELET (Lost electronic device), 3 = DISC (Disclosure), 4 = INSD (Insider attack);
Cross-breach type: 1 = BSF (Banking and insurance), 2 = GOV (Governmental entity), 3 = MED (Medical service), 4 = BSE (Retail/Merchant and other business), 5 = EDU (Educational institution).

4.3 Modeling zero loss dependence

If we want to derive risk measures from a portfolio loss distribution with excessive zeros, zero value dependence needs to be separately considered (Brechmann et al., 2014). As mentioned in Section 2.2, each variable is binary distributed, where 1 is given to a zero value and 0 is given to a non-zero value. We implement Gaussian copula modeling to identify the dependence of binary distributions, displaying how dependent zero value arrival processes are. For the Gaussian copula model, we generate parameter matrices for both cross-industry and cross-breach type (see Table 7).

We observe very weak dependences in zero loss arrivals where many of the parameters are close to 0. Here, a positive dependence means that risks are more likely to simultaneously arrive than arrive not simultaneously, whereas a negative dependence indicates a higher likelihood of counter-monotonicity in zero loss arrivals. It can be inferred that there is no significant dependency between zero losses. We generate the probability of zero losses for each loss distribution from the Gaussian dependence structure and use it to measure risk levels and price insurance premiums by the equation (11) in the next section.

Table 7. Dependence Parameters of Gaussian Copula for Zero Loss Arrival

Cross-industry					*Cross-breach type*					
	HACK	ELET	DISC	INSD		BSF	GOV	MED	BSE	EDU
HACK	1				BSF	1				
ELET	-0.017	1			GOV	0.004	1			
DISC	-0.009	0.0002	1		MED	-0.008	-0.003	1		
INSD	-0.006	0.0005	-0.007	1	BSE	0.010	0.018	0.004	1	
					EDU	-0.009	0.001	-0.006	-0.012	1

Note: The table displays the correlation matrices of zero-loss processes estimated by Gaussian copula model. The acronyms of variables are described as follows:
Cross-industry: HACK = Hacking; ELET = Lost electronic device; DISC = Disclosure; INSD = Insider attack
Cross-breach type: BSF = Banking and insurance; GOV = Governmental entity; MED = Medical service; BSE = Retail/Merchant and other business; EDU = Educational institution

4.4 Statistical difference in loss aggregate distributions

Prior to the applications, we want to evaluate whether there is a difference in the estimated loss aggregate distributions. To achieve this, we conduct two statistical tests: the Wilcoxon signed-rank test, which is a non-parametric statistical hypothesis test to compare two independent samples by using population mean ranks, and the Kolmogorov-Smirnov test (Wilcoxon, 1945; Smirnov, 1948). We thus test whether two loss aggregate distributions from different dependence structures are significantly different. Each aggregate distribution of breached records is generated based on the equation (11) with 500,000 simulations. Four comparisons are tested among the R-Vine structure, the independence structure, the Gaussian structure and the empirical structure, which are of main interest in the applications. The null hypothesis is:[31]

$$H_0: L_i - L_j = 0 \ (i \neq j),$$

where L_i is a loss vector from a dependence structure considered ($i = 1, \dots, 3$).

The results in Table 8 show that the differences in aggregate distributions between the R-Vine structure and the independence structure, between the independence structure and the empirical structure and between the R-Vine structure and the Gaussian structure are statistically significant for both cross-industry and cross-breach type settings; no significant difference between the R-Vine structure and the empirical structure is identified. Tests in both cross-

[31] All tests are two-sided paired difference tests as related to the null hypothesis.

sectional settings arrive at the same conclusion and the testing results are confirmed at the 10% critical level.[32] The R-Vine model is thus close to the empirical model, but different from the independence and Gaussian models.

Table 8. The Result of Statistical Difference Test

		Wilcoxon test	K-S test
Cross-industry	$L_{Rvine} - L_{emp}$	242,297	0.024
	$L_{Rvine} - L_{ind}$	268,598**	0.069***
	$L_{Rvine} - L_{Gauss}$	231,743*	0.059*
	$L_{Ind} - L_{emp}$	229,505**	0.069***
Cross-breach type	$L_{Rvine} - L_{emp}$	259,209	0.035
	$L_{Rvine} - L_{ind}$	279,855***	0.078***
	$L_{Rvine} - L_{Gauss}$	265,948*	0.068**
	$L_{Ind} - L_{emp}$	229,635**	0.076***

Note: The table shows how statistically close different aggregate distributions are based on two statistical tests. Wilcoxon test is a non-parametric statistical test using ranks of two distributions of interest and K-S test is a Goodness-of-Fit test using an empirical distribution to measure the distance. *,**,*** indicate p-value less than a significant level at 10%, 5% and 1% respectively.

5 Applications to risk measurement and pricing

We now apply the estimated PCC dependence structure from Section 4 to risk measurement and insurance pricing. The applications are based on the aggregate distributions from different dependence models with 500,000 copula-simulated values. The risk measures and prices using PCC are then compared with the measures under the a) independence assumption, b) linear dependence assumption (Gaussian copula), c) linear and symmetric tail dependence assumption (Student-t copula) and d) the dependence structure from the empirical copula.[33]

The applications are again carried out in two cross-sectional settings: cross-industry and cross-breach type. With regard to insurance pricing, since the values of the aggregated distribution are the number of breached records, it is necessary to convert them to dollars to derive insurance prices. Several cyber security companies offer data breach cost calculators (e.g., Imperva, 2016; eRiskHub, 2016; FireEye, 2016). However, there is no established method on how to conduct

32 The test result between R-Vine and Gaussian is determined at the 10% level, whereas other test results are confirmed at the 5% level. The statistical distance between R-Vine and Gaussian is relatively closer than the distances between R-Vine and independence and between independence and empirical structure.

33 The independence assumption is a baseline benchmark without any dependence modeling. Gaussian and student-t copula models are used in many fields; hence, the measures by these two models might be close to the values that are used in practice. Lastly, the empirical setting can serve as a benchmark for our estimated model on how close our model is to the historically, empirically observed dependency.

these calculations.[34] Jacobs (2014) analyzed the Ponemon datasets[35] for 2013 and 2014 and proposed the following regression model to compute data breach cost that we use to calculate insurance prices:

$$Dollar\ loss = \exp[7.68 + 0.76 * \ln(breach\ records)]. \tag{8}$$

Two risk measures are derived from the aggregate distribution of breached records:

Value at Risk: $$VaR_{(1-\alpha)}(X) = \inf\{X \in \mathbb{R}: X \geq F_X^{-1}(1-\alpha)\}, \tag{9}$$

Expected Shortfall: $$ES_{(1-\alpha)}(X) = E[X \in \mathbb{R}:\ X \mid X \geq F_X^{-1}(1-\alpha)], \tag{10}$$

where X is a non-negative random variable with finite variance, α is a risk threshold and F is the aggregated loss distribution. We calculate risk measures at critical levels of 90%, 95%, 99% and 99.5% (i.e., α is 10%, 5%, 1% and 0.5%).[36] In addition, we derive insurance premiums using three pricing principles that incorporate different expected utility functions (see Embrechts, 2000):

Fair Premium: $$P = EX, \tag{11}$$

Standard Deviation Principle: $$P = EX + \delta \cdot \sqrt{Var(X)}, \tag{12}$$

Exponential Principle: $$P = \frac{1}{\gamma}\ln\big(E(e^{\gamma X})\big), \tag{13}$$

where δ is a cost-loading and γ is the risk-aversion parameter.[37]

In addition to risk measurement and insurance pricing, we investigate the effect of each dependence model on portfolio diversification. The effect demonstrates the extent to which risk

[34] The cost per data breach might be different for company size, industry or other factors and insured losses by a cyber-insurance provider typically consist of data recovery/replacement of intellectual property, third-party liability and forensics (Allianz, 2015). Thus, the estimated risk measures in this section do not perfectly reflect the actually insured losses in reality.

[35] Ponemon Institute (2016) provided parameters about cost of data breach by different years, industries and type of attacks in its annual reports. The approximation of Jacobs (2014) to transfer the number of records breached into actual loss data is useful to carry out the first applications, but is clearly only a crude and rough approximation of the real loss.

[36] We choose 99.5% as the most extreme critical level estimated since Solvency II requires VaR at 99.5% as the equity capital. Swiss Solvency Test (SST) requires Tail VaR (Expected shortfall) at 99% level (FINMA, 2016).

[37] The equation (11) is based on the *zero-utility* model where the expected utility before paying the insurance premium is the same as that after paying the insurance premium, if the policyholder is risk neutral (Kaas et al., 2008). The equation (12) includes a cost-loading, δ, representing the level of transaction cost and requiring some risk aversion to accept the insurance contract. Following Mukhopadhyay et al. (2013), we assume a cost loading of 0.1 for the standard deviation principle. The equation (13) explicitly includes the exponential utility function $U(X) = -\gamma e^{-\gamma X}$, where increasing γ augments the premium, implying that a more risk averse insured is willing to pay a bigger premium to cover the risk. If γ converges to 0, the premium converges to the fair premium. We specify the risk aversion level as 1/1000, 1/10000, 1/100000 to see the difference in premiums with different risk aversion levels.

measures of the aggregate distribution under the estimated model are reduced compared to the sum of the risk measures from the individual loss distributions. That is the relative risk reduction compared to the comonotonicity assumption (the equally weighted summation of individual risks; Brechmann et al., 2014). We estimate the effects using the expected shortfall since this measure is coherent and satisfies sub-additivity (Artzner et al., 1999; Acerbi and Tasche, 2002).[38] The loss aggregation is based on the equally weighted aggregation that can be derived (Jorion, 2007):

$$\varphi_{j,1-\alpha} = \frac{ES_{j,1-\alpha}\left(\sum_{k=1}^{d} \lambda_{jk}\right) - \sum_{k=1}^{d} ES(\lambda_{jk})}{\sum_{k=1}^{d} ES(\lambda_{jk})}, \tag{14}$$

where $\varphi_{j,1-\alpha}$ is a diversification effect of j-th model at $1-\alpha$ quantile ($j = 1, \ldots, 3$) and λ_{jk} is a loss vector of k-th risk from j-th model ($k = 1, \ldots, d$).

The risk measures in panel A of Table 9 indicate the potential monthly breach records at different critical levels for the entire U.S. market consisting of risks from five industries (BSF, GOV, MED, BSE and EDU) and exposed to four breach types (HACK, ELET, DISC and INSD). For instance, if we use the pair copula model for cyber risk aggregation with the cross-industry risk categorization, we expect 59.54 million monthly breach records at the 90% critical level and around 1.03 billion at the 99.5% level from the potential U.S. risk pool.

The independence structure produces a lower level of risk than other structures do in both cross-sectional settings, resulting from the fact that the structure does not consider correlated risk among the risk factors in the portfolio. The student-t provides higher values at a more extreme level (e.g., 99 or 99.5%), especially in case of the expected shortfall. This might be because the restrictive model using student-t copula with only one tail dependence parameter can hinder the accurate estimation of the true tail dependency and causes the risk measure to be overestimated (Brechmann et al., 2014). In line with the statistical tests from Section 4.4, the risk measures with the R-Vine structure are closer to the empirical structure than in other cases, demonstrating

[38] Sub-additivity can be defined in the following. Let X and Y be two risk factors and let ρ be a function of a risk measure. The risk measure, ρ, which can be defined in the real space of random variables is sub-additive if it satisfies the following property (Artzner et al., 1999):

$$\rho(\mathrm{X} + \mathrm{Y}) \leq \rho(\mathrm{X}) + \rho(\mathrm{Y}).$$

This property does not hold in case of Value-at-Risk so that there could exist the following case in Value-at-Risk:

$$\rho(\mathrm{X} + \mathrm{Y}) > \rho(\mathrm{X}) + \rho(\mathrm{Y}).$$

that the pairwise dependence structure with different copula functions serves as a useful tool for this modeling.[39]

Table 9. Applications to Risk Measurement and Insurance Pricing

Panel A: Risk measurement *(in million breached records, monthly time horizon)*

Data Type	Dependence Structure	Value-at-Risk 90%	95%	99%	99.5%	Expected Shortfall 90%	95%	99%	99.5%
Cross-industry	Indep	46.87	157.88	764.80	994.87	288.41	472.25	1,011.61	1,050.55
	PCC	**59.54**	**163.22**	**984.61**	**1,041.85**	**313.98**	**522.75**	**1,033.12**	**1,053.11**
	Empirical	59.77	163.03	986.18	1,045.82	310.11	515.85	1,035.70	1,055.18
	Gaussian	55.91	161.43	980.39	1,040.28	301.51	502.24	1,033.78	1,055.87
	Student-t	57.20	161.06	983.76	1,045.13	306.85	512.52	1,039.03	1,062.97
Cross-breach type	Indep	45.39	152.59	732.02	946.05	276.49	452.80	957.41	983.84
	PCC	**51.60**	**153.99**	**930.56**	**930.88**	**276.09**	**467.91**	**933.49**	**936.16**
	Empirical	51.74	154.35	930.57	931.04	274.47	463.43	933.93	937.09
	Gaussian	51.78	156.06	930.62	931.25	280.95	475.16	934.82	938.79
	Student-t	50.88	154.35	930.60	932.09	279.24	474.29	942.20	953.38

Panel B: Insurance pricing *(monthly premium in million \$)*

Data Type	Dependence structure	Fair Premium	Standard Dev. Principle	Exponential Premium Principle $\gamma = 10^{-3}$	$\gamma = 10^{-4}$	$\gamma = 10^{-5}$
Cross-industry	Indep	724.81	938.58	10,924.60	1,065.05	748.35
	PCC	**761.97**	**979.94**	**10,911.78**	**1,114.47**	**786.59**
	Empirical	758.67	975.03	10,927.01	1,106.02	782.93
	Gaussian	739.42	952.01	10,906.24	1,074.83	762.84
	Student-t	749.61	964.88	11,024.94	1,094.51	773.63
Cross-breach type	Indep	688.44	894.99	10,257.92	999.85	710.38
	PCC	**689.86**	**889.76**	**9,763.18**	**978.17**	**710.52**
	Empirical	686.77	885.50	9,748.62	971.25	707.19
	Gaussian	695.86	897.84	9,812.90	990.53	716.96
	Student-t	685.96	888.05	9,929.22	982.63	707.09

Note: The risk measurements are specified by Value-at-Risk (VaR) and Expected Shortfall (ES) at three critical levels, 90%, 95%, 99% and 99.5%. For insurance pricing, we calculate the premium by three different pricing principles on an annual basis: fair premium principle, standard deviation principle and exponential premium principle. γ is the risk aversion parameter, where $\gamma \to 0$ indicates risk neutrality. The bold model is the preferred model from Section 4.

The risk measures estimated in the cross-industry setting are larger than the measures in the cross-breach type setting, although the number of risk factors in the cross-industry setting is smaller.[40] There are three plausible reasons for this outcome. First, it can result from the estimated dependence models and the dependence structure of the cross-industry risk pool incorporates a higher correlated risk (i.e., less number of pairs are modeled by independence

[39] Compared to these industry-level estimates, company-level estimates are derived in Appendix G, informing a cyber-insurer of the potential loss amount per event in the cyber-insurance portfolio. The company-level estimation is more complicated than the industry-level estimation due to additional assumptions required. For example, we need to reduce the dimension from aggregate level to individual level by using an estimator how many companies are breached in the U.S. market. Furthermore, we also need to break down the aggregate level of premium size into the individual level by specifying a certain industry and a certain risk type.

[40] Brechmann et al., (2014) also compare the estimated risk measures from two settings of operational risk (business line: BL and event type: ET). The difference between the log-scaled estimates in their paper (see Figure 6) is not significant and comparable with the difference between the estimates in panel A of Table 9 with log-transformation.

copula in the cross-industry setting than in the cross-breach type setting; see Table 6). Second, the likelihood of zero loss occurrence is affected by different zero loss dependences when integrated by the equation (7). Lastly, a smaller size in the cross-breach type setting could be addressed by the key factor of the dependence structure, EDU, which demonstrates the lowest severity among all considered risk factors. We thus suggest aggregating the cyber risks in the cross-industry setting, which could lead a cyber-insurer to be on the safe side against cyber-insurance claims.

Based on the results of risk measurement, cyber-insurance premiums are estimated in panel B of Table 9, again on a monthly and an aggregated industry level. Note that there are no cover limits or deductibles assumed in the pricing application and current cyber-insurance policies usually provide the protection against a certain type of risk with restricted coverage and consider internal factors of an insured.[41] Therefore, the estimated premium size will be different from that in practice. If we assume independence, \$724.81 million would be needed as fair premium to cover the possible monthly loss in the cyber-insurance portfolio with four types of risk and five industries. The value estimated for the PCC model (\$761.97 million) is 5.1% higher and very close to the empirical value (\$758.67 million); it thus seems that not enough premium could be collected when independence is assumed, especially in the cross-industry setting. Gaussian and student-t models lead to an underestimation of the insurance premium in the cross-industry setting, illustrating the need to be accurate in describing the dependence structure.[42] As with the risk measurement, the premium size in the cross-industry setting is generally higher than in the cross-breach type setting due to the stronger dependence.

Diversification effects of the estimated models are presented in Table 10 and Figure F2 (Appendix F). In line with the literature (see Brechmann et al., 2014) we observe diversification effects in all structures in both cross-industry and cross-breach type modeling and this benefit becomes larger at more extreme levels. Overall, the pair copula structure shows a bigger diversification effect across quantiles for both cross-sectional settings than other structures do, accounting for the strength of the pairwise dependence model.

41 We do not specify any details such as company size, revenue or type of security system in place. In addition, a range of risk types including malicious and accidental risks are considered in the price. However, risk classification in practice is based on the specific cyber risk profile of a customer, which relies on company size, industry, existing security systems and other factors (Allianz, 2015; KPMG, 2016).

42 In the cross-breach type setting, it is observed that the Gaussian model generates a higher fair premium than others do. This can be explained by a higher density of the aggregate distribution by the Gaussian model at less extreme levels, which can influence the expectation value of the distribution.

Table 10. Diversification Effects on ES per Quantile

Model	90%	95%	99.5%
Panel A: Cross-industry			
PCC	**-5.6%**	**-9.6%**	**-19.9%**
Empirical	-5.1%	-9.5%	-19.0%
Gaussian	-5.5%	-9.5%	-19.6%
Student-t	-5.4%	-9.5%	-19.0%
Panel B: Cross-breach type			
PCC	**-7.8%**	**-9.7%**	**-13.6%**
Empirical	-7.6%	-9.6%	-13.3%
Gaussian	-7.7%	-9.6%	-13.4%
Student-t	-7.0%	-9.0%	-11.9%

Note: The diversification effects are derived using expected shortfalls. The bold model is the preferred model from Section 4.

6 Conclusion and further research

In this paper, we implement the pair copula construction (PCC) with a range of parametric copulas to investigate the dependence structure of data breach losses. Since monthly breach records include excessive zero losses, we model the dependence by using non-zero pair dependence and zero-loss dependence. We describe the modeling results in two cross-sectional settings: cross-industry by four breach types (HACK, ELET, DISC and INSD) and cross-breach type by five industries (BSF, GOV, MED, BSE and EDU). We find a significant asymmetric tail dependence among risk factors, especially lower tail dependency dominated by small losses. This asymmetric tail dependency is estimated pairwise, resulting in a more accurate tail dependence estimation than the widely used elliptical models or Archimedean copula models. The likelihood of simultaneous small losses might in general not be a big concern for a cyber-insurance provider, however, when aggregating the estimated loss distributions, the insurer must take the potential dependence into account to avoid an underestimation of the risk and necessary premiums. We also show that on an aggregate industry level the U.S. could be faced with approximately 1 billion monthly breach records at 99.5% critical level (one event with such amount likely to occur over 200 months), in our case considering four breach types and five industries.

The estimated pair copula structure produces a higher level of potential loss than under independence assumption. Hence, there might be a possibility for a risk manager to underestimate the risk level when neglecting the potential correlated risk. In addition, the correlated risk in the cross-industry setting (between risk factors by breach types) turns out to be higher than the risk in the cross-breach type setting (between risk factors in different industries). This result illustrates the importance of determining the risk factors considered in underwriting and risk management of cyber risk. We propose considering the risk aggregation in the cross-industry setting, since this setting incorporates higher correlated risks in the portfolio, leads a cyber-insurer to be on the safe side with higher capital requirement against cyber risk claims and provides a higher diversification benefit at more extreme levels (see Figure F2). Given that cyber insurance policies typically cover different types of risk and the policies are then aggregated across companies from different industries, risk managers in insurance companies might follow this approach. We also show that if an insurer offers different types of data breach risks in the coverage, simultaneous losses on the tails might have different impacts on the capital basis when different dependence structures are considered. This finding is relevant for recent equity capital standards such as US Risk-Based Capital (RBC), Solvency II or the Swiss Solvency Test, which typically do not model non-linear dependence among risk factors.

This study identifies a non-linear dependence in data breach losses, but there are several limitations which open directions for future research. For example, the cyber-insurance market development must clearly define damage by a single risk and identify how this damage can be accurately measured. Moreover, the dependence structures in data breach loss could be affected by different characteristics of industries or other causes, such as geographical variation or the degree of the development of security system. For instance, we might expect the costs of data breaches to vary by industry, e.g., when comparing banking and healthcare. If more data on breach events are available across the globe, we might compare the dependence among different regions, because there could be geographically different appearance, size and frequency in cyber risk events. Moreover, the time variation in data breach risk should be studied in more detail, especially given the dynamic nature of cyber risk and the risk of change; it is thus not clear whether a dependence structure observed in historical data will also hold in the future. As indicated above, we could not include all types of data breaches in our analyses (payment card fraud, unknown attacks) and not all types of companies (NGO's), because the samples are too small. When enough data on those events and industries are accumulated, those should be also analyzed in detail. Additionally, the insurance pricing example presented here should not be interpreted as more than a first rough indication, because it is based on the number of breached data and not on actual loss information that needs to be verified when more and better information becomes available. Another interesting avenue for future research could be to dig deeper into the potential drivers of correlations.

Appendix A. Overview of literature and comparison with the present paper

Table A1. Summary of Literature on Dependence of Cyber Risk Analyzed with Copula Modeling

		BK06	HH11	MCSMS13	S16	PXXH18	Present paper
Modeling perspective	Data type	Honeypot data (worms and virus infection) – 183,000 data (Feb, 2003 – Sep, 2005)	ICSA survey – 15 data (number of viral infected computers and dollar loss)	Log data from security appliances (not details on the data provided)	Simulated copula-based prices with presumed parameters	Honeypot data – 1123 hours between Nov 4, 2010 and Dec 21, 2010	Data breaches (PRC) – 3,327 data (Jan, 2005 – Dec, 2016)
	Focus of Study	· Simulation-based estimation on premium pricing. · Empirical estimation of correlation on loss arrival.	Focusing on specifying Copula-based pricing models in the actuarial aspect.	Theoretical modeling by Copula-based Bayesian Belief Network (CBBN) and suggesting pricing model with utility theory.	Simulation-based estimation on dependence among sub-cyber risk losses affecting CLI pricing.	Modeling high-dimensional cyber attack losses.	**High-dimensional risk structures with Pair Copula Construction (PCC).**
	Copula modeling	t-copula to model the distribution of cross-firm risks	Archimedean copula (Clayton & Gumbel)	Gaussian copula with normality and linear correlation of Bayesian nodes	Gaussian, t and Gumbel copulas to model six-dimensional dependence	R-Vine with different copulas, Gaussian and t	**R-Vine with different copulas**, Gaussian, t and Archimedean
	Copula-dimension	Multivariate (High-dimension)	Bivariate	Bivariate	Multivariate (High-dimension)	Multivariate (High-dimension)	Multivariate (High-dimension)
	Evaluated Dependence	· Cross-attack risks · Cross-industry risks	Cross-attack risks	Cross-network vulnerability dependence	Cross-attack risks	Cross-attack risks	· **Cross-breach type risks** · **Cross-industry risks**
Outcome	Main points	· The most suited classes of risk correlation for the cyber risk insurability are identified. · Risk-averse firms are inclined to insure risks if intra-firm correlation is high and cross-firm correlation is low.	· The marginal distributions of two variables, the number of infected computer and the dollar loss, are Weibull (non-normal). · Independence assumption in cyber-insurance pricing can make pricing errors substantial.	· Four different nodes about security elements are evaluated and the firewall is the most vulnerable. · By Utility-based preferential pricing (UBPP) model, a risk-averse entity is willing to transfer its cyber risk to an insurer and pay a higher premium.	The effectiveness of CLI contracts on risk mitigation is increasing as intra-firm and upper tail correlations are increasing.	Copula-GARCH model is used to model different cyber attack losses in a high-dimensional setting.	· Information on data breach events has been classified in different cross-sectional settings and analyzed in different dependent structures. · It shows how different high-dimensional dependence constructions influence on cyber-insurance premiums and risk measures.
	Implication & Limitation	The more reliable outcomes and accurate risk measures can be realized by more appropriate and abundant data.	The size of the data is relatively too small to obtain the meaningful result of copula modeling.	This study does not cover different types of attack risks and the curse of dimensionality is implied in the model.	Empirical evidence is not enough to support its proposal of cyber risk index.	The dataset covers cyber incidents, but not real loss data.	A comprehensive analysis on dependence structure of data breach risks is suggested.

Note: The references introduced in the table are specified in the following. BK06: Böhme and Kataria (2006); HH11: Herath and Herath (2011); MCSMS13: Mukhopadhyay et al. (2013); S16: Shah (2016); PXXH18: Peng et al. (2018). The bold indicates the contributing points of the present paper to the literature.

Appendix B. Vine copula models

We consider three types of vine models in the dependence modeling: C-Vine, D-Vine and R-Vine. Among them, the R-Vine is a generalized model that is more flexible than the C-Vine and the D-Vine (Cooke, Joe and Aas, 2011). In what follows, we define three models to describe the difference in the formulation of the structure. Forming the copula distribution function starts with the density factorization. Let us consider a vector of random variables, $X = (X_1, \dots X_d)$, which has a joint density function, $f(x_1, \dots, x_d)$. Then, the density f can be factorized as:

$$f(x_1, \dots, x_d) = f_1(x_1) \cdot f_{2|1}(x_2 | \, x_1) \cdots f_{d|1,\dots,d-1}(x_d | \, x_1, \dots, x_{d-1}). \tag{B.1}$$

Using Sklar's theorem (Sklar, 1959) and the chain rule with continuous marginal function, $F_i(x_i), i = 1, \dots d$, we can obtain

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d) \cdot c_{1,\dots,d}[F_1(x_1), \dots, F_d(x_d)], \tag{B.2}$$

where $c_{1,\dots,d}[\cdot]$ is a d-dimensional copula density.

Specifically, the equivalence between the equation (B.1) and the equation (B.2) can be proven by factorizing the conditional density in the equation (B.1). For example, the conditional density in the bivariate setting, $f_{2|1}(x_2 | \, x_1)$, can be defined as

$$f_{2|1}(x_2 | \, x_1) = \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} = \frac{c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_1(x_1) \cdot f_2(x_2)}{f_1(x_1)} = c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_2(x_2), \tag{B.3}$$

where $c_{1,2}[\cdot,\cdot]$ is a copula density function for the pair of X_1 and X_2, each of which has a density function, $f_i(x_i)$, $i = 1,2$, and a probability function, $F_i(x_i)$.

A three-dimensional case with more than one variable given to the condition in a conditional density can be expressed as an example,

$$f_{1|23}(x_1 | \, x_2, x_3) = c_{1,2|3}\left[F_{1|3}(x_1 | x_3), F_{2|3}(x_2 | x_3)\right] \cdot f_{1|3}(x_1 | x_3). \tag{B.4}$$

Depending on the dependence structure of variables, one can find different factorization of the conditional density from the factorization in the equation (B.4) as follows

$$f_{1|23}(x_1 | \, x_2, x_3) = c_{1,3|2}\left[F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2)\right] \cdot f_{1|2}(x_1 | x_2). \tag{B.5}$$

The equation (B.5) can be further decomposed to the following by factorizing the conditional density function, $f_{1|2}(x_1 | x_2)$,

$$f_{1|23}(x_1 | \, x_2, x_3) = c_{1,3|2}\left[F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2)\right] \cdot c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_1(x_1). \tag{B.6}$$

Based on the development of the decomposition, we can generalize the equation (B.1) by constituting the pairwise copula construction and a conditional marginal density as follows (Aas et al., 2009):

$$f(x|\,\mathbf{\Theta}) = c_{x,\Theta_j|\mathbf{\Theta}_{-j}}\left[F\left(x|\mathbf{\Theta}_{-j}\right), F\left(\Theta_j|\mathbf{\Theta}_{-j}\right)\right] \cdot f\left(x|\mathbf{\Theta}_{-j}\right), \tag{B.7}$$

where $\mathbf{\Theta}$ is a d-dimensional vector, Θ_j is an arbitrarily selected component of the vector $\mathbf{\Theta}$ and $\mathbf{\Theta}_{-j}$ is a vector of $\mathbf{\Theta}$ without the j-th component.

Vine models consist of a number of trees, T_j, $j = 1, \ldots, d-1$, as the decomposition is developed, the trees where starts from unconditional marginal densities. Each tree, T_j, incorporates $d-j$ nodes and $d-j-1$ edges and in the end the entire decomposition of the vine density is defined by $n(n-1)/2$ edges and the marginal densities of d variables. The labels of the edges in the tree T_{j+1} are defined by the nodes in the tree T_j. Specifically, the case that two edges in the tree T_j have a common node forms an edge in the tree T_{j+1} as can be seen in Figure 1.

As abovementioned, R-Vine flexibly links the variables by dependency without fixing a certain structure, providing a general form of a vine model. The following definition illustrates a regular vine model based on Bedford and Cooke (2001, 2002), Cooke, Joe and Aas (2011) and Dissmann et al. (2013).

Definition B1. (A regular vine) $\mathbf{X} = \{x_1, \ldots, x_d\}$ is a d-dimensional set. $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$ is a nested set of trees in a regular vine structure on d components if

(1) T_1 is the first tree with nodes $D_1 = \{1, \ldots, d\}$ and a set of edges, E_1.
(2) For $i = 2, \ldots, d-1$, T_i is a following tree with nodes $D_i = E_{i-1}$ and a set of edges, E_i.
(3) (Proximity condition) for $i = 2, \ldots, d-1$ and $\{a, b\} \in E_i$, $\#(a \cap b) = 1$ holds, where # indicates the cardinality of a set.

The definitions in the following describe how to formulate the vine densities of the C-Vine and the D-Vine, which are special forms of a regular vine model (Aas et al., 2009).

Definition B2. (Canonical vine density) $\mathbf{X} = \{x_1, \ldots, x_d\}$ is a d-dimensional set. In a canonical vine model (C-Vine) in the d-dimensional setting, it consists of T_j, $j = 1, \ldots, d-1$, trees, each of which incorporates $d-j$ nodes and $d-j-1$ edges. Then, the joint density of d variables can be described as:

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1}\left(F\left(x_j|x_1, \ldots, x_{j-1}\right), F\left(x_{j+i}|x_1, \ldots, x_{j-1}\right)\right), \tag{B.8}$$

where index j stands for the j-th tree and index i represents the i-th edge in each tree.[43]

The C-Vine consists of the trees T_j, each of which has a unique node linked with $d-j$ edges, implying that a core variable placed in the center connects all other variables. In contrast, the D-Vine is a structure showing the dependency in a row and forming a hierarchical tree.

Definition B3. (Drawable vine density) $\mathbf{X} = \{x_1, \dots, x_d\}$ is a d-dimensional set. In a drawable vine (D-Vine) model in the d-dimensional setting, it consists of $T_j, j = 1, \dots, d-1$, trees, each of which incorporates $d-j$ nodes and $d-j-1$ edges. Then, the joint density of d variables can be described as:

$$f(x_1, \dots, x_d) = \prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}\left(F\left(x_i \middle| x_{i+1}, \dots, x_{i+j-1}\right), F\left(x_{i+j} \middle| x_{i+1}, \dots, x_{i+j-1}\right)\right), \quad \text{(B.9)}$$

where index j stands for the j-th tree and index i represents the i-th edge in each tree.[44]

Aas et al. (2009) provide more detail in specific algorithms of the D-Vine and the C-Vine and Bedford and Cooke (2001, 2002), Cooke, Joe and Aas (2011) and Dissmann et al. (2013, Section 2) for more detail in representation and specification of R-Vine and the statistical inference on the vine model.

[43] As an example, the 3-dimensional density function of the C-Vine can be expressed as:

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) \cdot c_{13}\big(F_1(x_1), F_3(x_3)\big) \cdot c_{23|1}\big(F(x_2|x_1) \cdot F(x_3|x_1)\big).$$

[44] As an example, the 3-dimensional density function of the D-Vine can be expressed as:

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) \cdot c_{23}\big(F_2(x_2), F_3(x_3)\big) \cdot c_{13|2}\big(F(x_1|x_2) \cdot F(x_3|x_2)\big).$$

Appendix C. Additional description of the method and risk factors

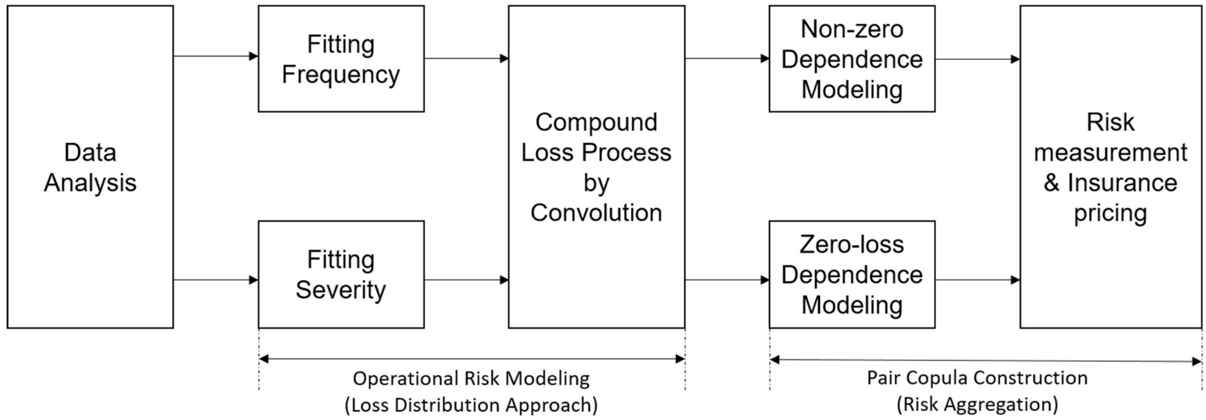
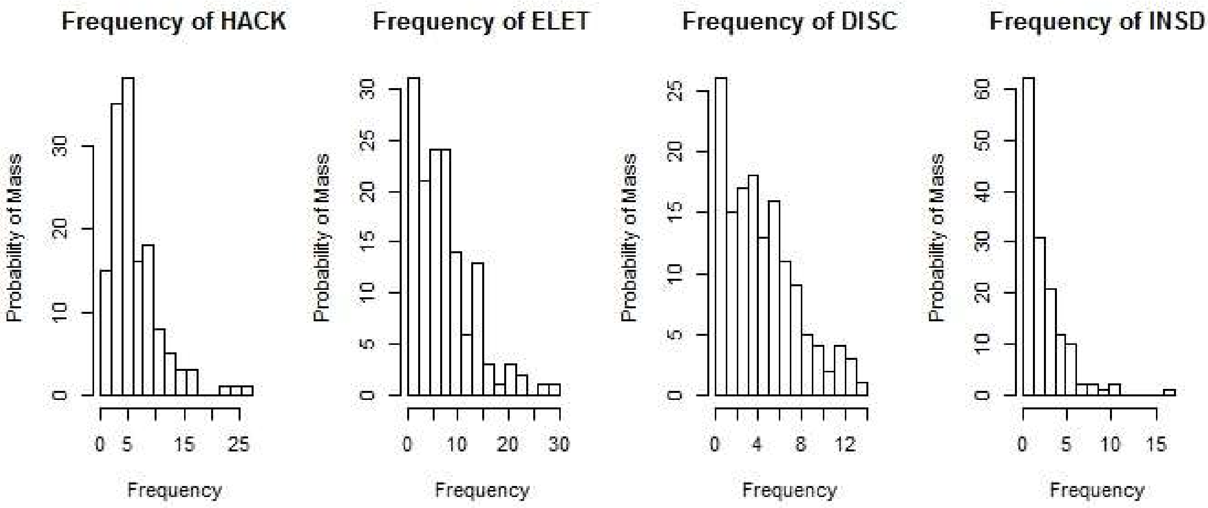


Figure C1. Graphical structure of methodology

Table C1 Augmented Dickey-Fuller Test Results

	Variable	Statistics	P-value
Cross-industry	HACK	-4.241	0.010
	ELET	-3.782	0.022
	DISC	-4.015	0.011
	INSD	-3.971	0.013
Cross-breach type	BSF	-3.095	0.098
	GOV	-3.861	0.018
	MED	-3.710	0.026
	BSE	-4.199	0.010
	EDU	-3.756	0.023

Panel A: Histograms of loss frequency data



Panel B: Histograms of loss severity data (original vs. log-scaled)

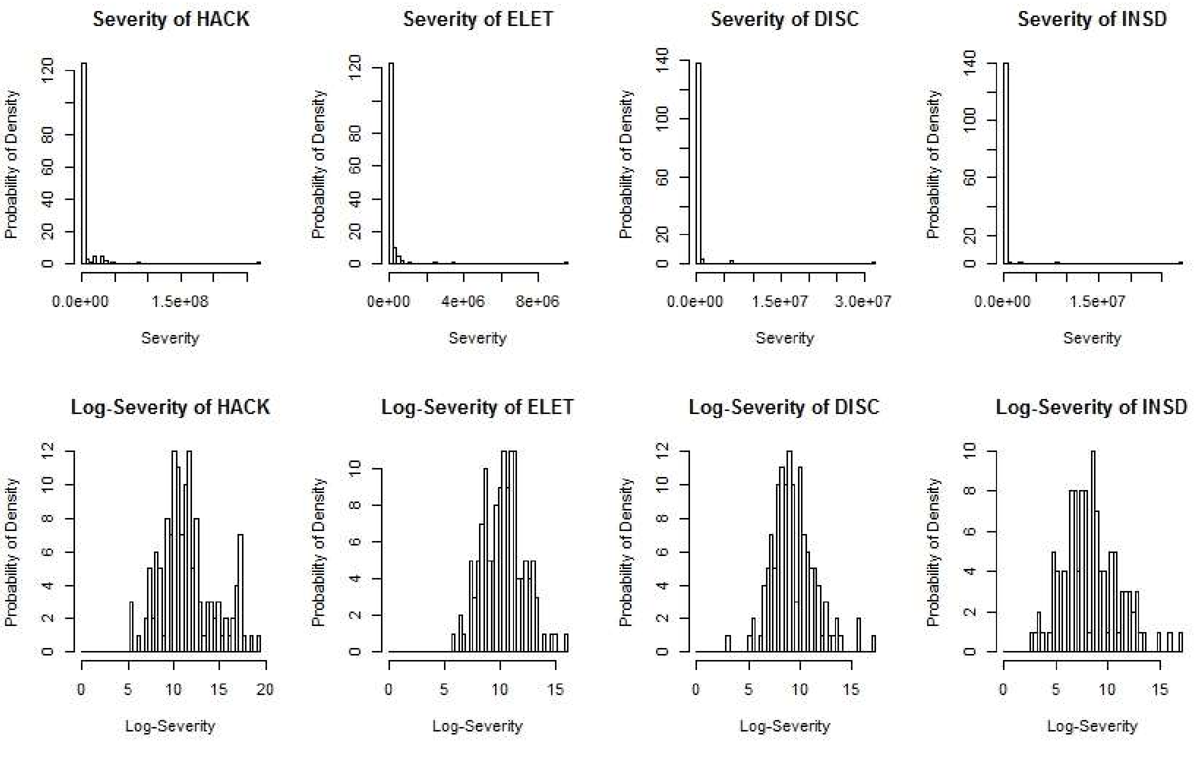
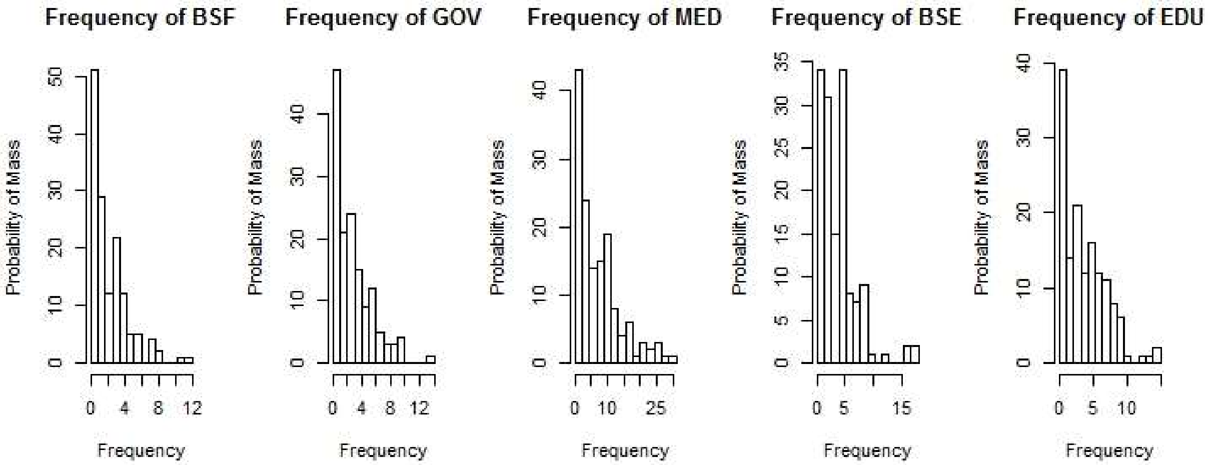


Figure C2. Graphical description on data (cross-industry)

Panel A: Histograms of loss frequency data



Panel B: Histograms of loss severity data (original vs. log-scaled)

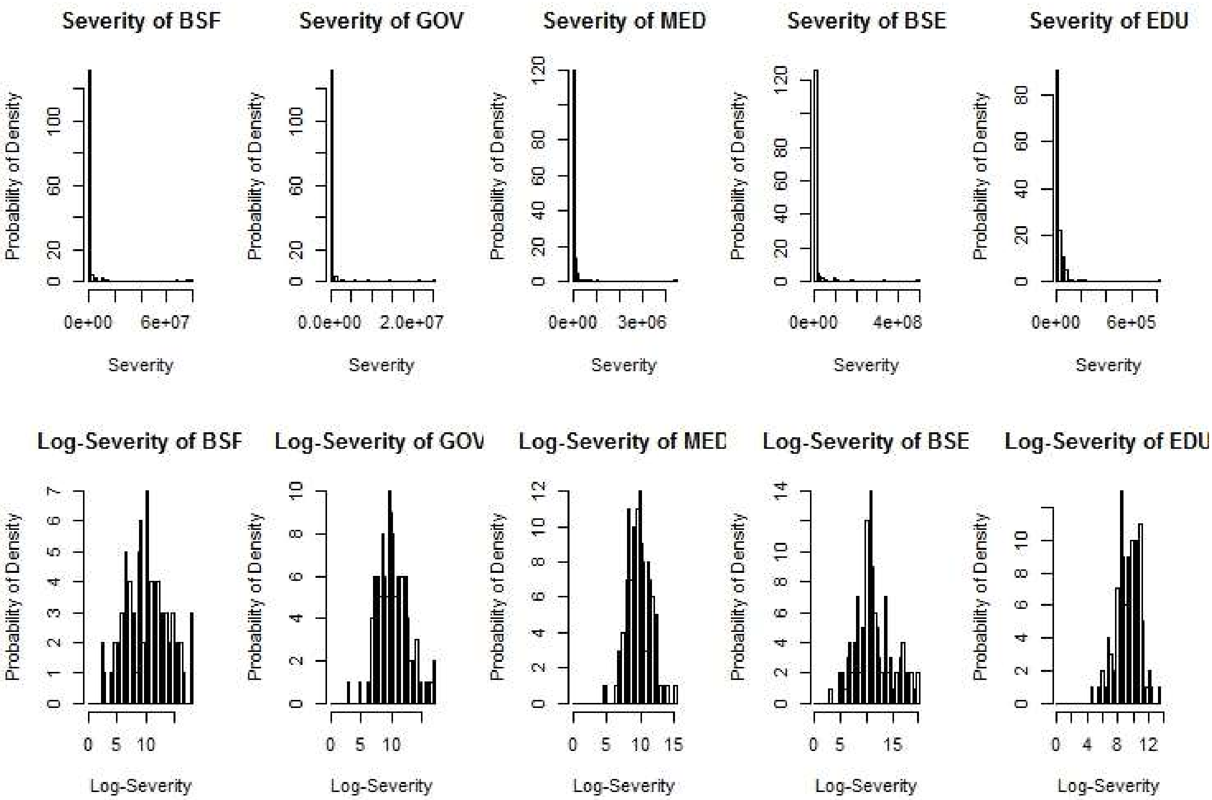
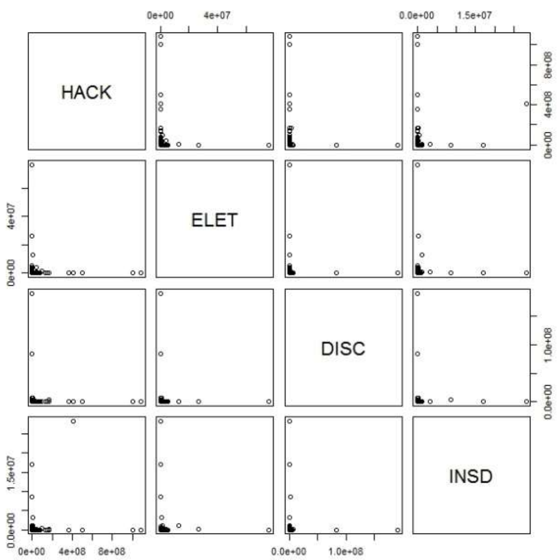


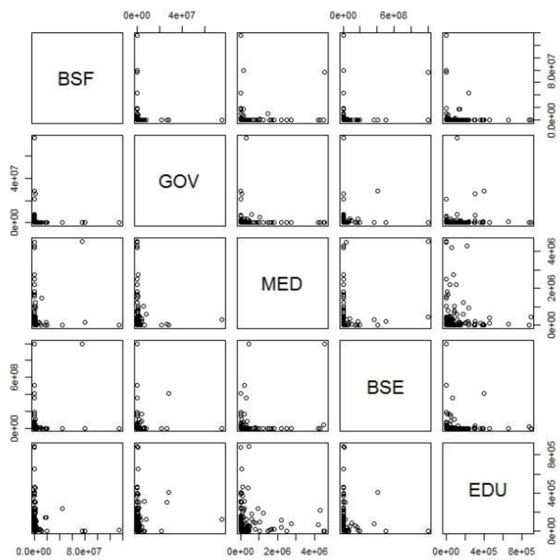
Figure C3. Graphical description on data (cross-breach type)

Panel A: Scatterplot of original monthly data

Cross-industry

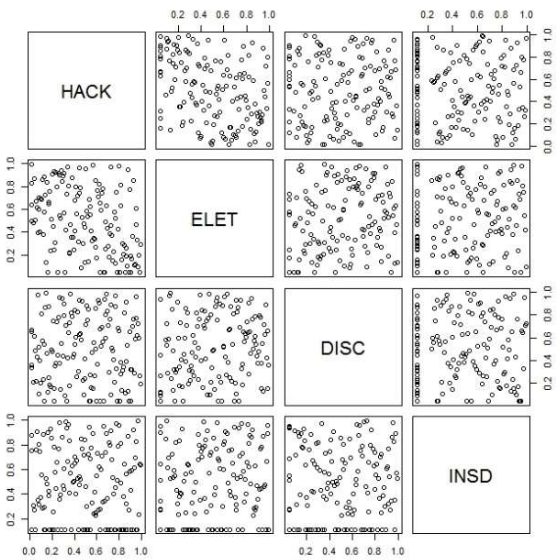


Cross-breach type



Panel B: Scatterplot of transformed uniform margins

Cross-industry



Cross-breach type

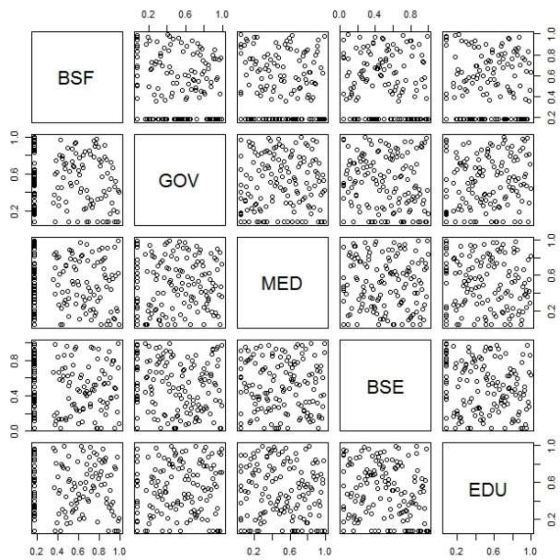


Figure C4. Pairwise plots in both cross-sectional settings

Appendix D. Dependence in zero-loss arrivals

Suppose that there are d-dimensional losses including excess zeros. We model the dependence between occurrences of zero loss events, following Erhardt and Czado (2012) and Brechmann et al. (2014). The k-th loss is denoted by $\lambda_k \geq 0,\ k = 1, \dots, d$. The occurrence of zero loss can be defined by a binary random variable, I_k given as:

$$v_k := \begin{cases} 1 & zero\ loss \\ 0 & otherwise \end{cases}. \tag{D.1}$$

For k-th loss, λ_k is either 1 or 0 as a binary component when it has zero or a positive number respectively. As Erhardt and Czado (2012) and Brechmann et al. (2014) indicate, we also denote a positive loss by $\lambda_k^+ > 0$ distinguished from zero loss, $\lambda_k^0 = 0$. We recall the equation (7) for k-th loss distribution with the binary distribution and the separated loss values as (Brechmann et al., 2014):

$$\lambda_k = v_k \times \lambda_k^0 + (1 - v_k) \times \lambda_k^+ = (1 - v_k) \times \lambda_k^+. \tag{D.2}$$

The cumulative density function of v_k, P_{v_k}, is composed by:

$$P_{v_k} := \begin{cases} p_{v_k}(0) & Zero\ loss \\ 1 & otherwise \end{cases}, \tag{D.3}$$

where p_{v_k} is a probability mass function of v_k and $p_{v_k}(0)$ is the probability of positive loss.

Based on the above expressions, we can construct the joint density of multivariate binary distribution and multivariate loss distribution, where we apply the mathematical descriptions from Erhardt and Czado (2012) and Brechmann et al. (2014). The d-dimensional joint density consists of $\Lambda := (\lambda_1, \dots, \lambda_d)'$, a vector of loss values, $\mathbf{V} := (v_1, \dots, v_d)'$, a binary vector. The joint density can be expressed:

$$l_{\mathbf{V},\Lambda}(v, \boldsymbol{\lambda}) = p_{\mathbf{V}}(\boldsymbol{v}) l_{\Lambda|\mathbf{V}}(\boldsymbol{\lambda}|\boldsymbol{v}), \tag{D.4}$$

where $\boldsymbol{V} := (v_1, \dots, v_d)' \in \{0,1\}^d$ and $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_d)' \in \mathcal{R}_{\geq 0}^d$.

The conditional probability density function, $l_{\Lambda|\mathbf{V}}(\boldsymbol{\lambda}|\boldsymbol{v})$, can be specified by:

$$l_{\Lambda|\mathbf{V}}(\boldsymbol{\lambda}|\boldsymbol{v}) = l_{\lambda_1^+, \dots, \lambda_k^+ \mid v_k = 0,\ k \in \{1, \dots, d\}}(\lambda_k, k \in \{1, \dots, d\}). \tag{D.5}$$

The joint probability mass function of binary variables, $p_{\mathbf{V}}(\boldsymbol{v})$, can be formed by copula functions. However, in the high-dimensional case, there might exist heterogeneous pairwise correlation between binary variables, which cannot be captured by Archimedean copulas with a single dependence parameter (Brechmann et al., 2014). Besides, no closed form has been developed for such a dependence by vine copula method so that we use Gaussian copula to build the joint binary distribution as modeled in Section 4.3.

Basically, a joint probability mass function can be generated as follows:

$$
\begin{aligned}
p_{\mathbf{V}}(\boldsymbol{v}) &= P(V_1 = v_1, V_2 = v_2 \dots, V_d = v_d) \\
&= P(V_1 \le v_1, V_2 \le v_2, \dots, V_d \le v_d) - P(V_1 \le v_1 - 1, V_2 \le v_2, \dots, V_d \le v_d) \\
&\quad - \cdots \\
&\quad -P(V_1 \le v_1 - 1, V_2 \le v_2 - 1, \dots, V_{d-1} \le v_{d-1} - 1, V_d \le v_d) \\
&\quad +P(V_1 \le v_1 - 1, V_2 \le v_2 - 1, \dots, V_{d-1} \le v_{d-1} - 1, V_d \le v_d - 1).
\end{aligned}
$$

2^d-components are required to generate the joint function. For each multivariate probability function, we can use a copula function as represented by Sklar's theorem (Sklar 1959):

$$
\begin{aligned}
p_{\mathbf{V}}(\boldsymbol{v}) &= C_{\mathbf{V}}\left(P_{V_1}(v_1), P_{V_2}(v_2), \dots, P_{V_d}(v_d)\right) - C_{\mathbf{V}}\left(P_{V_1}(v_1 - 1), P_{V_2}(v_2), \dots, P_{V_d}(v_d)\right) - \cdots \\
&\quad -C_{\mathbf{V}}\left(P_{V_1}(v_1 - 1), P_{V_2}(v_2 - 1), \dots, P_{V_{d-1}}(v_{d-1} - 1), P_{V_d}(v_d)\right) \\
&\quad +C_{\mathbf{V}}\left(P_{V_1}(v_1 - 1), P_{V_2}(v_2 - 1), \dots, P_{V_{d-1}}(v_{d-1} - 1), P_{V_d}(v_d - 1)\right) \\
&= C_{\mathbf{V}}\left(u_1^{\{1\}}, u_2^{\{1\}}, \dots, u_d^{\{1\}}\right) - C_{\mathbf{V}}\left(u_1^{\{2\}}, u_2^{\{1\}}, \dots, u_d^{\{1\}}\right) - \cdots \\
&\quad -C_{\mathbf{V}}\left(u_1^{\{2\}}, u_2^{\{2\}}, \dots, u_{d-1}^{\{2\}}, u_d^{\{1\}}\right) + C_{\mathbf{V}}\left(u_1^{\{2\}}, u_2^{\{2\}}, \dots, u_{d-1}^{\{2\}}, u_d^{\{2\}}\right),
\end{aligned}
$$

where $C_{\mathbf{V}}$ is a copula function, $u_k^{\{1\}}=P_{V_k}(v_k)$ and $u_k^{\{2\}} = P_{V_k}(v_k - 1)$. When $v_k = 0$, $P_{V_k}(v_k - 1) = P_{V_k}(-1) = 0$.

Appendix E. Distribution fitting results

Table E1. Goodness-of-fit and Model Comparison for Loss Frequency (cross-breach type)

Distribution	Log-likelihood	AIC	Chisq-Test	
Panel A: Banking and Insurance (BSF)				
Poisson	-323.185	648.370	170.542	***
Zero-inflated Poisson	-285.813	575.625	293.466	***
Negative Binomial	-275.371	554.742	10.041	**
Zero-inflated Neg. Binomial	-274.285	554.571	13.283	
Geometric	**-275.379**	**552.758**	**14.605**	
Panel B: Governmental and Military Entities (GOV)				
Poisson	-356.099	714.199	69.186	***
Zero-inflated Poisson	-338.971	681.942	430.380	***
Negative Binomial	**-322.991**	**649.983**	**1.560**	
Zero-inflated Neg. Binomial	-322.706	651.413	11.839	
Geometric	-330.648	663.296	23.148	*
Panel C: Medical Service (MED)				
Poisson	-656.099	1,314.197	3,479.017	***
Zero-inflated Poisson	-608.074	1,220.147	>10,000	***
Negative Binomial	**-434.950**	**873.901**	**5.462**	
Zero-inflated Neg. Binomial	-434.950	875.901	25.882	
Geometric	-436.423	874.846	26.856	
Panel D: Retail/Merchant and Other Business (BSE)				
Poisson	-383.642	769.284	52.480	***
Zero-inflated Poisson	-378.542	761.085	>10,000	***
Negative Binomial	**-339.426**	**682.852**	**5.827**	
Zero-inflated Neg. Binomial	-339.426	684.852	67.508	***
Geometric	-352.259	706.518	49.869	***
Panel E: Educational Institution (EDU)				
Poisson	-395.706	793.413	166.410	***
Zero-inflated Poisson	-364.743	733.485	214.055	***
Negative Binomial	-351.012	706.025	13.769	*
Zero-inflated Neg. Binomial	**-348.082**	**702.164**	**11.490**	*
Geometric	-358.493	718.987	30.455	**

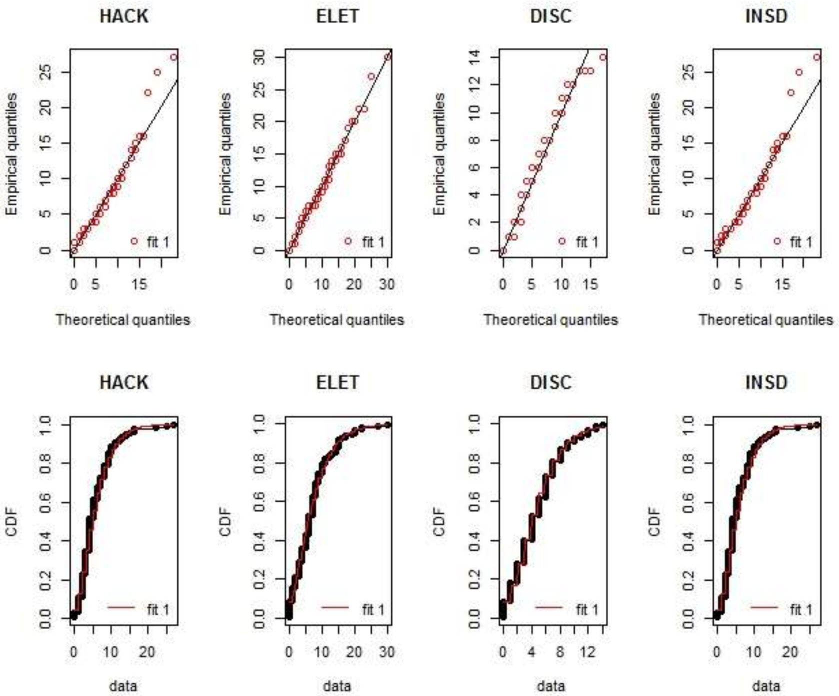
Note: *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit distribution for each loss distribution based on AIC and goodness-of-fit test result.

Table E2. Goodness-of-fit and Model Comparison for Loss Severity (cross-breach type)

Distribution	Log-likelihood	AIC	K-S Test	
Panel A: Banking and Insurance (BSF)				
Lognormal	**-1,186.534**	**2,377.069**	**0.051**	
Skew normal	-1,657.241	3,318.481	0.935	***
Skew t	-1,580.384	3,168.769	0.978	***
Weibull	-1,194.651	2,393.302	0.099	
Gamma	-1,216.283	2,436.567	0.239	***
Inverse Gaussian	-1,271.820	2,547.640	0.489	***
Cauchy	-1,307.130	2,618.260	0.369	***
Burr	-1,187.885	2,383.769	0.061	
GPD	-1,262.033	2,530.065	0.389	***
POT (lognormal-GPD)	-1,186.796	2,381.593	0.129	
Panel B: Governmental and Military Entities (GOV)				
Lognormal	**-1,529.995**	**3,063.990**	**0.060**	
Skew normal	-2,006.018	4,016.036	0.959	***
Skew t	-1,869.513	3,891.964	0.984	***
Weibull	-1,549.831	3,097.662	0.114	
Gamma	-1,578.088	3,160.176	0.248	***
Inverse Gaussian	-1,603.662	3,211.324	0.382	***
Cauchy	-1,616.086	3,236.172	0.309	***
Burr	-1,528.178	3,064.357	0.035	
GPD	-1,672.186	3,350.372	0.470	***
POT (lognormal-GPD)	-1,528.819	3,065.638	0.049	
Panel C: Medical Service (MED)				
Lognormal	**-1,558.599**	**3,121.197**	**0.052**	
Skew normal	-1,902.317	3,808.634	0.947	***
Skew t	-1,646.361	3,300.721	0.985	***
Weibull	-1,574.132	3,152.265	0.116	
Gamma	-1,595.772	3,195.545	0.185	***
Inverse Gaussian	-1,592.513	3,189.025	0.206	***
Cauchy	-1,621.456	3,246.912	0.266	***
Burr	-1,557.608	3,123.217	0.045	
GPD	-1,593.763	3,193.527	0.263	***
POT (lognormal-GPD)	-1,558.205	3,124.411	0.091	
Panel D: Retail/Merchant and Other Business (BSE)				
Lognormal	-1,856.833	3,717.666	0.100	
Skew normal	-2,656.840	5,317.680	0.955	***
Skew t	-2,554.446	5,116.892	0.985	***
Weibull	-1,875.289	3,754.578	0.171	**
Gamma	-1,912.092	3,828.184	0.275	***
Inverse Gaussian	-4,830.702	9,665.404	1.000	***
Cauchy	-2,014.406	4,032.811	0.338	***
Burr	**-1,854.728**	**3,717.456**	**0.049**	
GPD	-1,975.322	3,956.643	0.374	***
POT (lognormal-GPD)	-1,856.936	3,721.872	0.127	
Panel E: Educational Institution (EDU)				
Lognormal	-1,376.489	2,756.979	0.049	
Skew normal	-1,473.857	2,951.714	0.772	***
Skew t	-1,377.429	2,762.858	0.984	***
Weibull	**-1,374.155**	**2,752.309**	**0.068**	
Gamma	-1,375.841	2,755.682	0.089	
Inverse Gaussian	-1,416.137	2,836.274	0.270	***
Cauchy	-1,428.622	2,861.243	0.248	***
Burr	-1,373.585	2,755.170	0.060	
GPD	-1,408.566	2,823.132	0.246	***
POT (lognormal-GPD)	-1,382.941	2,757.360	0.146	

Note: *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit distribution for each loss distribution based on AIC and goodness-of-fit test result.

Panel A: Frequency QQ-plots and CDF plots



Panel B: Severity QQ-plots and CDF plots

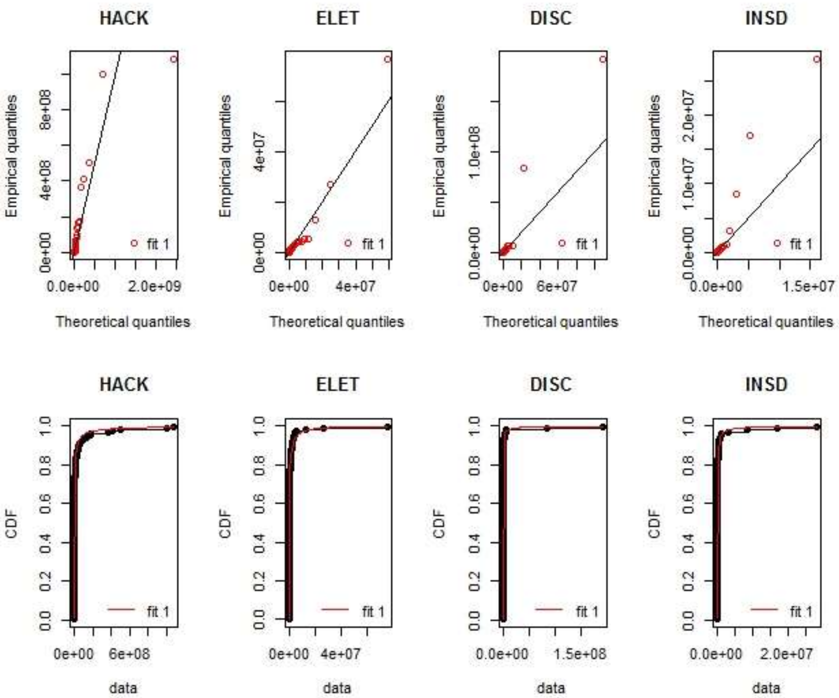
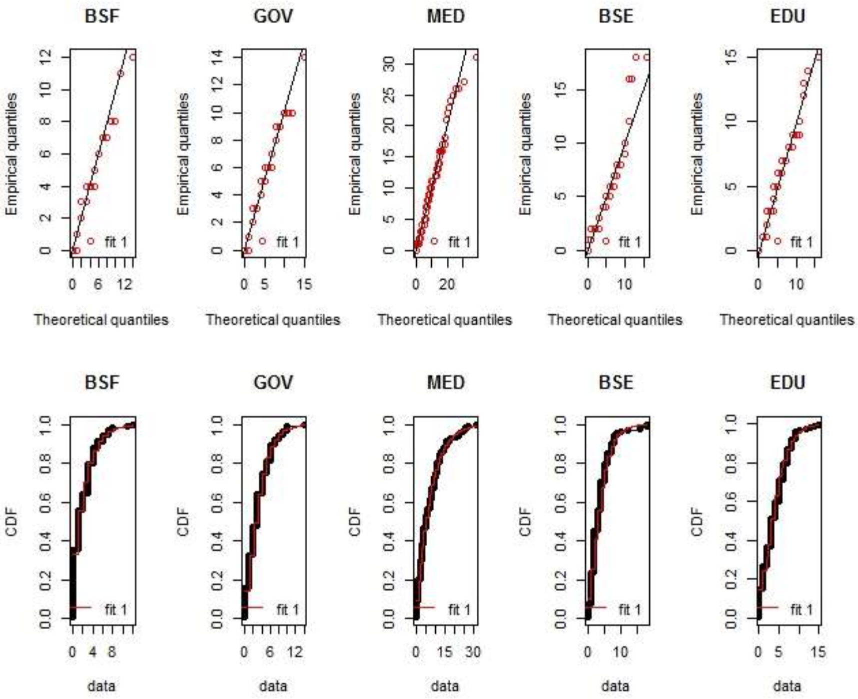


Figure E1. QQ-plots and CDF plots for the best fitted distributions (Cross-industry)

Panel A: Frequency QQ-plots and CDF plots



Panel B: Severity QQ-plots and CDF plots

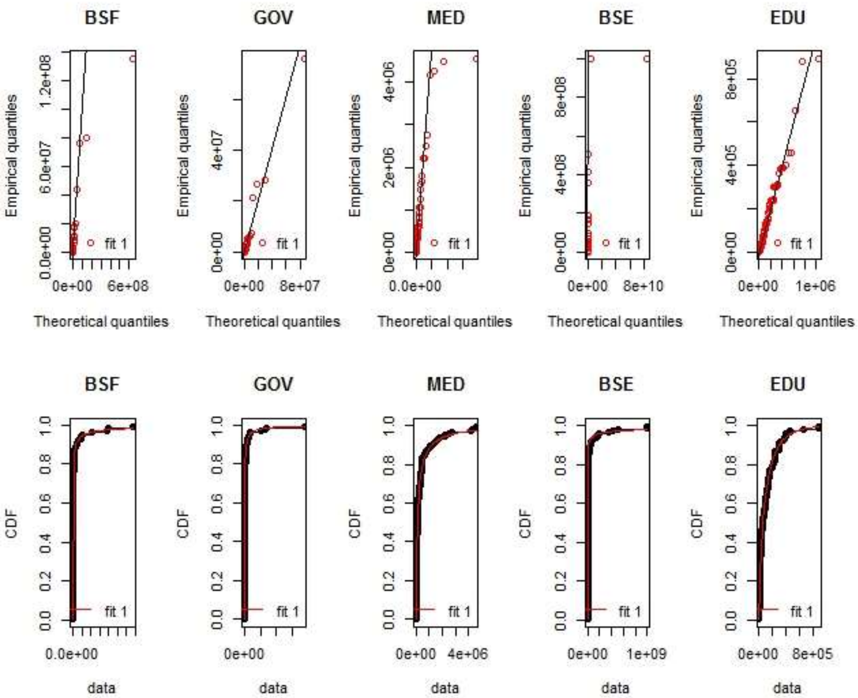


Figure E2. QQ-plots and CDF plots for the best fitted distributions (Cross-breach type)

Appendix F. Graphical description of tree structures and diversification effects

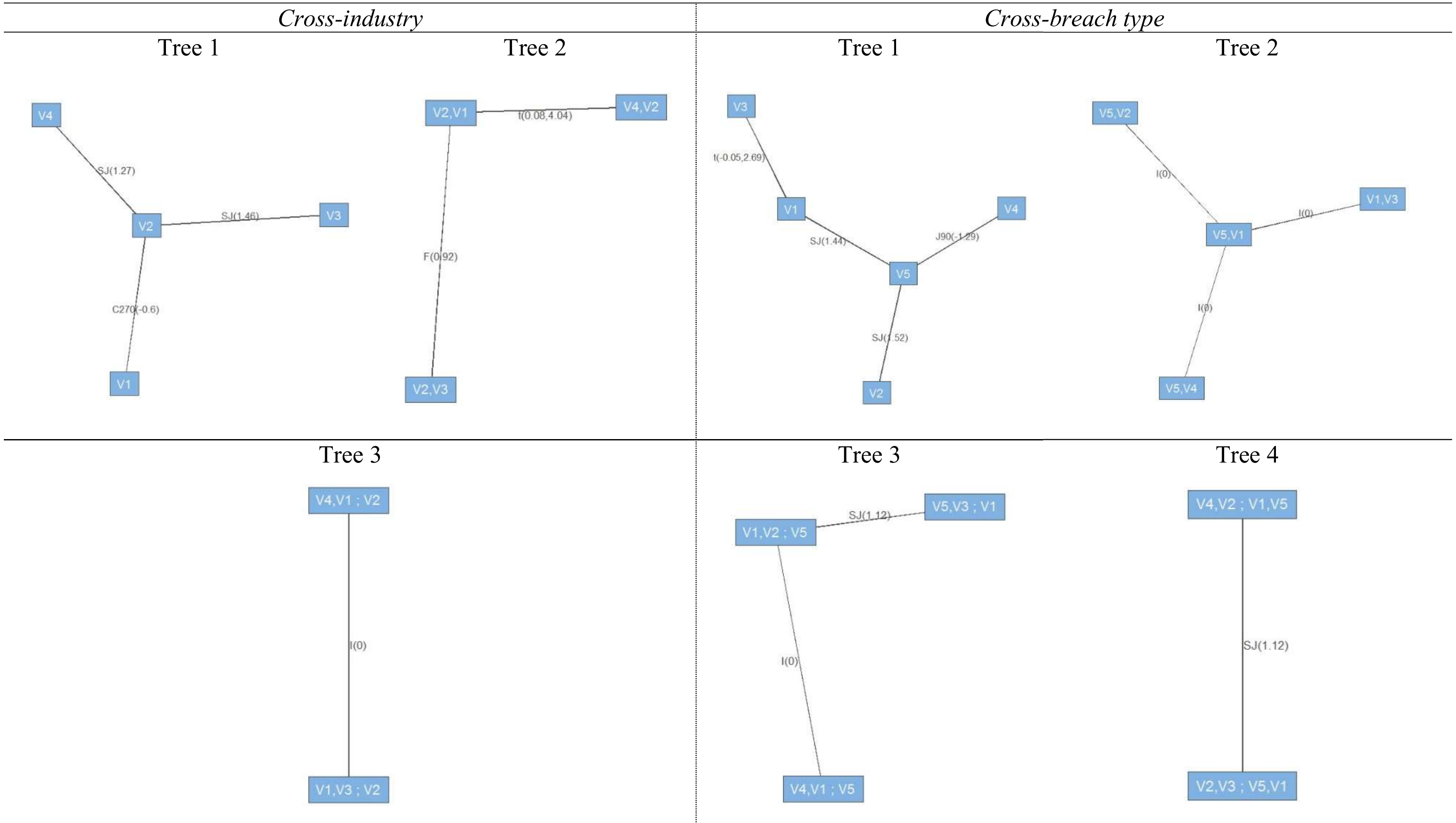


Figure F1. The estimated dependence structure for each cross-sectional setting. The estimated copula functions and their dependence parameters are displayed at edges. The names of copula functions at edges and the variable names in vertices are specified as follows:
SJ: Survival Joe; C270: 270° Rotated Clayton; t: Student-t; F: Frank; I: Independence
Cross-industry: V1 = HACK, V2 = ELET, V3 = DISC, V4 = INSD; *Cross-breach type*: V1 = BSF, V2 = GOV, V3 = MED, V4 = BSE, V5 = EDU.

Cross-industry ***Cross-breach type***

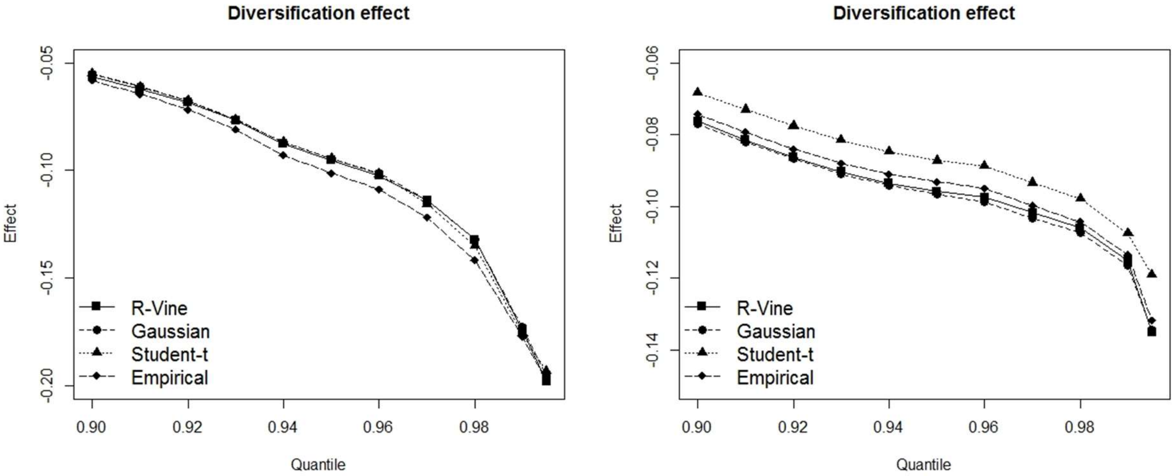


Figure F2. Comparison of diversification effects. The figures indicate the diversification levels over the quantiles between 90% and 99.5%. The diversification level varies from 5% to 20% in the cross-industry setting and from 7% to 14% in the cross-breach type setting. The measures are estimated upon the equation (14) to show how expected shortfalls of different dependence models are reduced from the comonotonicity assumption (simple sum of individual expected shortfalls).

Appendix G. Applications on a company level

The results of risk measurement in panel A of Table G1 indicate the potential breach records per event on a single company level. In other words, a cyber-insurer could be confronted with such estimates from an individual insured in its cyber-insurance portfolio. Besides of the risk measurement, two cases of annual insurance premiums in panel B of Table G1 are considered: one is for an educational institution suffering from hacking risk (company #1) and the other for a hospital exposed to lost electronic device risk (company #2).[45] These cases are designed to be comparable with the examples in Table 9 of Eling and Loperfido (2017).[46] To compute annual insurance premiums (annual loss probability) and calibrate the values on a company level, we construct the following estimator as the annual loss probability (π_L) of a cyber-insured (Hess, 2011):

$$\pi_L = \frac{N_r}{N_y \cdot N_e}, \tag{G.1}$$

where N_r is the number of incidents, N_y is the number of years and N_e is the number of entities. Applying the above equation to the complete data breach dataset leads to a conditional transformation factor, because our dataset only contains companies that actually have been breached. Risk measures and insurance prices in this case might be too high. To overcome this problem, we use the S&P500 index (as a reference for the number of entities) and filter our dataset by looking into how many firms in the index have been breached (to identify the number of incidents). Through the filtering process, we find 0.0165 as the annual loss probability used to price the fair premium with N_r=99, N_y=12 and N_e=500.

[45] To generate the premium estimation for a particular insured exposed to a specific risk type, we define a weighting parameter, w_{ij}, to generate a loss variable for a considered policyholder applied to insurance pricing, the parameter where i is the dimension of cross-industry setting (i=1,…4) and j is of cross-breach type setting (j=1,…,5). The weighting parameter is described in the following:

$$w_{ij} = \frac{\mathrm{E}(x_{ij})}{\sum_{i=1}^{4}\sum_{j=1}^{5}\mathrm{E}(x_{ij})},$$

where x_{ij} is a vector of industry j suffering from risk i. For an educational institution suffering from a hacking risk, the weighting parameter turns out to be 0.0019 and for a hospital exposed to lost electronic device risk it is 0.0065.

[46] Eling and Loperfido (2017) derive the premium value for a university suffering from a hacking risk from univariate distributions under different parametric assumptions; hence, the result does not incorporate any dependence structure. However, our result is derived under the estimated dependence structures, which implicitly contains a correlated risk. The difference in the results between two studies comes from different parameter assumptions on the loss probability, that is, Eling and Loperfido (2017) assume 10% as the loss probability, which is around 6 times larger than the loss probability used in this application.

Table G1. Applications to Risk Measurement and Insurance Pricing

Panel A: Risk measurement *(in million breached records)*

Data Type	Dependence Structure	Value-at-Risk 90%	95%	99%	99.5%	Expected Shortfall 90%	95%	99%	99.5%
Cross-industry	Indep	2.34	7.89	38.24	38.24	14.42	23.61	50.58	52.53
	PCC	**2.98**	**8.16**	**49.23**	**52.09**	**15.70**	**26.14**	**51.66**	**52.66**
	Empirical	2.99	8.15	49.31	52.29	15.51	25.79	51.78	52.76
	Gaussian	2.80	8.07	49.02	52.01	15.08	25.11	51.69	52.79
	Student-t	2.86	8.05	49.19	52.26	15.34	25.63	51.95	53.15
Cross-breach type	Indep	2.27	7.63	36.60	47.30	13.82	22.64	47.87	49.19
	PCC	**2.58**	**7.70**	**46.53**	**46.54**	**13.80**	**23.40**	**46.67**	**46.81**
	Empirical	2.59	7.72	46.53	46.55	13.72	23.17	46.70	46.85
	Gaussian	2.59	7.80	46.53	46.56	14.05	23.76	46.74	46.94
	Student-t	2.54	7.72	46.53	46.60	13.96	23.71	47.11	47.67

Panel B: Insurance pricing *(annual premium in $)*

Case	Data Type	Dependence structure	Fair Premium	Standard Dev. Principle	Exponential Premium Principle $\gamma = 10^{-3}$	$\gamma = 10^{-4}$	$\gamma = 10^{-5}$
Company #1	Cross-industry	Indep	102,214	132,360	102,259	102,219	102,215
		PCC	**107,454**	**138,192**	**107,501**	**107,459**	**107,455**
		Empirical	106,990	137,500	107,036	106,994	106,990
		Gaussian	104,275	134,254	104,320	104,279	104,275
		Student-t	105,711	136,068	105,757	105,716	105,711
	Cross-breach type	Indep	97,085	126,214	97,127	97,089	97,085
		PCC	**97,285**	**125,475**	**97,325**	**97,289**	**97,285**
		Empirical	96,850	124,874	96,889	96,854	96,850
		Gaussian	98,131	126,615	98,172	98,135	98,132
		Student-t	96,736	125,234	96,740	96,740	96,736
Company #2	Cross-industry	Indep	260,798	337,715	261,092	260,827	260,801
		PCC	**274,167**	**352,595**	**274,475**	**274,198**	**274,171**
		Empirical	272,982	350,829	273,285	273,012	272,985
		Gaussian	266,055	342,547	266,348	266,085	266,058
		Student-t	269,720	347,176	270,020	269,750	269,723
	Cross-breach type	Indep	247,710	322,032	247,984	247,737	247,713
		PCC	**248,221**	**320,147**	**248,480**	**248,247**	**248,224**
		Empirical	247,111	318,615	247,367	247,137	247,114
		Gaussian	250,380	323,056	250,645	250,407	250,383
		Student-t	246,820	319,532	247,084	246,846	246,822

Note: The risk measurements are specified by Value-at-Risk (VaR) and Expected Shortfall (ES) at three critical levels: 90%, 95%, 99% and 99.5%. For insurance pricing, we calculate the premium by three different pricing principles on an annual basis: fair premium, standard deviation and exponential premium principles. γ is the risk aversion parameter, where $\gamma \to 0$ indicates risk neutrality. The bold model is the preferred model from Section 4.

References

Aas, K., and Berg, D. (2009). Models for construction of multivariate dependence - A comparison study. *The European Journal of Finance, 15*(7-8), 639-659.

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics, 44*(1), 182-198.

Acerbi, C., and Tasche, D. (2002). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes, 31*(2), 379-388.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*. Budapest, Akademiai Kiado, 267-281.

Allianz. (2015). *A guide to cyber risk*. Munich: Allianz Global Corporate & Specialty Communications.

Araichi, S., Peretti, C., and Belkacem, L. (2016). Solvency capital requirement for a temporal dependent losses in insurance. *Economic Modelling, 58*, 588-598.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203-228.

Bedford, T., and Cooke, R. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence, 32*(1-4), 245-268.

Bedford, T., and Cooke, R. (2002). Vines: A new graphical model for dependent random variables. *The Annals of Statistics, 30*(4), 1031-1068.

Belgorodski, N. (2010). Selecting pair-copula families for regular vines with application to the multivariate analysis of European stock market indices. *Diploma Thesis*. Technische Universität München.

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

Böhme, R., and Kataria, G. (2006). Models and measures for correlation in cyber-insurance. *Workshop on the Economics of Information Security (WEIS)*, University of Cambridge, UK.

Böhme, R., and Schwartz, G. (2010). Modeling cyber-insurance: Towards a unifying framework. *Workshop on the Economics and Insurance Security (WEIS)*, Havard University, US.

Brechmann, E. C., and Schepsmeier, U. (2013). Modeling dependence with C- and D-vine copulas: The R-package CDVine. *Journal of Statistical Software, 52*(3), 1-27.

Brechmann, E. C., Czado, C., and Paterlini, S. (2014). Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *Journal of Banking and Finance, 40*, 271-285.

Bühlmann, H. (2007). *Mathematical methods in risk theory*. Heidelberg: Springer Science & Business Media.

Cebula, J., and Young, L. (2010). *A taxonomy of operational cyber security risks*. Software Engineering Institute, Carnegie Mellon.

Chen, X., and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics, 135*(1-2), 125-154.

Chiou, S., and Tsay, R. (2008). A Copula-based approach to option pricing and risk assessment. *Journal of Data Science, 6*(3), 273-301.

Clarke, K. (2007). A simple distribution-free test for nonnested model selection. *Political Analysis, 15*(3), 347-363.

Conover, W. (1971). *Practical nonparametric statistics*. New York: John Wiley & Sons.

Cooke, R. M., Joe, H., and Aas, K. (2011). Vines arise. In Kurowicka. D., and Joe. H., *Dependence modeling: vine copula handbook* (pp. 37-71). Singapore: World Scientific.

Czado, C., Jeske, S., and Hofmann, M. (2013). Selection strategies for regular vine copulae. *Journal de la Société Française de Statistique, 154*(1), 174-191.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula theory and its applications* (pp. 93-109). Berlin Heidelberg: Springer.

Dissmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis, 59*, 52-69.

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

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

Eling, M., and Wirfs, J. (2016). *Cyber risk: too big to insure? Risk transfer options for a mercurial risk class*. Institute for Insurance Economics, University of St. Gallen.

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

Embrechts, P. (2000). Actuarial versus financial pricing of insurance. *Journal of Risk Finance, 1*(4), 17-26.

Embrechts, P., and Hofert, M. (2013). Statistical inference for copulas in high dimensions: A simulation study. *ASTIN Bulletin, 43*(2), 81-95.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). *Modelling dependence with copulas and applications to risk management. Rapport technique*, Département de mathématiques, Institut Fédéral de Technologie de Zurich.

Embrechts, P., and Puccetti, G. (2008). Aggregating risk across matrix structured loss data: The case of operational risk. *Journal of Operational Risk, 3*(2), 29-44.

Erhardt, V., and Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal, 2012*(2), 106-129.

eRiskHub. (2016). NetDiligence® *Mini data breach cost calculator*. Retrieved from https://eriskhub.com/mini-dbcc.

European Union. (2016). *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC* (General Data Protection Regulation). Regulation (EU) 2016/679 of the European Parliament and of the Council.

FINMA. (2016). *FINMA Circular 2017/3 SST*. Bern: Swiss Financial Market Supervisory Authority (FINMA).

FireEye. (2016). *Calculate your cyber security cost*. Retrieved from https://www.fireeye.com/current-threats/tco/calculator.html.

Fletcher, D., Mackenzie, D., and Villouta, E. (2005). Modelling skewed data with many zeros: A simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics, 12*(1), 45-54.

Frachot, A., Georges, P., and Roncalli, T. (2001). Loss distribution approach for operational risk. *Working paper*. Retrieved from https://ssrn.com/abstract=1032523.

Frees, E. W., and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association, 103*(484), 1457-1469.

Frees, E., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks, 4*(1), 4.

Fu, L., and Moncher, R. (2004). *Severity distributions for GLMs: Gamma or lognormal? Evidence from Monte Carlo simulations*. Casualty Actuarial Society Discussion Paper Program, 149-230.

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika, 82*(3), 543-552.

Genest, C., Remillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Ecnomics, 44*(2), 199-213.

Genest, C., and Rivest, L. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association, 88*(423), 1034-1043.

Giacometti, R., Rachev, ST., Chernobai, A., and Bertocchi, M. (2008). Aggregation issues in operational risk. *Journal of Operational Risk, 3*(3), 3-23.

Haff, I. H. (2013). Parameter estimation for pair-copula constructions. *Bernoulli, 19*(2), 462-491.

Harder, M. (2016). Exchangeability of copulas. *Doctoral dissertation*, Universität Ulm.

Herath, H., and Herath, T. (2011). Copula-based actuarial model for pricing cyber-insurance policies. *Insurance markets and companies: Analyses and Actuarial Computations, 2*(1), 7-20.

Hess, C. (2011). The impact of the financial crisis on operational risk in the financial services industry: Empirical evidence. *Journal of Operational Risk, 6*(1), 23-35.

Hogg, R., McKean, J., and Craig, A. (2005). *Introduction to mathematical statistics*. New Jersey: Pearson Education International.

Imperva. (2016). *DDoS downtime cost calculator*. Retrieved from https://lp.incapsula.com/ddos-downtime-cost-calculator.html.

Jacobs, J. (2014). *Analyzing Ponemon cost of data breach*. Retrieved from http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m − 1)/2 bivariate dependence parameters. *Distributions with Fixed Marginals and Related Topics, 28*, 120-141.

Joe, H., and Xu, J. (1996). *The estimation method of inference functions for margins of multivariate models*. Technical Report 166, Department of Statistics, University of British Columbia.

Joe, H. (1997). *Multivariate models and dependence concepts*. New York: Chapman & Hall.

Jorion, P. (2007). *Value-at-Risk: The new benchmark for managing financial risk*. New York: McGraw-Hill.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern actuarial risk theory*. Heidelberg: Springer-Verlag.

Kleiber, C., and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*, Hoboken, NJ: John Wiley & Sons.

Klugman, S., Panjer, H., and Wilmot, G. (2012). *Loss models: from data to decisions*, 3rd ed. New Jersey: John Wiley & Sons.

KPMG. (2016). *Cyber Insurance: Are insurers finding growth or looking for trouble?*.

Kurowicka, D., and Cooke, R. (2004). *Distribution - free continuous Bayesian belief nets*. In Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice.

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

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. New Jersey: Princeton University Press.

Moscadelli, M. (2004). *The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee*. Retrieved from https://ssrn.com/abstract=557214.

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., and Sadhukhan, S. (2013). Cyber-risk decision models: To insure IT or not? *Decision Support Systems, 56*(1), 11-26.

National Conference of State Legislative (NCSL). (2019). *Security breach notifications laws*. Retrieved from http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx.

Ogut, H., Raghunathan, S., Menon, N. (2011). Cyber security risk management: Public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection. *Risk Analysis, 31*(3), 497-512.

Panjer, H. (2006). *Operational risk: Modeling analytics*. New Jersey: John Wiley & Sons.

Peng, C., Xu, M., Xu, S., and Hu, T. (2018). Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, *45*(15), 1-23.

Ponemon Institute. (2016). *Cost of data breach study*. Traverse City, MI: Ponemon Institute.

Privacy Rights Clearinghouse (PRC). (2016). *Data breaches*. Retrieved from https://www.privacyrights.org/data-breaches.

Rasoulian, S., Gregorie, Y., Legoux, R., and Senecal, S. (2017). Service crisis recovery and firm performance: Insights from information breach announcements. *Journal of the Academy of Marketing Science, 45*(6), 1-18.

Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. (2017). Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?. *Working paper*.

Rosenberg, J., and Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics, 79*(3), 569-614.

Savu, C., and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance, 10*(3), 295-304.

Scarrott, C., and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal, 10*(1), 33-60.

Scheffer, M., and Weiss, G. N. (2017). Smooth nonparametric Bernstein vine copulas. *Quantitative Finance, 17*(1), 139-156.

Shah, A. (2016). *Pricing and risk mitigation analysis of a cyber liability insurance using Gaussian, t and Gumbel copulas - A case for cyber risk index*. In Canadian Economic Association (CEA) Ottawa meetings paper.

Shevchenko, P. V. (2010). Implementing loss distribution approach for operational risk. *Applied Stochastic Models in Business and Industry, 26*, 277-307.

Shevchenko, P. V. (2011). *Modelling operational risk using Bayesian inference*. Heidelberg: Springer.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Inst. Statist. Univ Paris 8*, 229-231.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics, 19*(2), 279-281.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307-333.

Wang, S. (1998). Aggregation of correlated risk portfolios: models and algorithms. *Proceedings of the Casualty Actuarial Society, 85*(163), 848-939.

Wheatley, S., Maillart, T., and Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B, 89*(1), 7.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.

World Economic Forum. (2016). *The global risks report 11th edition*. Geneva: World Economic Forum.

Xu, M., and Hua, L. (2017). *Cybersecurity insurance: modeling and pricing*. Schaumburg, IL: Society of Actuaries (SoA).

Zuur, A., Leno, E., Walker, N., Saveliev, A., and Smith, G. (2009). Zero-truncated and zero-inflated models for count data. In *Mixed effects models and extensions in ecology with R* (pp. 261-293). New York: Springer.

Essay II

Probable maximum cyber loss: Empirical estimation and reinsurance design with public intervention

Abstract

This study defines the probable maximum cyber loss (PMCL), which stands for the worst cyber loss likely to occur, with an alternative approach to estimating the potential size of a next worst cyber risk event. We determine that the series of cyber loss maxima on weekly, biweekly and monthly bases are stationary, that short-range temporal dependency exists and that the Fréchet type of generalized extreme value distribution (GEV) well describes fat-tailedness of cyber loss maxima, and we observed extreme dependency. We find that the predicted cyber loss likely to occur in the next five years is almost seven times larger than the estimate presented by the recent literature based on a widely used Pareto model. In particular, the comparison between the estimates from the data for the more recent period (after 2014) and those for the older period (before 2014) shows a significant increase with a structural break. Applying PMCL estimates, we further provide an empirical benchmark of premiums under reinsurance design with public intervention. Our findings are important for risk managers, actuaries and policymakers concerned about the enormous cost of the next extreme cyber event.

Keywords: Cyber risk, Block maxima, Generalized extreme value distribution, Extreme value copula, Cyber-reinsurance, Public intervention

---

Kwangmin Jung (2019).

This paper was presented at the 2019 annual meeting of American Risk and Insurance Association (ARIA) in San Francisco, the 2019 S. S. Huebner Doctoral Colloquium in San Francisco, the 2019 annual meeting of the Asia-Pacific Risk and Insurance Association (APRIA) in Seoul, the 2019 German Insurance Congress (DVfVW) in Berlin, the Ph.D. in Finance Seminar at the University of St. Gallen and the 8th research seminar at the Institute of Insurance Economics at the University of St. Gallen.

1 Introduction

In the interconnected network environment, society has constantly been exposed to large-scale cyber risk over the last decade. As seen in Figure 1, large-scale cyber losses are already considered in the risk category of high likelihood and high severity (World Economic Forum, 2016). Large-scale extreme losses mostly result from extreme scenarios potentially triggering a significant systemic loss in an industry beyond the classical estimate of extreme risk (Lloyd's, 2017). For instance, the loss amount stemming from the "WannaCry" attack in 2017 is expected to reach nearly $4 billion around the globe (Berr, 2017), and Lloyd's (2017) estimates the potential economic damage by extreme cyber risk to be more than $120 billion.[47] An extreme cyber loss event could cost society more than a major natural catastrophe would, such as super storm Sandy in 2012 (Ralph, 2017).

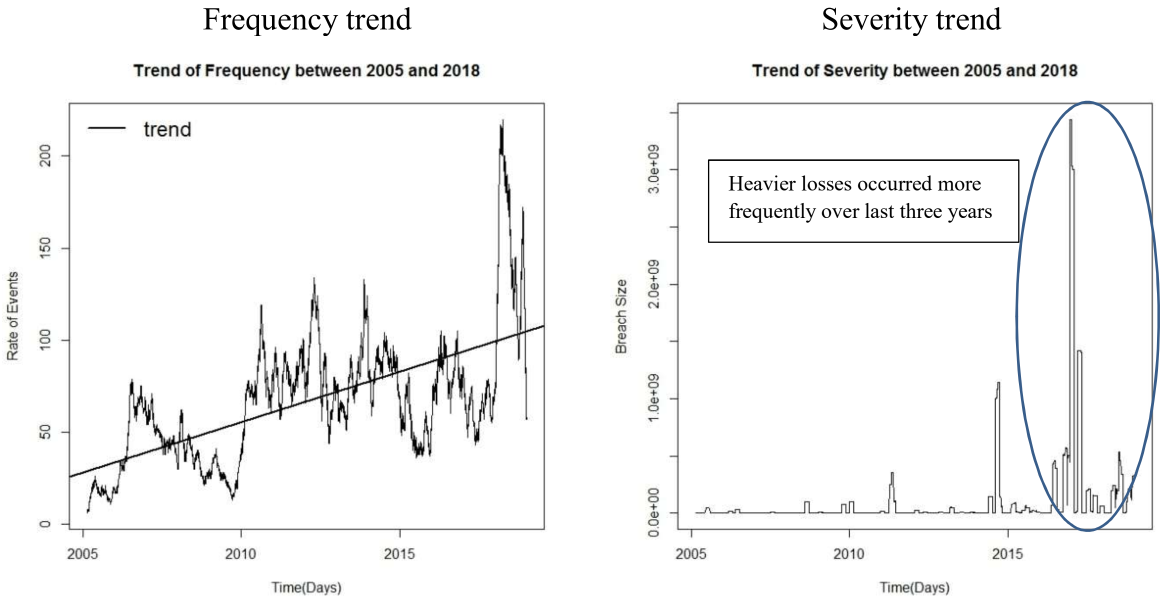


Figure 1. Trends in frequency and severity of cyber risk between 2005 and 2018 from Privacy Rights Clearinghouse (PRC). Rolling windows for 50 days are applied to both data sets to see the cluster.

This extreme cyber risk can also trigger a huge accumulation risk to insurance companies. Extreme cyber risk becomes more catastrophic, complex and rapidly evolving; hence, one of the challenges for insurers and risk managers is to estimate the size of the next extreme loss simultaneously hitting many policyholders in other property and casualty pools, called silent cyber risk (Orcutt, 2017). However, most cyber-insurance policies in the market are conservatively underwritten and lack a standard approach to estimating the coverage limit and the premium so that the coverage limit of the policies is usually set at a low level (Romanosky

47 Extreme scenarios possibly triggering a huge economic cost could include, for example, a cyberattack on critical infrastructure (Paté-Cornell et al., 2018) and cyber hacking to disrupt a global IT supply chain (Zheng and Albert, 2019).

et al., 2017), keeping potential insureds (organizations) from purchasing cyber-insurance. Thus, there is a need to provide empirical evidence on how much the coverage limit of cyber-insurance could be set.

A number of studies on empirical processes have already sought to identify features of extreme cyber losses (Maillart and Sornette, 2010; Edwards, Hofmeyr and Forrest, 2016; Wheatley, Maillart and Sornette, 2016; Eling and Jung, 2018; Eling and Wirfs, 2019; Hofmann, Wheatley and Sornette, 2019).[48] Such studies have commonly found that cyber risk is heavy-tailed and needs to be modeled using extreme value techniques. However, all of them have mainly focused on the loss distribution of cyber risk across the entire spectrum of the loss severity, in which predominantly small losses are observed while only a few are extremely large.[49] From the insurer's perspective, small losses can be considered uninsurable as they are generally classified into the risk category of high frequency and low severity (Eling and Jung, 2018).

The extreme value technique mostly employed in the literature and in the industry for the entire spectrum is a threshold-based approach that splits a random variable into two parts: the main body and the tail above a threshold. For instance, Wheatley et al. (2016) conclude that the tail behavior of the loss severity is extremely heavy so that the tail distribution beyond a threshold needs to be evaluated by Pareto distribution (power-law), estimating the expected maximum breach size to be around 200 million records based on the data from 2007 to 2015. Yet nine events triggering more than 200 million breaches have already occurred over last five years (see Table B2 in Appendix B), accounting for losses that Wheatley et al. (2016) call are "dragon kings."[50] If insurers estimate the required capital using what the classical method suggests, as is done in most literature, they would fail to manage a cyber dragon king and would face an increasing possibility of insolvency. Thus, the main interest of cyber-(re)insurers would be to

[48] A detailed description of the literature is presented in Table A1 in Appendix A to compare the findings with the present paper.

[49] Likewise, in financial risk management, the entire spectrum of financial returns cannot help one accurately capture the maximum likely loss of a financial institution under stress market conditions (e.g., the 2008 financial crisis). Bali (2007), thus, suggests an approach to estimating the maximum likely loss based on the distribution of extreme financial returns with generalized extreme value method, which the author argues can provide a better prediction of catastrophic market risk.

[50] Wheatley et al. (2016) are concerned about the possibility of dragon kings that are beyond their estimated distribution function. As mentioned, their concern has already become the reality, which can increase the demand for a different approach to constructing a predictive model for the maximum loss. An alternative approach can be utilized to help insurers (or reinsurers) set the coverage limit for cyber-insurance. In addition, a dragon king event is different from a black swan event in that the latter is regarded as an "unknown unknown" (Aven, 2013), whereas the former is a "known unknown" so that it can be predictable to some extent. In this sense, we aim to construct an empirical estimation process for cyber dragon king losses.

determine how much probable maximum loss (PML).[51] that cyber risk as a dragon king could generate, which can significantly influence the estimation of capital requirement against catastrophe cyber risk.

In this regard, we suggest an alternative approach to a more accurate, widening prediction for the coverage limit by focusing on maximum losses in certain blocks (time blocks) and investigating their statistical property.[52] Thus, the aim of this paper is to provide a holistic framework in a clearly distinct manner from the existing literature to determine how big the next cyber loss would be. We apply time-series and extreme value analyses to the loss maxima for certain time blocks (weekly, biweekly, and monthly) to ascertain whether time-varying and autoregressive features exist in the cyber loss maxima series and what distributional properties the series incorporate.[53] This procedure is also applied to a bivariate setting, where two variables are defined from the database: malicious risk (hacking, insider breach and payment card fraud risk source) and negligent risk (lost portable device, unintended disclosure, physical loss and lost stationary device). We apply extreme value copulas (Gumbel-Hougaard, Galambos, Husler-reiss, and Tawn) to aggregate the bivariate risk factors and estimate a potential maximum size of the correlated risk that a cyber-insurer is likely to be confronted in a cyber risk pool.

All steps of the above process are conducted within three data periods: the entire (between 2005 and 2018), the pre-2014 and the post-2014 periods. The rationale of the time separation is based on our finding of a structural break between pre-2014 and post-2014, as observed in the right panel of Figure 1. This analysis is material for figuring out the recent trend of a possible worst cyber loss and how different probable maximum cyber losses between two periods are in the fast-changing digitalized society (risk of change). Based on the statistical estimation on loss maxima, the probable maximum cyber losses are predicted for the next one, three, and five years and applied to an empirical design of reinsurance portfolio with public intervention, where

[51] Probable maximum loss (PML) is a term that has been used in the insurance field for decades, indicating the possible largest loss that could be caused by a catastrophe (Wilkinson, 1982). This concept has usually been applied to property insurance under the assumption that self-protective measures functioning against possible fire losses exist.

[52] Modeling the upper limit of the cyber loss is crucial, particularly for the current cyber-insurance market, in that low demand for cyber-insurance observed in the market can result from uncertainty in insurance coverage and this uncertainty can be reduced by increasing the coverage limit for cyber loss, which can lead to a higher demand for cyber-insurance (Wang, 2017).

[53] We consider temporal dependency in cyber loss maxima series. The extremes with trends have been modeled in some fields, such as in hydrologic studies (Gilli, 2006), and we take this approach for different time frames to more accurately model the statistical process of the cyber loss maxima.

a quota share treaty is signed with a reinsurer and a form of deposit insurance contract with the government as the last resort.[54]

The rest of the paper is structured as follows. In Section 2, we first present the theoretical background of this study, followed by the methodological framework of the paper. The data used in the empirical study are described in Section 3. Section 4 identifies the procedure and result of the modeling, and Section 5 presents applications to the economic values and a reinsurance design with public intervention. Finally, the conclusion and potential future research opportunities are shown in Section 6.

2 Theoretical background and methodology

Most extreme value models assume that a random variable is independent and identically distributed (i.i.d.), particularly a threshold-based extreme model (i.e., Pareto-based or power-law) with the estimation of entire densities. Diebold, Schuermann, and Stroughair (2000) claim that this basic assumption can be violated and may even undermine the performance of the tail estimator. They suggest fitting GEV to the maxima series as one of two possible ways to improve the performance, which can particularly mitigate the problem in violating the assumption. In addition to this argument, Chavez-Demoulin and Davison (2012) demonstrate that short-range dependence of extremes often arises in many cases, whereas long-range dependence seems implausible in most contexts. Thus, it can also be expected in the cyber risk context that short-range dependence of cyber loss maxima can emerge. In the following sections, we briefly describe the theoretical background of extreme value theory and dependency in extremes and introduce four types of extreme value copula in the bivariate setting.

Extreme value theory

Two main approaches in extreme value theory (EVT) exist: the block maxima (BM) and the peaks-over-threshold (POT). The former model leads one to focus on the maximum observation in a certain block (usually time block), showing that such maxima approximately follow an extreme value distribution (Ferreira and De Haan, 2015). One using POT method needs to set a threshold level to filter the observations in excess of the threshold, then model two separate domains by the threshold. The observations over the threshold are typically approximated by a

[54] Our analysis results from the database at the aggregate level of cyber risk in American industries, which supports the idea that there is a need for an industry or national level plan against cyber risk. Hence, our results suggest a need for an industry- or regional-level risk pooling (see, e.g., Dumm, Johnson and Watson, 2015). We also expect that the modeling outcomes and the procedure help risk managers and cyber-insurers, who still suffer from a lack of standardization in modeling and underwriting cyber risk.

generalized Pareto distribution using a tail index estimator (De Haan and Ferreira, 2006, Chapter 3). The POT method is frequently used in many studies working on rare events, as POT can usually provide better flexibility than BM due to a flexible choice of the threshold (Ferreira and De Haan, 2015). However, as the POT method heavily depends on the choice of the threshold and it is hard to identify the optimal level of the threshold, the tail fit by GPD might not be optimal. In contrast, although the choice of the block size is also an issue in the estimation, the BM method is implemented with the fixed size of the observations to approximate the maxima distribution (Ferreira and De Haan, 2015).[55] Unlike most previous studies on cyber loss data using the POT method, here we focus on the BM method to identify the predictive distribution of the maximum cyber losses in different time blocks.

Let $X_i, i = 1, \ldots, n$ be an i.i.d. random variable in the real space, $\mathbb{R}$, following a distribution function, F. Let us consider the maximum of the process in a certain timeframe, $M_j, j = 1, \ldots, b$, where b denotes the number of blocks. The size of n is determined by the choice of the timeframe, for example, $n = 365$ (366 in leap years) for the annual timeframe. The total number of observations, $b \times n$, is split into b blocks with size n. The maximum process is then defined as follows (Ferreira and De Haan, 2015):

$$M_j = \max_{(j-1)n < l \leq jn} X_l \tag{1}$$

The distribution function for the maximum process is

$$\mathrm{P}\big(M_j \leq x\big) = \mathrm{P}\big(X_{(j-1)n+1} \leq x, \ldots, X_{jn} \leq x\big) = F^j(x) \tag{2}$$

The maxima M_j converges to $\bar{x}_j$ in probability ($M_j \xrightarrow{p} \bar{x}_j$) as $n \to \infty$ almost surely, where $\bar{x}_j$ is the right end point of the distribution function $F^j(x)$ (i.e., $\bar{x}_j = \sup\{x \in \mathbb{R}: F^j(x) < 1\} \leq \infty$). To determine the asymptotic feature of the sample maxima, M_j, it is customary to normalize the maxima by using constants $\sigma > 0$ and $\mu \in \mathbb{R}$ so that

$$\lim_{n \to \infty} F^j\big(\sigma M_j + \mu\big) \xrightarrow{d} G\big(M_j\big), \tag{3}$$

where $\xrightarrow{d}$ indicates the convergence in distribution.

[55] As typically found in many statistical estimation problems, there exists a trade-off in choosing the block size: bias and variance (Embrechts, Klüppelberg, and Mikosch, 1997, Chapter 6). Usually, the small size of the block causes a bias issue whereas a large block triggers a larger variance.

We call the distribution function, G, a *max-stable distribution* if its i.i.d. random variable, $X_i, i = 1, \dots, n$, satisfies $\max(X_1, \dots, X_n) \xrightarrow{d} \sigma X + \mu$ for appropriate constants $\sigma > 0$ and $\mu \in \mathbb{R}$ (Embrechts, Klüppelberg, and Mikosch, 1997, Definition 3.2.1). Equation (3) comes from the following theorem by Fisher and Tippett (1928), called the Fisher–Tippett theorem or extreme value theorem:

Theorem 1 (Fisher–Tippett theorem) Let X_n be an i.i.d. random variable. If there exist constants $\sigma > 0$ and $\mu \in \mathbb{R}$ and a non-degenerating function, K, such that

$$\frac{X_n - \mu}{\sigma} \xrightarrow{d} K(x), \tag{4}$$

then K belongs to the following distribution family for extreme values:

$$G_\gamma(x) = \begin{cases} \exp\left[-(1+\gamma x)^{-\frac{1}{\gamma}}\right], & \gamma \neq 0 \\ \exp[-\exp(-x)], & \gamma = 0 \end{cases} \tag{5}$$

where $\gamma \in \mathbb{R}$ and $1 + \gamma x > 0$.[56]

This family is a limiting distribution, as shown in equation (3), and called a generalized extreme value distribution (GEV). Combining the max-stable property for the maxima, we refer to Coles (2001, p. 50) for the following theorem:

Theorem 2 A distribution is max-stable if and only if it is a GEV distribution.

The parameters of the family consist of location (μ), scale (σ) and shape (γ), where the location parameter determines the position of the distribution's maximum and the scale parameter describes the distribution's spread. Depending on the choice of the shape parameter, this family ends up with one of the following three distributions:

Type I (Gumbel, $\gamma = 0$): $$\exp[-\exp(-x)], \quad -\infty < x < \infty \tag{6}$$

Type II (Fréchet, $\gamma > 0$): $$\begin{cases} 0, & x \leq 0 \\ \exp[-x^{-\alpha}], & x > 0, \alpha > 0 \end{cases} \tag{7}$$

Type III (Weibull, $\gamma < 0$): $$\begin{cases} \exp[-(-x)^{\alpha}], & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases} \tag{8}$$

Type I (Gumbel) is light-tailed on the right side, Type II (Fréchet) is heavy-tailed, and Type III (Weibull) is short-tailed. The location, scale, and shape parameters of a GEV distribution can

[56] The proof of Theorem 1 can be found in De Haan and Ferreira (2006, p.7).

be estimated by maximizing the following log-likelihood function derived from the probability density function of the GEV distribution:

$$l(\mu, \sigma, \gamma; x) = -b\log\sigma - \left(\frac{1}{\gamma} + 1\right)\sum_{i=1}^{b} \log\left[1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{b}\left[1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}. \tag{9}$$

Temporal dependency in extremes

When one discusses the presence of temporal dependency in an extreme series, the following definition, called the *distributional mixing condition* or *Leadbetter's $D(u_n)$ condition*, is referred to (Leadbetter, 1983):

Definition 1 (Distributional mixing condition) Let $\tilde{X}_i, i = 1,2, \ldots$ be a stationary series and δ_n be a sequence of constants. A stationary series, $\tilde{X}_i$, satisfies the distributional mixing condition if, for all $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$\begin{aligned} &\left| F\left\{\tilde{X}_{i_1} \le \delta_n, \ldots, \tilde{X}_{i_p} \le \delta_n, \tilde{X}_{j_1} \le \delta_n, \ldots, \tilde{X}_{j_q} \le \delta_n\right\} \right. \\ &\left. - F\left\{\tilde{X}_{i_1} \le \delta_n, \ldots, \tilde{X}_{i_p} \le \delta_n\right\} F\left\{\tilde{X}_{j_1} \le \delta_n, \ldots, \tilde{X}_{j_q} \le \delta_n\right\}\right| \le \alpha_{n,l}, \end{aligned} \tag{10}$$

where $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ such that $l_n = o(n)$ as $n \to \infty$.

In the case of an independent series, the subtraction in equation (10) is exactly zero for any sequence $\{\delta_n\}$. It is required in a general case that the above condition holds only for a certain sequence of $\{\delta_n\}$ that increases with *n*. For this sequence, the condition holds in a way that, for a set of clearly distinct variables, the subtraction in equation (10) is not zero, but sufficiently close to zero so that it does not influence the limiting laws for extremes (Leadbetter, 1983). Based on Definition 1, the following theorem describes the limiting distribution of an extreme series in the presence of temporal dependency (Chavez-Demoulin and Davison, 2012).

Theorem 3 Let $\tilde{X}_i, i = 1,2, \ldots$ be a stationary series satisfying the distribution mixing condition in Definition 1 and let $\tilde{M}_n = \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$. Suppose there exists an independent series $X_i, i = 1,2, \ldots$ with X, which has the same distribution as $\tilde{X}_i$, and $M_n = \max\{X_1, \ldots, X_n\}$. If M_n has a non-degenerate limiting distribution G under the non-degenerate limit law, $\mathrm{P}[(M_n - b_n)/a_n \le x] \to G(x)$, then it follows that

$$\mathrm{P}\left[\frac{\tilde{M}_n - b_n}{a_n} \le x\right] \to G^{\theta}(x), \tag{11}$$

for some extremal index, $0 \le \theta \le 1$.

The extreme parameter, θ, quantifies the degree of dependency in extremes, showing that $\theta = 1$ indicates a completely independent process and $\theta = 0$ addresses an increasing level of dependency in extremes. Importantly, a non-degenerate limiting distribution, G, is necessarily an extreme value distribution so that the distribution of maxima with short-range temporal dependency is also an extreme value distribution with the max-stability property in Theorem 2. This implies that the limiting distribution of the maxima of a stationary series is linked with the limiting distribution of an independent series if the distribution mixing condition is fulfilled. For more detail on the theoretical background, see Beirlant et al. (2004, Chapter 10).

Extreme value copula

The joint distribution of bivariate extremal random variables, X and Y, can be expressed as (Sklar, 1959):

$$G(X, Y) = \mathrm{P}[X \leq x, Y \leq y] = C[F_X(x), F_Y(y)], \tag{12}$$

where $x, y \in \mathbb{R}$ and C is some extreme-value copula, which is an asymptotic dependence structure of component-wise maxima (Ghorbal, Genest and Neslehova, 2009).

If a copula, C, is max-stable in the d-dimensional setting expressed as,

$$C(\mathbf{u}) = \left\{C\left(u_1^{1/r}, \dots, u_d^{1/r}\right)\right\}^r, \tag{13}$$

where $\forall\, \mathbf{u} \in [0,1]^d$ and $\forall\, r > 0$, then the function C is an extreme-value copula (Kojadinovic, Segers and Yan, 2011).

Any extreme-value copula C can be expressed with its Pickands dependence function (Pickands, 1981) as follows:

$$C(u, v) = \exp\left[log(uv) \cdot A\left\{\frac{log(v)}{log(uv)}\right\}\right], \tag{14}$$

where $u, v \in [0,1]$ are uniform distributions from $F_X(x), F_Y(y)$, respectively, through probability integral transform and $A\{\cdot\}: [0,1] \to [0.5, 1]$ is the Pickands dependence function that is convex and fulfills the conditions $\max(t, 1-t) \leq A(t) \leq 1$ for all $t \in [0,1]$ and $A(0) = A(1) = 1$ (Ghorbal, Genest, and Neslehova, 2009; Kojadinovic and Yan, 2010).

The Pickands dependence function, $A\{\cdot\}$, in equation (14) is determined by the extreme-value copula family to which it belongs (Genest et al., 2011). The following specifies the Pickands dependence function for each extreme-family:

Gumbel Hougaard: $A(t) = \left(t^{\theta} + (1+t)^{\theta}\right)^{1/\theta}, \theta \geq 1$ (15)

Galambos: $A(t) = 1 - \left(t^{-\theta} + (1+t)^{-\theta}\right)^{-1/\theta}, \theta > 0$ (16)

Tawn: $A(t) = \theta t^2 - \theta t + 1, \theta \in [0,1]$ (17)

Husler-Reiss: $A(t) = t\Phi\left(\frac{1}{\theta} + \frac{1}{2}\theta \ln\left[\frac{t}{1-t}\right]\right) + (1-t)\Phi\left(\frac{1}{\theta} + \frac{1}{2}\theta \ln\left[\frac{1-t}{t}\right]\right), \theta \geq 0$ (18)

Based on the theoretical description above, Figure 2 highlights the overall methodological structure of this study with the theoretical background for each step described above. Detailed empirical procedures are illustrated in Section 4.

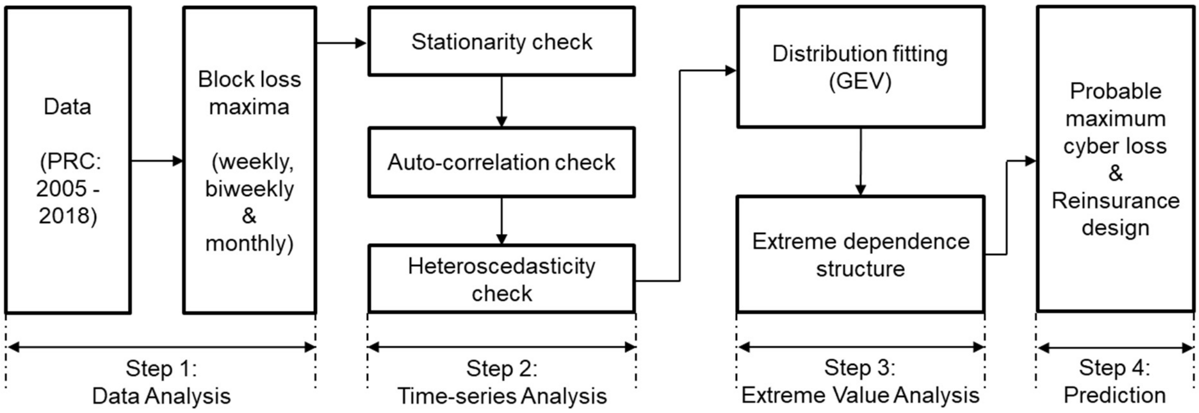


Figure 2. The methodological process for the statistical estimation of cyber loss maxima.

3 Data

The dataset in the empirical study is derived from Privacy Rights Clearinghouse (PRC), which is a nonprofit organization collecting information on data breach events in the U.S. since 2005 (PRC, 2019).[57] PRC has accumulated 9,002 observations (as of January 31, 2019) from the beginning of 2005 to the end of 2018. The PRC database has recently been used in many academic studies and is widely accepted in practice, ensuring the reliability of the database (see, e.g., Edwards et al., 2016; Eling and Loperfido, 2017; Rasoulian et al., 2017; Eling and Jung, 2018; Kamiya et al., 2018; Hofmann, Wheatley and Sornette, 2019). Forming the largest public database for cyber loss, PRC continuously increases its data pool by adding and updating the

[57] The definition of cyber risk can incorporate different types of risks associated with information technology and security assets. Data breach risk as part of the operational risk (Eling and Wirfs, 2019) is one of the most prominent types in the category of cyber risk and is a main loss event covered by cyber-insurance in the current market. This paper deals with data breach risk by representing the majority of cyber risk incidents as most literature in the cyber risk context have done (see, e.g., Edwards et al., 2016; Wheatley et al., 2016; Eling and Jung, 2018; Hofmann et al., 2019).

events, each of which can be found by either a government agency or a verifiable media source.[58]

The dataset incorporates information on the breach date, the event location, the entity level, the loss type and the total number of breached records, among which the total number of breached records is the only numerical variable representing the loss severity. The number of records includes zero values that demonstrate either no loss amount in the events or records still under ongoing investigation. Therefore, the inclusion of zero values might hinder the reliability of the modeling result (Edwards et al., 2016; Eling and Jung, 2018) and trigger the self-selection bias, which—according to Kamiya et al. (2018) using the same dataset—does not meet cyberattack notification laws and might not be compulsory for firms to disclose information. For this reason, we do not count zero values in the empirical study. We further exclude the losses in the category of “Unknown” risk type, which does not include any information on loss events. As a result, we end up with 6,047 data items. Table 1 describes the variables from the database and the bivariate categorization of the variables.

Table 1. Bivariate Risk Setting

Risk type	Variable	Explanation
Malicious	Hacking (HACK)	Hacking attack by outsiders or infection by malware
	Insider (INSD)	Breached by an insider (e.g., employee or contractor)
	Payment card fraud (CARD)	Fraud involving debit and credit cards
Negligent	Portable device (PORT)	Lost, discarded or stolen portable devices
	Stationary device (STAT)	Lost stationary computers
	Unintended disclosure (DISC)	Privacy information disclosed unintentionally
	Physical loss (PHYS)	Lost, discarded or stolen non-electronic information

Note: The information in this table is derived from PRC (2019), and the bivariate categorization (malicious vs. negligent) is based on Edwards et al. (2016).

Descriptive statistics of cyber loss maxima

We consider weekly, biweekly and monthly timeframes in the statistical fitting procedure.[59] The block length (time length) is important to extract the maxima from the dataset; if the length is too short, Theorem 1 and 2 supporting the maxima process might not be valid, whereas too lengthy of a block might trigger a lack of data with which to work (Bücher and Segers, 2014). For this reason, we consider weekly, biweekly and monthly loss maxima series. As previously

[58] Despite the provision of the largest data pool for cyber risk, there is a possibility of bias in the data completeness (backfilling bias) so that the data collected more recently might be relatively less complete than the data from an older period. However, the impact of such bias might be small as the database is continuously updated daily using reliable data source standards (see, e.g., Eling and Wirfs, 2019, for a similar argumentation with SAS OpRisk database).

[59] We do not take into account time frames shorter than a week or longer than a month because 1) the size of observations is generally small for timeframes shorter than a week, so that the maxima distribution is not significantly different from the entire spectrum of the loss density, and 2) the number of observations for timeframes longer than a month, particularly for sub-periods, is not sufficient.

mentioned, we attempt to split the data period into two parts based on the trend of the loss severity with a statistical test for a structural break in the entire loss density (see Appendix B). We find that there is a significant change in the loss size between 2013 and 2014, which leads us to split the data into the pre- and post-2014 periods. This change can be evidenced by the historical extreme loss sizes in Table B2, where most extreme losses occurred over last five years (i.e., after 2014). Panel A of Table 2 describes the statistics for the loss maxima of the complete dataset, where it is observed that no loss occurrence exists in at least one biweekly period, whereas an economic entity can be faced with a breach at least once a month based on the data. Moreover, this historical data show that the largest monthly loss to an entity could be at least the size of 18,000 records. The losses by malicious risks are in general severer than those by negligent risks, which is observed in panels B and C of Table 2.

Table 2. Descriptive Statistics

Panel A: Complete dataset (number of breached records)

Data	Block	mean	SD	skewness	kurtosis	Max	median	min
Entire	Weekly	13.53m	132.29m	17.97	372.20	3,000m	37,000	0
	Biweekly	26.59m	186.14m	12.67	184.60	3,000m	130,000	0
	Monthly	55.90m	272.03m	8.52	82.85	3,000m	625,000	18,000
Pre-2014	Weekly	1.69m	13.84m	13.97	227.74	250m	36,500	0
	Biweekly	3.24m	19.43m	9.84	112.46	250m	100,000	0
	Monthly	6.13m	26.78m	7.41	61.72	250m	310,500	32,000
Post-2014	Weekly	34.26m	219.31m	10.76	132.01	3,000m	38,000	0
	Biweekly	67.52m	306.50m	7.53	64.17	3,000m	309,079	1,359
	Monthly	145.47m	442.12m	4.95	27.19	3,000m	3,960,000	18,000
Panel B: Malicious risk								
Entire	Weekly	10.58m	121.59m	21.28	503.54	3,000m	10,868	0
	Biweekly	20.91m	171.32m	15.02	250.08	3,000m	59,710	0
	Monthly	44.01m	251.08m	10.12	112.74	3,000m	350,000	0
Pre-2014	Weekly	2.04m	15.83m	11.75	156.82	250m	5,389	0
	Biweekly	3.98m	22.22m	8.25	76.56	250m	43,764	0
	Monthly	7.60m	30.97m	5.99	39.05	250m	176,500	0
Post-2014	Weekly	25.48m	201.72m	12.82	180.44	3,000m	24,188	0
	Biweekly	50.68m	283.02m	8.99	88.08	3,000m	220,000	0
	Monthly	109.54m	412.19m	5.93	37.69	3,000m	2,050,000	18,000
Panel C: Negligent risk								
Entire	Weekly	3.58m	53.22m	23.60	594.15	1,370m	11,300	0
	Biweekly	7.10m	75.09m	16.64	294.69	1,370m	38,580	0
	Monthly	15.00m	110.42m	11.18	132.46	1,370m	128,000	565
Pre-2014	Weekly	0.54m	5.24m	14.00	201.55	80m	14,875	0
	Biweekly	1.05m	7.39m	9.81	98.09	80m	43,000	0
	Monthly	2.23m	10.81m	6.50	42.16	80m	151,000	565
Post-2014	Weekly	9.03m	88.63m	14.12	210.88	1,370m	7,693	0
	Biweekly	17.71m	124.69m	9.89	102.97	1,370m	19,898	684
	Monthly	37.98m	182.94m	6.51	43.98	1,370m	66,399	911

Note: m stands for million.

4 Empirical estimation

4.1 Time series analysis of extremes

First, we test whether there exists a dynamic feature of maximum cyber losses through the following steps.

Step 1: Detect the evidence of a dynamic feature and pre-select potential fit models for the data.

Step 2: Estimate the parameters and the likelihoods of the potential models and conduct statistical tests to determine the best fit model.

Step 3: Check whether the estimated model is appropriate by carrying out diagnostic tests suggested by Box and Jenkins (1976).

Step 1 typically consists of the graphical identification using time series plots of the data and autocorrelation function plots. As observed in the right graph of Figure 1, a clear distinction in the trend exists between pre-2014 and post-2014, and a cluster of large losses is observed, especially in the post-2014 period, whereas a cluster of small losses is found in the pre-2014 period.[60] This trend showing evidence of temporal dependency is supported by the autocorrelation function in the lower panel of Figure C1, which confirms the presence of serial correlation in the maximum loss process, particularly in the complete data as well as in bivariate cases on a shorter time horizon (weekly and biweekly).

Testing for stationarity

Prior to investigating the fitted model for autocorrelation, we conduct three statistical tests to see whether the cyber maxima processes are stationary: augmented Dickey-Fuller (ADF) test (Said and Dickey, 1984), Phillips-Perron (PP) test (Phillips and Perron, 1988) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992). The null hypothesis of the first two tests indicates the presence of a unit root, whereas that of the KPSS test addresses the stationarity of a series. Thus, if a cyber loss maxima series is stationary, the conclusions of ADF and PP tests would be opposite that of the KPSS test. Table C1 indicates that all series tested for the entire, pre-2014 and post-2014 periods can be regarded as stationary series confirmed by at least two tests. The stationarity of the cyber loss maxima series with a mean-reverting process is graphically observed in the upper panel of Figure C1, where no

[60] Appendix B indicates that a clear distinction exists between pre-2014 and post-2014 in severity trend and between pre-2010 and post-2010 in frequency trend over the data period. This finding is supported by three statistical tests investigating the presence of a structural break in the time series data. See Appendix B for more detail.

significant trend is found. Note that our separation of the time period with a structural break is based on the entire spectrum of the loss severity density, and the maxima series for three data periods turns out to be stationary independent of the structural break in the entire density. The separation in time period is important in the cyber risk modeling because it can help capture the magnitude of a loss event due to the risk of change over time.

Testing for autocorrelation and heteroscedasticity in extremes

With stationary maximum cyber losses, Step 2 leads us to test the autoregressive model with a range of lags to find the best fit for the loss process. The autoregressive model with p lags for the maximum cyber loss process is defined as:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{19}$$

where ϕ_i, $i = 1, \ldots, p$, is the autocorrelation coefficient and $\varepsilon_t \sim N(0, \sigma^2)$ is the error term with homogeneous volatility over time under the normality condition.

As the maximum cyber losses are proven to be stationary, the autocorrelation coefficient, ϕ_i $(i = 1, \ldots, p)$, should be smaller than 1. Three information criteria are used to determine the best fit model: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the corrected Akaike information criterion (AICc).[61] AICc is especially significant when the sample size is small because, in such a case, AIC might overfit by selecting the model with the most parameters (Hurvich and Tsai, 1989). The testing procedure is carried out with up to 24 lags to determine whether it is a long memory series; Panels A and B of Table C2 display the fitted models and three measures for different maximum loss vectors over three data periods.[62]

As evident in Figure C1, complete dataset and malicious variables are autocorrelated, whereas negligent variables have no temporal dependency. Furthermore, from the malicious variables, temporal dependency can exist on a shorter time horizon as loss occurrences might be affected by a common shock in a shorter timeframe. For the post-2014 loss maxima, the same

---

[61] The three criteria can be mathematically expressed as (Akaike, 1973; Schwarz, 1978; Hurvich and Tsai, 1989):

$$AIC \coloneqq -2Loglik(\Theta) + 2k$$
$$BIC \coloneqq -2Loglik(\Theta) + k \cdot \ln(n)$$
$$AICc \coloneqq AIC + \frac{2k^2 + 2k}{n - k - 1}$$

where $Loglik(\Theta)$ is the log-likelihood function with a parameter set of $\Theta = (\theta_1, \ldots, \theta_k)$, k is the number of parameters and n is the number of observations.

[62] The figures in the table are scaled down by dividing the losses by 1 million because, with a large number of losses, the inverse of the Hessian matrix might crash when calculating the covariance matrix of the coefficients. The results are indifferent from the adjustment of the scale.

specification of autoregressive models is identified for the complete dataset, whereas no autocorrelation is observed in the bivariate setting. In contrast, weekly and biweekly series of the complete and malicious data are found to be serially correlated, particularly between small losses for the pre-2014 period.

A time series can be characterized by volatility clustering, which features periods of low volatility followed by periods of high volatility. In this case, the error term in equation (19) is changing volatility over time; hence, the volatility at time *t* would be affected by the past values of the volatility. Given the fitted autoregressive models in panels A and B of Table C2 are applied, the presence of heteroscedasticity is investigated using the ARCH test with different lags developed by Engle (1982).[63] The test is based on the fitted autoregressive models; thus, we test variables that have temporal dependency in the series in panel C of Table C2. The number of lags is determined by an integer sequence of 1:min(24,n) with a step of 4, matching the lagged effects of considered periods (e.g., four weeks are equivalent to one month). The results verify the absence of autocorrelation of innovations for the entire, pre-2014 and post-2014 periods, thereby concluding homoscedasticity of the error term with i.i.d. assumption.

As a final step, we check the fitted models (see panels A and B of Table C2) by looking at the residuals using two diagnostic tests: Box-Pierce test and Ljung-Box test.[64] The test results in Table C3 confirm that the residuals from the fitted models are serially correlated for three periods by rejecting the null hypothesis, except for the maximum malicious loss process on a biweekly basis (entire period) that has, however, a sufficiently low p-value of the tests.

---

[63] Suppose the maximum cyber loss is following AR(p) defined by the following auxiliary regression:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$$

The squared residuals are regressed on p of its own lags, which is constructed by

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_p \varepsilon_{t-p} + u_t,$$

where $u_t \sim i.i.d.$ is white noise.

The null and alternative hypotheses of the test are described as:

$$H_o: \alpha_0 = \alpha_1 = \cdots = \alpha_p = 0,$$
$$H_1: \alpha_0 \neq 0 \ or \ \alpha_1 \neq 0 \ or \cdots or \ \alpha_p \neq 0.$$

The test statistic is defined as the number of observations multiplied by the coefficient of multiple correlation from the auxiliary regression, which follows $\chi^2(p)$.

[64] Two tests are closely linked to each other as follows (Box and Pierce, 1970; Ljung and Box, 1978):

$$BP_Q = n \sum_{k=1}^{p} \hat{\rho}_k^2$$
$$LB_Q = n(n+2) \sum_{k=1}^{p} \frac{\hat{\rho}_k^2}{n-k},$$

where n is the number of observations, p is the number of lags, and $\hat{\rho}_k$ is the sample autocorrelation at lag k. The null hypothesis of two tests is that the residuals are independent. Thus, if temporal dependency exists in the residuals, the null hypothesis would be rejected.

4.2 Fitting extreme value distribution

As we discussed in Section 4.1, the maximum cyber loss process is stationary and serially dependent, but not heteroscedastic. We fit the GEV distribution for stationary loss maxima and investigate which type of the distribution the process follows. In panel A of Table D1 (Appendix D), we estimate distribution parameters using the maximum likelihood estimation method and determine negative log-likelihood and Akaike information criterion. We also perform two goodness-of-fit tests to check whether the fitting is good: Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D).[65] All variables considered in three periods demonstrate a positive value of the shape parameter ($\gamma > 0$), showing that all maximum cyber losses in these timeframes are heavy-tailed by following type II (Fréchet) of GEV distribution. Importantly, the size of the shape parameter for the post-2014 period increased by 47.5%[66] on average compared to that for the pre-2014 period, indicating a heavier tail of the cyber loss maxima in recent years. This result clearly demonstrates that the magnitude of an extreme cyber event continues to grow, thereby implying the necessity of a modeling framework for such an event.

However, several maxima series do not perfectly fit GEV distribution, as proven by two goodness-of-fit tests (see panel A of Tables D1 and D2). This result is also evident in the graphical diagnosis using the QQ plot, probability density plot and cumulative density plot displayed in Figure D1, where the empirical lines do not exactly match the theoretical lines. In particular, looking at the probability and the cumulative density plots, we observe significant deviations between empirical values and theoretical values for the weekly malicious series. The result of the diagnosis can lead us to reject GEV estimation for these series, but test other types of skewed distribution.

To identify a better fit, we check the following right-skewed distributions[67]: log-normal, gamma, Cauchy, inverse Gaussian, Burr XII, and generalized Pareto. Panel B of Table D1 displays Akaike information criteria for the distributions to be compared. The Fréchet type of

65 The tests are specified as follows (D'Agostino and Stephens, 1986):

$$T_{KS} = \sup|F_n(x) - F(x)|,$$

$$T_{AD} = \mathrm{n}\int_{-\infty}^{\infty}\left(F_n(x) - F(x)\right)^2 dx,$$

where $F_n(x)$ is the empirical distribution of a random variable, x, and $F(x)$ is the fitted distribution function of x. The null hypothesis of the above tests is as follows:

$$H_0: X \sim F_\theta(x: \theta),$$

where θ is a set of parameters for the distribution F.

66 This number comes from the calculation based on the shape parameters of fitted GEV loss series and fitted GPD series that fail to fit GEV.

67 They are used to model non-negative variables that are usually right-tailed and widely applied in the literature (see, e.g., Frachot, Georges and Roncalli, 2001; Moscadelli, 2004; Fu and Moncher, 2004; Shevchenko, 2011; Eling, 2012; Frees, Lee and Yang, 2016; Eling and Jung, 2018).

GEV distribution as the fitted model for the cyber loss maxima is also a good fit compared to other right-tailed distributions as it shows the lowest level of AIC. However, as demonstrated by the graphical depiction, GPD can be a better fit for seven series of loss maxima (see Appendix D) according to the goodness-of-fit test in panel C of Table D1 and graphical depiction in Figure D2. Therefore, we can conclude that the cyber loss maxima for the considered timeframes follow the Fréchet type of GEV distribution in general, but maxima series of malicious risk factors for shorter timeframes fit GPD better.

4.3 Dependence structure of extremes

The bivariate extreme risks defined in Table 1 are possible risk factors considered when underwriting cyber-insurance. Two risk factors are clearly distinct with respect to the nature of loss occurrence: Malicious risk can be a source of systemic risk affecting a wide range of systems in the interconnected network environment, and negligent risk could be limited to a loss event involving a single entity. Therefore, from the insurer's perspective, malicious risk needs to be considered as the more significant factor in a potential cyber-risk pool. Indeed, many policies traded in the current market (e.g., the U.S. market) mainly cover the losses by malicious attacks (Romanosky et al., 2017). However, negligent risk should not be disregarded as this risk shows a similar pattern in loss severity to that of malicious risk, as seen in Figure B1; hence, it is also expected to extend the coverage by underwriting the loss by negligent risk source. In this section, we assume that a cyber-insurer pools two possible cyber-insurance lines (malicious vs. negligent) that are heterogeneous in the risk source underwritten in the policy.

When estimating the size of risk that an insurer could face, the insurer can consider two possibilities in pooling: i.i.d. assumption on risk factors (lines of business) and dependency in risk factors. Obviously, the first algorithm is more tractable for pooling the risks and carrying out the estimation procedure both theoretically and practically. However, statistical properties of risk factors are usually heterogeneous, especially in the tail behavior, and a statistically significant dependence structure might exist in losses. In addition, the i.i.d. assumption might underestimate the level of risk, which does not reflect any correlated risk, whereas the dependence structure in a risk pool can provide a diversification benefit compared to the sub-additive assumption on the risk factors. Based on the finding in the previous section that maximum cyber losses follow the generalized extreme value distribution with the Fréchet family, we fit the loss maxima in the bivariate setting into an extreme dependency.

First, when modeling extreme dependency, it is necessary to test whether the copula represented by Sklar's theorem (1959) can also be expressed by the Pickands dependence function A in equation (14), positing the null hypothesis in the following[68]:

$$H_0 : C \in C_{ev}, \tag{20}$$

where C_{ev} stands for an extreme value copula with the Pickands dependence function, A.

Thus, if the null hypothesis is rejected at a critical level, the copula function between variables cannot be represented by an extreme value copula, implying that an extreme dependency (in the right tail) does not exist. The test result is illustrated in panel A of Table E1 (Appendix E), showing that dependency in the weekly and biweekly setting does not present the extreme trend (except for the biweekly maxima of the post-2014 cyber loss). This might result from the fact that the level of maxima on a weekly or biweekly basis is not significantly extreme, especially compared to other series on a longer time basis (see Table 2). However, the monthly maxima series turns out to be extremely dependent so that extreme value copulas fit well with Pickands dependence functions.

Panel B of Table E1 provides Akaike information criteria and the results of the goodness-of-fit test[69] for copulas to determine the most appropriate function for the dependence structures. Based on the results in panel A, we fit extreme value copulas into three maxima series with extreme dependency and also take into consideration other types of copula family for the series not showing extreme dependency. For series with extreme dependency, four copula functions fit each extreme dependency based on the lowest AIC and the goodness-of-fit test results. For series that are not extremely dependent, we find a clear tendency that lower dependency with

---

68 We employ the nonparametric rank-based test for bivariate extreme value dependency developed by Kojadinovic and Yan (2010). The test statistic is formulated by comparing the empirical copula C_n (nonparametric copula estimation) with a nonparametric copula estimation, $C_{A_{n,c}}$, under extreme value dependency defined in equation (14) with the Pickands dependence function, $A_{n,c}$. The test statistic is described as:

$$S_n = \int_0^1 \int_0^1 n \left| C_n(u,v) - C_{A_{n,c}}(u,v) \right|^2 du dv.$$

69 The goodness-of-fit test for extreme value copulas is developed by Genest et al. (2011). Suppose there exists a bivariate random sample $(M_i^X, M_i^Y), i = 1, \dots, n$ from some unknown continuous distribution G, whose underlying copula is an extreme value copula with Pickands dependence function A. The null hypothesis to test whether the underlying extreme value copula is appropriately specified is as follows (Genest et al., 2011):

$$H_0 : A \in \mathcal{A} = \{A_\theta : \theta \in \Theta\},$$

where $\mathcal{A}$ is a set of parametric extreme value copulas with the parameter, θ.
The test statistic developed by Genest et al. (2011) is based on the Cramer-von-Mises statistic by comparing a nonparametric estimator A_n with a parametric estimator A_{θ_n} as follows:

$$S_n = \int_0^1 n \cdot \left| A_n(t) - A_{\theta_n}(t) \right|^2 dt.$$

Clayton copula is observed from the weekly setting, which has relatively smaller loss maxima. In contrast, longer-range series (biweekly) and the setting for the more recent period (post-2014) demonstrate upper dependency and symmetry fitting the Joe and Frank copulas, respectively, which could be important for insurers concerned about the correlated risk triggering a huge extreme loss.

5 Probable maximum cyber loss and reinsurance design

The modeling procedure in the previous sections illustrates that cyber loss maxima series are stationary, that short-range temporal dependency (weekly and biweekly) exists, that GEV distribution fits the series well, and extreme dependency exists for monthly series that can fit extreme value copula models. With the outcome from the procedure, in this section we estimate a potential level of the worst risk using a quantile-based measure called probable maximum cyber loss (PMCL) and propose a reinsurance portfolio with public intervention.

5.1 Probable maximum cyber loss

The probable maximum cyber loss can be defined as the worst cyber loss likely to happen as determined using the following expressions:

$$\mathrm{P}\big[\widetilde{M}_n \leq \xi_p\big] = 1 - p, \tag{21}$$

$$\xi_p = {G_{\widetilde{M}_n}^{\theta}}^{-1}(1 - p) \tag{22}$$

for some small $p \in [0,1]$, where $\widetilde{M}_n$ is a series of the cyber loss maxima, ξ_p is the probable maximum cyber loss, and $G_{\widetilde{M}_n}^{\theta}$ is the probability function of the cyber loss maxima series with the parameter of θ based on equation (11).

The above quantile-based measure is based on Value-at-Risk (VaR) at the quantile $1 - p$, but the loss vector consists of the maximum values for a certain time period, indicating a probable worst loss of extreme cases likely to occur p times out of 100 corresponding time units (e.g., years, months and days). To show how large the estimates of PMCL are, our estimates are compared with the measures from three recent references: Edwards et al. (2016), Wheatley et al. (2016), and Eling and Jung (2018). These authors use well-known right-skewed distributions to estimate extreme cyber loss, such as lognormal and generalized Pareto models.[70] It is

[70] For the sake of comparison, we implement another extreme value model, Peaks-over-Threshold (POT), with the same dataset. We consider the POT model with lognormal distribution in the body and generalized Pareto distribution in the tail, which is widely used for cyber risk analysis in the literature (Wheatley et al., 2016;

important to note that the generalized Pareto model in the tail theoretically incorporates the infinite moment of the distribution function, which means that the maximum level is by definition unmeasurable. However, Wheatley et al. (2016) employ a doubly truncated Pareto distribution to model the large breach sizes—a function that distinguishes itself from the model with the infinite moment.

Table 3 displays monthly PMCL estimates with the fitted GEV distributions[71] and the estimates of extreme losses derived from three references. The quantile, $1-p$, is set at 91.7%, 97.2% and 98.3%, indicating the worst case likely to happen once every one, three and five years, respectively. For example, considering the composite dataset over the entire period, we find that the maximum loss amount within the next year is likely to be around 61.79 million data breaches. The size of the maximum loss becomes much bigger over the longer time period, showing 692.2 million and 2,053 million breaches within the next three and five years respectively. This breach size might happen to a single entity or multiple parties based on the property of the database, which requires the industry- or national-level reaction to a possible systemic risk. A single insurer cannot afford to recover this size of loss; rather, co-insurance or a collective reinsurance plan seems more appropriate. Thus, in the next section, we attempt to practically design a reinsurance portfolio with public intervention, which is expected to provide insights into potential sizes of premiums and economic burdens to involved parties.

One of the key results in Table 3 shows that, compared to the estimate in the literature, a cyberattack successfully occurring once every five years could lead to the loss of 2,053 million data records in the U.S., which is around seven times larger than the expected maximum size (= 300 million[72]) that Wheatley et al. (2016) predicted for in the next five years. Compared to Edwards et al.'s (2016) estimate, we can observe that the gap increases to five times for the next three years. However, the largest level of Edwards et al. (2016) does not result from their estimation process, but from the largest level in their dataset; therefore, the gap between our estimate and their potential estimate might be much bigger (GEV vs. lognormal).

Eling and Jung, 2018; Eling and Wirfs, 2019). This comparison can be a good complement to the idea of the possibility of cyber dragon kings beyond Pareto-based estimation by the recent literature with a slightly older dataset than ours. The fit results with the estimated parameters for POT are described in Table G1 of Appendix G, and the graphical description is presented in Figure G1. The model is not block-based, but considers the entire spectrum of the cyber loss.

71 In order to make a reasonable comparison in this application, we estimate PMCL with monthly series that fit the extreme dependence model across three time periods (entire, pre-2014, and post-2014).

72 Wheatley et al. (2016) find that the maximum breach size grows over time at the rate of $t^{0.84}$ and expect to see such growth reach 50% over the next five years, resulting in 300 million (200 million × 150%) maximum breaches.

Table 3. Comparison of Extreme Loss Prediction

Panel A: PMCL estimates					(million breach)
		Composite	Malicious	Negligent	Dependence
Next 1 yr	Entire Period	61.79	85.22	6.33	142.83
	Pre-2014	8.53	17.18	2.99	26.73
	Post-2014	1,333.90	785.04	18.26	1,347.07
Next 3 yrs	Entire period	692.21	1,539.94	52.46	2,241.70
	Pre-2014	50.72	227.04	15.19	284.48
	Post-2014	62,693.28	20,533.20	313.14	33,004.61
Next 5 yrs	Entire period	2,053.21	5,987.12	140.76	8,723.67
	Pre-2014	117.62	784.80	32.63	876.31
	Post-2014	371,964.44	98,198.51	1,179.38	132,992.70

Panel B: Estimates of the recent literature			(million breach)
	Edwards et al. (2016) (Lognormal)	Wheatley et al. (2016) (Truncated Pareto)	Eling and Jung (2018) (Correlated risk)
Data period	Jan, 2005 – Feb, 2015	Jan, 2007 – Apr, 2015	Jan, 2005 – Dec, 2016
Data source	PRC	Open Security Foundation & PRC	PRC
Breach size estimate	**130.00**	**300.00**	**1,053.11**
Time prediction	Next 3 yrs	Next 5 yrs	1 out of 200 cases (99.5%)

Panel C: Threshold-based estimation (Pareto density in the tail)				(million breach)
		99%	99.5%	99.9%
Entire	Composite	10.81	263.64	1,001.86
	Malicious	8.18	382.35	1,407.30
	Negligent	0.50	110.21	478.38
Pre-2014	Composite	0.94	29.91	111.47
	Malicious	4.24	46.58	149.14
	Negligent	1.22	13.87	56.66
Post-2014	Composite	2.43	315.90	1,240.32
	Malicious	34.19	524.63	1,761.11
	Negligent	0.63	202.33	734.77

Note: PMCL estimates in panel A are based on the monthly maxima series, which demonstrate stationarity, GEV fit and extreme dependency. Panel C shows extreme loss estimates based on the raw data on a daily basis and, thus, the value-at-risk measures.

One might argue that our dataset includes more recent observations than the dataset in the literature, which can account for this difference to some extent. For the sake of fair comparison, we implement the POT method with lognormal distribution in the body and generalized Pareto distribution in the tail, as used in most literature. We derive the risk measures (Value-at-Risk) at three extreme quantiles (i.e., 99%, 99.5% and 99.9%) and simply compare them with our estimates in three time horizons. Based on this comparison, our estimates from the composite risk for the entire data period are on average 3.46 times larger than the POT estimates from the same risk type, and a huge gap between estimates from the post-2014 period exists. Thus, the

POT method heavily underestimates the change in the magnitude of risk over time, showing almost 98% of the underestimation on average compared to our approach.

Our result further highlights that PMCL by a malicious cyberattack could be even more extreme than the one by a negligent risk—almost 28.5 times bigger for the entire period. This might result from the fact that a loss by a negligent risk is relatively limited in the effect on a single business entity, whereas a loss event by a malicious risk is usually triggered by an infectious attack simultaneously affecting multiple agents. Thus, a malicious case could cause a silent cyber risk in numerous P&C policies (systemic risk in non-life insurance business) so that insolvency issues might be raised for an insurer. Furthermore, this difference is prominent for cyber-insurers required to consider cyber risk classification in the underwriting process to set the coverage limit.

The structural break leads to the finding that the PMCL estimates for the post-2014 period are on average 433 times larger than those for the pre-2014 period—a gap that is the numerical realization of the right panel in Figure 1. It demonstrates that cyber loss process is changing quickly over time as the information technology is more advanced and dependency on such technology is rapidly increasing. Thus, insurers must develop an underwriting process that can capture rapidly changing risk dynamics over time. Otherwise, insurers would fail to predict the size of an extreme loss in the coming years, which could be much larger than the size for the post-2014 period. In this regard, our statistical procedure can provide insights for considering temporal dependency and the distributional property of cyber loss maxima, thereby leading insurers in the right approach.

Based on the assumption that a cyber-insurer is pooling two possible cyber risk factors, the PMCL estimates with the correlated risk are on average 3.27 times larger than those without the correlated risk for the entire data period. Comparing Eling and Jung's (2018) estimate with the correlated risk, we observe that our estimates for the next three and five years are almost fivefold. Overall, the defined probable maximum loss by cyber risk outweighs the extreme loss estimates in the recent literature with lognormal and Pareto-based models. This finding implies that a fundamental change in modeling to estimate a possible extreme cyber loss is needed for (re)insurers and regulators in order to be able to predict a cyber dragon king to the industries.

5.2 Design of cyber-reinsurance portfolio with public intervention

As previously mentioned, an extreme event might trigger a number of claims simultaneously, possibly leading a potential cyber-insurer to become insolvent. Furthermore, the nature of risk is heterogeneous depending on the source of risk, network environment, the security level, and other cyber-specific factors; hence, a variation in the loss ratio could occur due to the fluctuations of the claim severity (Eling and Toplek, 2009). Our finding implies that the size of such a catastrophic cyber loss might be uncontrollable for the private sector (e.g., for a single insurer).

Using a database of 477 organizations across 15 nations, the Ponemon Institute (2018) found that the financial service industry (e.g., banking) is most exposed to data breach risk and has one of the highest per capita costs (= $208 per breached record on average). If a catastrophic cyber loss occurred in the financial industry, the cost of claims imposed on a cyber-insurer might result in the insurer's insolvency with insurance runs. For example, if Wheatley et al.'s (2016) 300 million maximum breach size occurred, the economic damage estimated using the average cost for the financial industry (= $208) would be $62.4 billion, which is almost equivalent to the shareholders' equity of Allianz Group in the 1st quarter of 2019 (= $67.2 billion), one of the biggest cyber-insurers in the U.S. market.[73] Thus, the "too-big-to-fail" problem could exist under the current circumstances surrounding the cyber-insurance market, thereby requiring the consideration of potential cyber-reinsurance and possible backstop by the government in the form of deposit insurance, where the government plays the role of the last resort to guarantee a certain amount of the loss.[74] We call this cyber-deposit insurance.

Based on this idea, we design a cyber-deposit insurance that requires the government to be responsible for the loss amount above our PMCL estimate, which is the coverage limit of a proportional reinsurance treaty with a reinsurer.[75] The proportional reinsurance is based on

---

[73] The number of breached records, for example 300 million, is not necessarily the number of customers breached, but can be total records including cases where multiple customers lose more than two records from different types. Furthermore, the $208 cost for each record is an average figure, meaning the cost could vary by type of information breached at the individual level. Thus, $62.4 billion is a rough idea of the size of economic damage by a single cyber event to a banking entity that is a global-level player in the industry.

[74] The huge risk of change, underdevelopment of modeling and pricing schemes, and scarcity of data have played a significant role of preventing the reinsurance market from offering sufficient capacity, which might cause insurers to be reluctant to enter or withdraw from the cyber-insurance market. This circumstance can drive more demand for the establishment of state-backed programs (national level), such as the California Earthquake Authority (CEA), the Terrorism Risk Insurance Act (TRIA) in the U.S., and the National Flood Insurance Program (NFIP) from other catastrophe risk cases (Kunreuther, 2015). To consider this possibility, we design a government-involved reinsurance portfolio in this section.

[75] An extreme risk event is traditionally covered by both insurers and reinsurers together under a conventional two-stage process, where loss distributions are estimated using catastrophe modeling and calibration in the first stage and the negotiation of the price between insurers and reinsurers in the second stage (Chang and Chang, 2017). The linkage between extreme value theory and the reinsurance can be found in a specific type of contract, known as catastrophe excess-of-loss coverage (CatXL), as described in Embrechts, Mikosch and Kluppelberg (1997, p. 503) and Eling and Toplek (2009). The contract is a non-proportional form that requires a reinsurer to be responsible for the loss amount above a certain catastrophic level (retention level for the ceding company) up to a certain limit. Specifically, extreme value theory is connected to the pricing with regard to the reference loss defined as a value that helps determine a CatXL coverage level. The reference loss chosen characterizes

sharing a fixed percentage of the loss from a particular risk with the reinsurance market, called a quota share treaty. In this situation, the reinsurance market is responsible for its share $1-q$ of the amount up to the coverage limit while the insurer takes up the share q of the amount. In this portfolio design, we denote the coverage limit of the quota share treaty by U and the loss amount by L.[76] Figure 3 illustrates the portfolio structure for a potential cyber-insurer. The graph below presents the amount of loss that each party needs to take up in the portfolio.

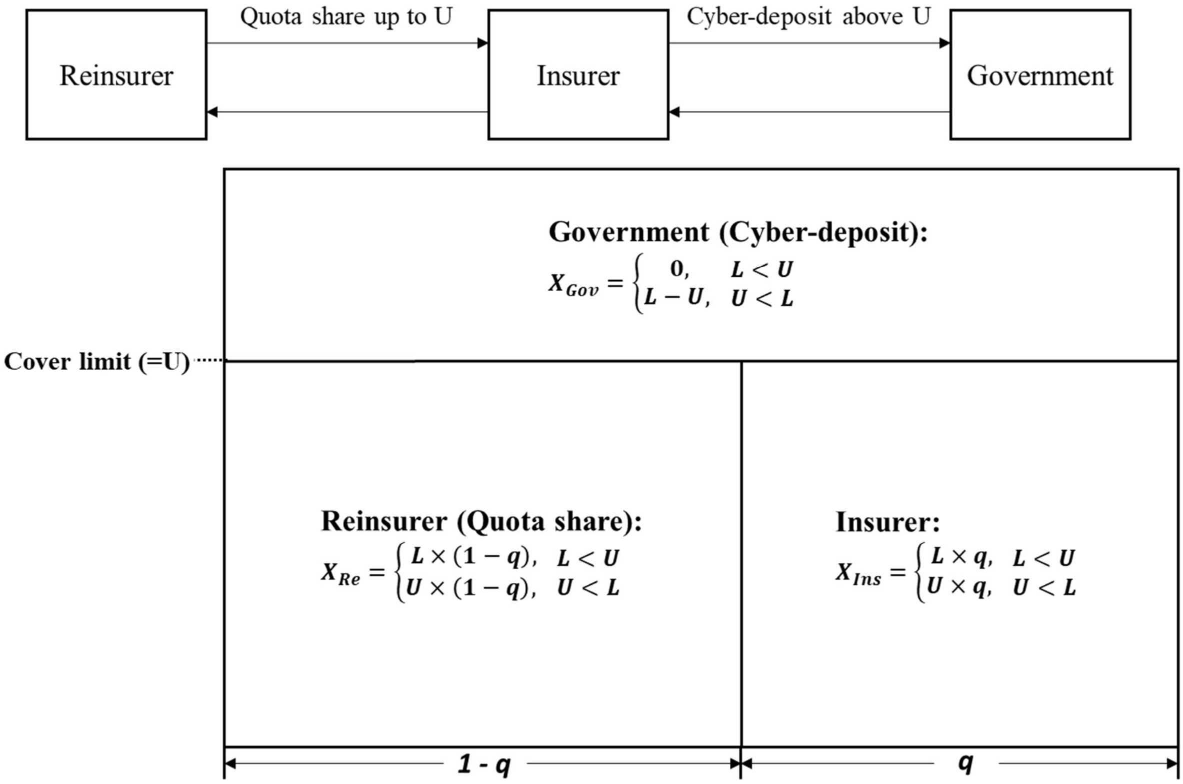


Figure 3. The structure of reinsurance portfolio for a cyber-insurer. The portfolio consists of a proportional reinsurance contract (quota share) with the reinsurance market and a non-proportional cyber-deposit insurance with the government above the coverage limit of the quota share treaty.

With the portfolio design in Figure 3, we calculate the aggregate premium size for the reinsurer and the insurer from a potential cyber risk pool by means of two premium principles with a loading factor: expectation premium and standard deviation premium, defined as (Embrechts, 2000):

rare, but possible, extreme loss, which is in line with our definition of probable maximum cyber loss. Here we utilize the CatXL design for the cyber-deposit insurance with the government.

76 The optimal reinsurance policy has been studied in the literature (see, e.g., Kaluszka, 2001, 2005; Gajek and Zagrodny, 2004; Guerra and Centeno, 2008). Among them, we utilize a combination of CatXL and quota share based on Embrechts, Mikosch and Kluppelberg (1997) with extreme value theory and Kaluszka (2005). Our reinsurance portfolio might not satisfy optimality, but it can provide insights for a possible reinsurance contract with the PMCL.

Expectation principle: $$H(X) = (1 + \delta) \cdot E(X), \quad (23)$$

Standard deviation principle: $$H(X) = E(X) + \delta \cdot \sqrt{Var(X)}, \quad (24)$$

where $H(\cdot)$ is a premium function of loss X and δ is a loading factor.[77] The aggregate premium calculated indicates the total amount that both the ceding company (insurer) and the ceded company (reinsurer) need to collect from the insureds in the pool for itself. The ceding company then pays the aggregate premium to the ceded company up to the agreed share ($= 1 - q$). The principles have been used for the optimal reinsurance problem in, for example, Kaluszka (2005) and Guerra and Centeno (2010). For the cyber-deposit insurance with the government, we provide the average size of loss per exposure that the government can potentially face.[78]

Table 4 shows the premium estimates for the reinsurer and the insurer and the estimated size of loss that the government can embrace. We derive the outcomes based on the PMCL estimates for the next one year (one-year contract) as the coverage limit (= U); for the sake of comparison, we calculate the measures using the estimates of maximum cyber loss from two references (Edwards et al., 2016; Wheatley et al., 2016). We find nearly \$627 million and \$209 million in premium sizes that the reinsurer and insurer need to collect, respectively, per month from their cyber risk pool in the U.S. based on the total loss data (= Comp) for the entire period. The projected annual premium (= \$10.03 billion) under the assumption of the simple arithmetic annualization is almost 60% larger than the value of cyber-insurance premiums worldwide for 2019 (= \$6.2 billion), as estimated in the industry (PwC, 2016). Thus, applying our approach to the estimation of PMCL can meet many industries' commercial needs for protection by extending the coverage limit, leading more insurers and reinsurers to enter the competition in underwriting cyber risk. If insurers are more confident in their underwriting competence due to

77 Romanosky et al. (2017) use a content analysis tool to investigate 100 cyber-insurance policies and show the current market situation and insurance designs. They find that 50% of the policies they examined provide a flat rate of premium equivalent to equation (23), which does not differentiate the premium rate by firm size or revenue. They also observe that the loading factor, δ, in equations (23) and (24) proportionally increasing the premium level is usually set between 25% and 35%, which represents the transaction cost (administration cost) for underwriting cyber-insurance. Following them, we assume a loading of 0.25, which is applied to the premium calculation for the reinsurer and the insurer in the portfolio for simplicity. Romanosky et al. also observe that all considered policies set the deductible and the coverage limit in accordance with company-specific parameters, implying that full coverage does not exist in the market. They find it surprising that many policies are priced by means of a simple, flat pricing scheme—that is, an expectation principle with a loading.

78 A typical deposit insurance in the banking sector is priced by a flat-rate system, imposing a given rate per unit of deposits to all insureds. However, the flat-rate premium cannot resolve the moral hazard problem inherent in the provision of deposit insurance. In other words, deposit insurance can reduce the cost of pursuing riskier strategies and incentivize excessive risk-taking. Although a risk-based pricing scheme has been developed to resolve this issue, in this application we do not implement any pricing scheme of deposit insurance, but rather provide an empirical benchmark on the average loss that the government needs to bear.

sufficient data and accumulated knowledge, our approach can scale up the capacity of the cyber-insurance market with higher limits.

This increase in the capacity needs to be backed by the government for a case when the possibility of a loss beyond our PMCL estimates might still exist. This possibility could be realized in a situation where a tremendous range of parties around the globe are affected by a malicious attack. In addition, this possibility becomes more feasible as internet-based technology evolves more, as already evidenced by the huge increase in the size of our PMCL estimates from pre-2014 to post-2014. Our results indicate that the government needs to take up on average $2.1 billion in losses per exposure beyond our PMCL estimate based on the total loss data (= Comp) for the entire data period.

Table 4. Reinsurance Pricing Results

Panel A: Aggregate premium size for the reinsurer and the insurer (on a monthly basis)

		Expectation principle			Standard deviation principle		
($ million)		Comp	Mal	Neg	Comp	Mal	Neg
Reinsurer	Entire	627.01	768.20	112.73	813.60	1,012.63	124.01
	Pre-2014	248.88	231.42	100.35	258.18	283.65	101.56
	Post-2014	2,212.40	4,316.11	150.21	4,344.84	6,773.16	184.50
	Reference 1	915.12	895.18	572.36	1,350.71	1,264.95	1,008.29
	Reference 2	1,436.40	1,291.78	886.52	2,390.52	2,125.88	1,707.68
Insurer	Entire	209.00	256.07	37.58	271.20	337.54	41.34
	Pre-2014	82.96	77.14	33.45	86.06	94.55	33.85
	Post-2014	737.47	1,438.70	50.07	1,448.28	2,257.72	61.50
	Reference 1	305.04	298.39	190.79	450.24	421.65	336.10
	Reference 2	478.80	430.59	295.51	796.84	708.63	569.23

Panel B: The average size of loss per exposure for the government

($ million)		Comp		Mal		Neg	
Govern-ment	Entire	2,091.21	(17,669.3)	1,959.31	(20,945.6)	1,285.63	(11,296.5)
	Pre-2014	486.94	(3,197.2)	200.88	(2,182.0)	210.58	(1,672.3)
	Post-2014	282.05	(10,233.0)	2,235.36	(25,920.6)	698.41	(8,953.4)
	Reference 1	1,783.89	(16,564.4)	1,823.87	(20,351.2)	795.35	(9,565.1)
	Reference 2	1,227.85	(14,188.7)	1,400.83	(18,301.8)	460.26	(7,834.9)

Note: References 1 and 2 indicate the outcomes based on the estimates of maximum loss by Edwards et al. (2016) (= 130 million records breached) and Wheatley et al. (2016) (= 300 million records breached), respectively. The parameters assumed for pricing are listed in the following: $\delta = 0.25$; $q = 0.25$; deductible for individual exposure = $500,000. The coverage limits are set by the PMCL estimates likely to happen in the next one year (see Table 3) and diversified depending on the data period and the type of risk. The figures in panel B indicate the mean and standard deviation (in parentheses) of the losses per exposure beyond our PMCL estimates. Appendix F provides the specification of the design in more detail.

The result further illustrates that differentiating the coverage limits depending on the risk type is significant for estimating the size of the premium. The premium level covering only the malicious risk is much higher (14.3 times higher on average from both expectation and standard deviation principles) than the level covering negligent risk. Both risk types are covered in the current cyber-insurance market, falling into specific property-casualty lines of business (Romanosky et al., 2017). In contrast, despite different time spans for the coverage limit between our estimate (next one year) and the estimates from the references (next three years and five years, respectively), the difference between the two risk types significantly decreases to 1.38 times larger for the malicious risk when adapting identical coverage limit from the two references to heterogeneous risk types.

In addition, the premium sizes based on the post-2014 database are almost 11.9 times larger on average than those based on the pre-2014 database across risk types and premium principles. This finding suggests that there is a need to establish a risk-adjusted pricing scheme for cyber risk, particularly considering time-varying risk dynamics of extreme cases. If the market stays in a classical manner with long-range historical data for pricing, it might heavily underestimate the premium size for cyber risk. Moreover, adapting the volatility of the losses into pricing increases the premium level in general, as shown by the deviation between the expectation principle and the standard deviation principle.

Discussion of public intervention

The proposed design in Section 5.1 consists of three layers[79] to deal with potential cyber losses. The first layer is taken by the primary insurer, who collects the premium from (organizational) insureds in the cyber risk pool. The second layer then goes to a reinsurer with an agreed coverage limit based on our statistical estimation and a contractual quota. The government is responsible for the last layer, as it absorbs losses above the coverage limit of the reinsurance contract. However, one might question whether public intervention is essential against a catastrophic cyber event that might have to be covered solely by the private sector. To answer this question, the suggested three-layer scheme can be more realistically applied to the banking industry, where customer information/records might be monetarily more valuable than those in, for example, the retail industry, thereby causing a severer systemic risk for society. In this case,

[79] A similar form as our three-layer program can be found in examples that the French government established a pool for terrorism coverage in 2002 (Kunreuther, 2002) and the Terrorism Risk Insurance Act (TRIA) passed by the U.S. Congress in November 2002, which provides a federal insurance backstop for U.S. property and casualty insurers against terrorism risk (Brown et al., 2004). Apart from the U.S. case, a number of countries are currently operating the three-layer government-backed program particularly against terrorism risk (e.g., Pool Re in the U.K., Terrorism Reinsurance & Insurance Pool [TRIP] in Belgium, and Dutch Terrorism Reinsurance Pool [NHT] in the Netherlands). In addition, some studies in the catastrophe risk context also describe that the government can provide a guarantee of the last resort (see, e.g., Brown et al., 2004; Smetters, 2005; Charpentier and Le Maux, 2014; Kunreuther, 2015).

public intervention against the extreme cyber event could be considered as a form of deposit insurance by the government with a premium in return.

In our design for the three-layer program, the premiums and the size of loss taken up by the government clearly rely on the level of the coverage limit, which in our case is defined as the PMCL estimate. This dependence is illustrated in Figure 4, where plots also play the role of a robustness check by conducting sensitivity analyses on any change in premiums and the average loss per exposure (for the government) above the coverage limit. The higher the coverage limit, the more the reinsurer and the insurer should be paid by potential insureds and the less the government needs to absorb. Organizations willing to purchase cyber-insurance could be more burdened to pay the premium as they face a larger loss as the network environment becomes more complicated, which can be observed by the huge difference between pre-2014 and post-2014 in Table 3.

However, providing the financing guarantee from the government enables an insurer to increase the insurability of cyber risk and makes it more feasible to offer a reasonably priced policy to insureds. Thus, a social discussion between (re)insurers and responsible government parties is encouraged to agree on the limit level to determine the size of financial backstop (deposit insurance) by the government as well as the corresponding premium size to be paid to the government. However, prior to this discussion, a certain amount of funds might be required for dragon king-size cyber claims, particularly in order to avoid the insurer's insolvency. These funds should be regulated so as to be available only for extreme cyber claims, which can prevent the insurer from cross-subsidizing catastrophic claims from other lines (Jaffee and Russell, 1997).

An alternative route to optimize public intervention against cyber risk could be a comprehensive offer for cyber risk management by an insurer already incorporated in some current cyber policies; such an offer can reduce the need for financing future loss by the public sector. This type of offer can include a prevention/protection measure implemented by the collaboration between the insurer and an IT security firm and a warning system/process with the interaction between the insured and the insurer. In addition, the government can construct a security system as a public good for industries, which can serve as boundary protection for companies. Another possible design is a contract with industry loss warranties (ILWs), especially for our dataset, which incorporates information on industry-level losses in the U.S.[80] ILWs are particularly relevant when a limited supply capacity in the reinsurance and retrocession markets against extreme events exists (Gatzert and Schmeiser, 2012); hence, it could be an optimal industry-level solution for managing extreme cyber losses.

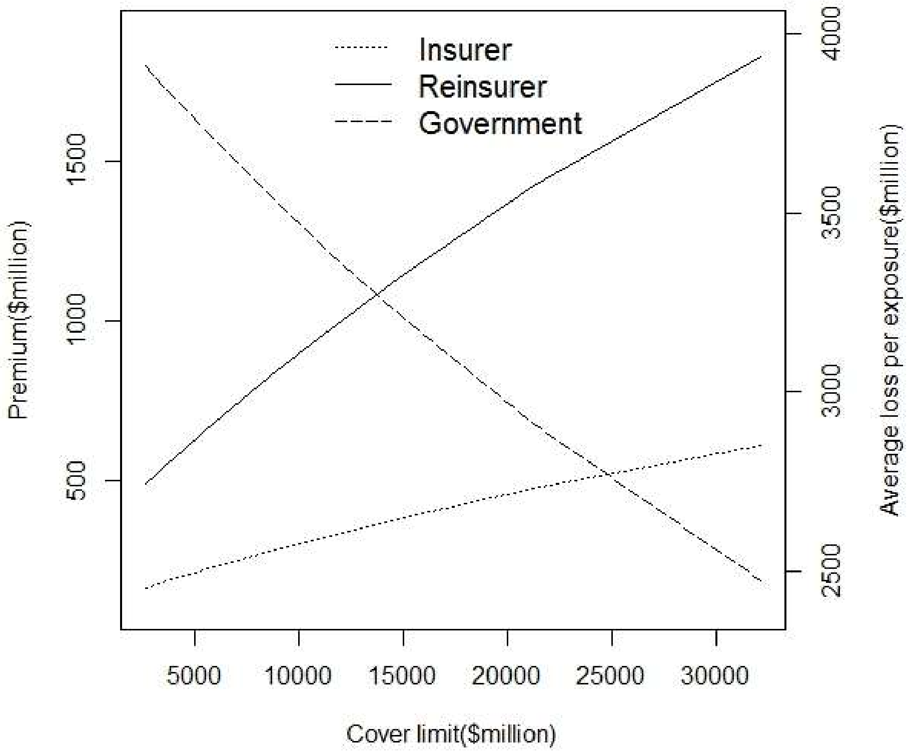


Figure 4. Sensitivity analysis of premiums and the average loss per exposure over the coverage limit with the data of the composite risk type for the entire period. Insurer and reinsurer lines display the premium estimates read on the left axis, which consistently describes the upward movement over increasing coverage limits. The government line illustrates the average loss per exposure read on the right axis, showing the downward movement over increasing coverage limits.

6 Conclusion

Regulators might have to be concerned about how to prevent insurers from becoming insolvent due to an uncontrollable cyber disaster triggering a huge accumulation risk. To set a capital requirement on cyber risk, they usually implement a threshold-based method to estimate the size of an extreme event based on what most literature suggests. Although the estimation procedure of an extreme cyber loss in the literature is well constructed and turns out to be robust, a possible catastrophic event much beyond the measure by this procedure is still likely to occur. Such an event, called a dragon king (Sornette and Ouillon, 2012) or black swan (Taleb, 2007),

[80] Industry loss warranties (ILWs) provide a fixed indemnity in case an industry-wide loss surpasses an agreed-upon threshold level, which varies by location, type of a loss event, line of business, and duration (Gatzert and Schmeiser, 2012). For more details on contract design and pricing approach, see Gatzert and Schmeiser (2012).

could lead to unpredictable level of cyber loss across multiple parties, resulting in a significant societal cost. This paper studies how to systematically estimate this type of event based on an empirical database (PRC) widely used in the cyber risk context. We model the cyber loss maxima in three different timeframes (weekly, biweekly, and monthly) using time series analysis, extreme value distribution fitting, and extreme value copula modeling. This procedure is developed to estimate probable maximum cyber loss (PMCL), which accounts for the worst cyber loss likely to happen once every specified time unit. The analysis is conducted for three data periods (entire period, pre-2014 and post-2014), with a structural break between pre-2014 and post-2014 identified by the graphical diagnosis and the statistical tests.

We identify that the maxima series of cyber loss are stationary and serially correlated, particularly short-range time series (weekly and biweekly). Cyber loss maxima series follow the Fréchet type of generalized extreme value distribution, demonstrating the heavy-tailedness of the cyber loss maxima. In addition, two risk categories of cyber loss, malicious and negligent, tend to be extremely dependent for the longer time period (monthly), implying that the maxima series of the two risk factors are dependent, especially in the extreme upper tail. This estimation leads to the findings in the application to probable maximum cyber loss that our estimates likely to happen in the next five years are around seven times larger than the expected maximum size (= 300 million breaches) predicted by Wheatley et al. (2016) with a widely used Pareto-based estimation. This difference implies that a threshold-based (Pareto) estimation might significantly undervalue the risk size and lead insurers to set a substantially low coverage limit, thereby keeping organizations from deciding on cyber-insurance. This implication is supported with our dataset by the comparison between the PMCL estimates and those based on the POT method, showing that the PMCL estimates are on average 3.46 times larger.

We further find that the loss by a malicious attack could be much more extreme than that by a negligent risk, showing around 28.5 times larger for the entire data period. This result can be reasonably interpreted by the fact that a single business entity is usually affected by negligent risk, whereas multiple entities can be simultaneously affected by malicious risk with an infectious attack. Thus, a malicious case is more likely to cause a silent cyber risk in numerous P&C policies than a negligent case. Therefore, we propose that a cyber-insurer needs model and price cyber risk by classifying the risk factors upon the market demand. Furthermore, the PMCL estimates based on the post-2014 database are on average 433 times larger than those based on the pre-2014 database, thereby indicating that the cyber risk landscape has been changing quickly over time with rapidly developing information technology and explosively expanding interconnected network environments. Thus, insurers must develop an underwriting process that can capture rapidly changing risk dynamics over time.

Applying PMCL estimates, we propose a possible risk transfer design with public intervention, where a primary insurer agrees on a quota share treaty with a reinsurer and a CatXL-based deposit insurance with the government is provided against losses above the coverage limit of the reinsurance contract. Our proposal leads to the findings that, with two premium principles (expectation and standard deviation), the aggregate premium level estimated for both the insurer and the reinsurer is already 60% larger than the value of cyber-insurance premiums worldwide for 2019 estimated in the industry. Thus, the current capacity of the cyber-insurance market is heavily limited, with a significantly low coverage limit. In addition, the government might encounter on average $2.1 billion claims per exposure beyond the coverage limit (PMCL) in the cyber-deposit insurance based on the total loss of data for the entire data period.

It seems obvious that extreme cases of cyber loss are penetrating our daily life and business. Current and potential cyber-insurers are more likely to face the possibility of extreme claims that might not be controllable. In this study, we suggest a first step for preventing a possible disastrous situation by providing extreme cyber risk pooling with a statistical process and the extreme risk measurement, probable maximum cyber loss (PMCL). The results of the paper are important for risk managers and actuaries designing cyber-insurance policies in primary insurers and reinsurers and for policymakers concerned about the social cost of the next extreme cyber loss. However, the dynamic nature of cyber risk as a significant risk of change can make the historical data less important for future predictions; therefore, (re)insurers need to develop a dynamic pricing/reserving method as more events occur. Furthermore, it would be fruitful to improve the estimation by identifying the relevant factors potentially affecting the extreme cyber cases in the model. This approach could explain the limitation of this paper as this study lacks a scenario-based analysis, thereby possibly showing the differences in extreme cases between scenarios.

Appendix A. Overview of literature and comparison with the present paper

Table A1. Summary of Literature about the Statistical Analysis on Extremes of Cyber Risk

		MS10	EHF16	WMS16	EJ18	EW19	HWS19	Present paper
Modeling perspective	Focus of Study	Statistical modeling of cyber risk	Statistical modeling of cyber risk	Statistical modeling of cyber risk	Dependence modeling of cyber risk	Statistical modeling of cyber risk	Statistical modeling of cyber risk	Statistical modeling of cyber risk
	Time variation	NA	NA	NA	NA	NA	NA	**Weekly, biweekly & monthly**
	Data type & period	Open Security Foundation (breached loss)	PRC data (breached loss)	Open Security Foundation & PRC (breached loss)	PRC data (breached loss)	SAS database (operational data)	Open Security Foundation & PRC (breached loss)	PRC data (breached loss)
		Jan 2000 – Nov 2008	Jan 2005 – Sept 2015	Jan 2007 – Apr 2015	Jan 2005 – Dec 2016	Jan 1995 – Mar 2014	Jan 2007 – Sept 2017	Jan 2005 – Dec 2018
	Methodology	Power law tail distribution	Lognormal and negative binomial distributions	POT with double truncated-Pareto	Collective risk model & *d*-dimensional copula method	POT and GLM (dynamic EVT model)	POT with lognormal and truncated-Pareto	**Temporal dependency, block maxima & extreme value copula method**
	Dependency modeling	NA	NA	NA	Multi-dimensional model (vine copula) for cyber loss	NA	NA	**Extreme value copulas**
Outcome	Main points	· Identify statistical properties of cyber risks by looking at power-law tail distribution (personal identity losses) · Find the existence of a size effect that the largest possible ID losses per event grow faster-than-linearly with the organization size.	· Fit the severity of data breach losses with lognormal distribution and the frequency with negative binomial. · Predict the likelihood of a cyber loss occurrence above a certain level in the next three years based on their model.	· Model the severity of personal data breaches using extremely heavy-tailed truncated Pareto distribution. · Find the tail exponent parameter decreasing over time, meaning that the size is heavier over time.	· Classify information on data breach events into different cross-sectional settings and analyze it in different dependent structures. · Show how different high-dimensional dependence constructions influence on cyber-insurance premiums and risk measures.	· Identify cyber losses from operational risk database and apply actuarial models to analyze extreme cyber losses. · Identify the impacts of covariates (country, industry, size and so on) on the loss using dynamic EVT method with heavy-tailed distributions.	· Complement the results from Wheatley et al. (2016) by using updated dataset by September 2017. · Fit the frequency of bigger breaches with negative binomial and the severity with POT with truncated-Pareto.	· Investigate the statistical features of data breach losses from a variety of time blocks and model extreme dependency of cyber loss maxima. · Estimate the probable maximum cyber loss with loss maxima and propose a reinsurance portfolio with public-private partnership.
	Limitation & Implication	The dataset is not up-to-date (2000 – 2008) so that the current trend is not reflected.	It might have reached a meaningful finding if the study had considered some extreme value distributions.	EVT method used in the study might not be able to capture a possible severer event beyond the usual degree of freedom.	A comprehensive analysis of the dependence structure of data breach risks is suggested.	The data are derived from the operational risk database with a certain classification standard.	Their suggestion on insurance pricing scheme is not supported by any empirical estimation.	Modeling cyber loss maxima while considering time variation is critical to figure out cyber loss dragon king.

Note: The references introduced in the table are specified in the following. MS10: Maillart and Sornette (2010); EHF16: Edwards, Hofmeyr and Forrest (2016); WMS16: Wheatley, Maillart and Sornette (2016); EJ18: Eling and Jung (2018); EW19: Eling and Wirfs (2019); Hofmann, Wheatley and Sornette (2019). The bold indicates the contributing points of the present paper to the literature.

Brief description on the literature

The studies in Table A1 are more specifically described here. Maillart and Sornette (2010) estimate a power-law tail distribution to fit the data of personal identity losses, thereby concluding that the extremely right-skewed cyber losses are identified. Edwards et al. (2016) investigate the distributional properties of cyber loss frequency and severity using a PRC dataset between 2005 and 2015. They find that negative binomial fits well the frequency of cyber loss and lognormal distribution for the severity. The authors further estimate the loss probability of the largest breach size (130 million) in the next three years to be around 16%. Using a larger amount of data on personal ID breaches than that of Maillart and Sornette (2010), Wheatley, Maillart and Sornette (2016) fit an extremely heavy-tailed truncated Pareto distribution by finding a decreasing tail parameter over time (from 2007 to 2015). They identify that the possible maximum size of cyber loss could be 200 million breaches at a growth rate of $t^{0.84}$.

Eling and Jung (2018) identify lognormal, Burr and the Peaks-over-Threshold (POT) with lognormal in the body and GPD tail in the tail as the best fit for the monthly severity of cyber risk using the data from Privacy Rights Clearinghouse (PRC) between 2005 and 2016. Eling and Wirfs (2019) use an operational risk database from SAS to derive cyber-related operational losses and analyze them using the POT technique with generalized Pareto distribution (GPD) above the threshold. Lastly, Hofmann et al. (2019) conduct distribution fitting for the frequency and severity of large breaches in a similar manner as Wheatley et al. (2016), using an updated dataset through September 2017. The authors determine that the frequency and severity of large breaches are growing over time, and there exists a huge gap (10 times) between the predicted risk size of the recent period and that of the older period.

Appendix B. Test for structural break and extreme events over last decade

As can be observed in Figure 1, structural breaks in frequency and severity trends might exist. To examine this possibility, we conduct three statistical tests that can prove the presence of a break: OLS-based cumulative sum test (OLS-CUSUM), recursive cumulative sum test (Rec-CUSUM), and Chow test.[81] Figure B1 highlights the graphical description of structural breaks in frequency and severity trends and the trends from two distinct periods. The test results in Table B1 confirm that the clear distinctions in frequency and severity between the two periods are presented by rejecting the null hypothesis. In particular, as seen on the right plot of Figure B1, a significant increase in the severity of cyber loss has taken place since 2014, showing that almost 0.1 million breaches increase every 50 days.

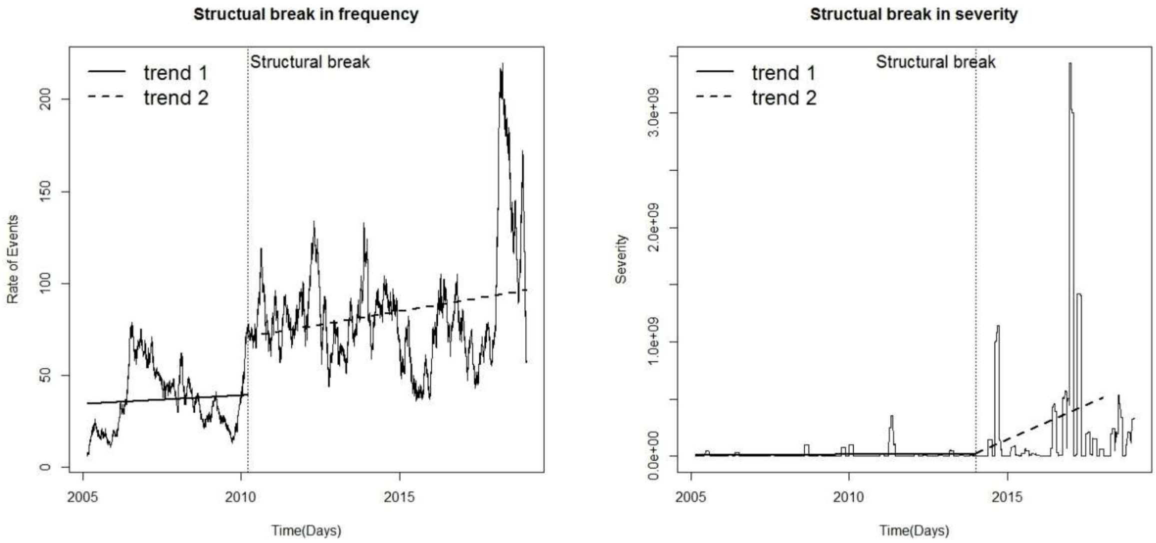


Figure B1. Structural breaks and trends of frequency and severity of cyber risk over the period between 2005 and 2018. Rolling windows for 50 days are applied to both data to see the cluster.

Table B1. Testing Structural Breaks in Frequency and Severity Trends

	Structural break	Test			Trend	
		OLS-CUSUM	Rec-CUSUM	Chow	Intercept	Slope
Frequency	Mar 2010	10.853***	7.443***	1,203.289***	1: 2.30 2: -45.38	1: 0.003 2: 0.008
Severity	Jan 2014	5.890***	3.851***	73.059***	1: -42m 2: -1.9b	1: 3,888 2: 0.1m

Note: Tests are implemented via `R` package, `strucchange`. m and b stand for million and billion respectively. *, ** and *** indicate that the p-value is less than the significance levels, 10%, 5% and 1%, respectively.

[81] These tests are designed to examine the null hypothesis of no structural change, $H_0: \beta_i = \beta_0$, where $\beta_i, i = 1, \dots, n$, is the vector of regression coefficients and the alternative hypothesis is that the coefficient vector varies over time. For more detail on the test specifications, see Kleiber et al. (2002).

Extreme events over the last decade

Some notoriously extreme losses (above 100 million breached records) due to cyberattacks (hacking) shown in Table B2.[82] have occurred simultaneously in different geographical regions and network systems (multiple entities in the column of breached entity), accounting for 0.3% of the total number of losses (= 6,780). Furthermore, most extreme losses have arisen over last five years (16 out of 20 events in Table B2), especially in the general business field, implying that such maximum losses are more likely to occur and disrupt businesses in the near future. This trend is supported by Figure B2, where the frequency of cyber losses soared in 2010 and has remained stable so far. Therefore, we can pose a question as to whether much severer losses cluster over time and, if so, how they could be modeled and predicted. This analysis is important not only for risk managers in primary cyber-insurers, but also for reinsurers looking for an opportunity to be involved in the cyber-insurance market. Moreover, it supports our analysis for post-2014 cyber losses, which can reflect the recent trend of loss severity in the fast-changing technological environment.

Table B2. List of Extreme Cyber Losses from 2005 to 2018

Date	Breached entity	Risk type	Industry	Breach records (million)
Dec 14, 2016	Yahoo	HACK	Business	3,000.0
Mar 8, 2017	Multiple entities	DISC	Business	1,370.0
Aug 5, 2014	Multiple entities	HACK	Business	1,000.0
Sep 22, 2016	Yahoo	HACK	Business	500.0
Nov 16, 2016	FriendFinder	HACK	Business	412.0
May 31, 2016	MySpace	HACK	Business	360.0
Jul 3, 2018	Exactis	DISC	Business	340.0
Nov 30, 2018	Marriott International	HACK	Business	327.0
Apr 2, 2011	Epsilon	HACK	Business	250.0
Jun 19, 2017	DeepRootAnalytics	DISC	Business	198.0
Dec 28, 2015	Multiple entities	DISC	Business	191.0
Jun 6, 2012	LinkedIn	HACK	Business	167.0
Mar 30, 2018	Under Armour	HACK	Business	150.0
Sept 7, 2017	Equifax	HACK	Financial service	145.5
May 21, 2014	eBay	HACK	Business	145.0
Jan 20, 2009	Multiple entities	HACK	Financial service	130.0
Jun 27, 2018	NameTests	DISC	Business	120.0
May 17, 2016	LinkedIn	HACK	Business	117.0
Oct 11, 2018	MindBody - FitMetrix	DISC	Business	113.5
Apr 27, 2011	Sony	HACK	Business	101.6

Note: The list is based on the data from PRC and sorted by the loss amount (breached records). In terms of risk type, HACK stands for a hacking risk (cyberattack) and DISC indicates an unintended disclosure of private data. In the industry column, "business" incorporates any type of business entity apart from retail/merchant including online retail and financial service providers, whereas "financial service" contains information from the banking and insurance sectors.

[82] The losses in Table B2 indicate data breached records that occurred in the U.S. over the last 14 years. Other types of losses, for example direct economic loss by a malicious hacking, are not counted in this dataset.

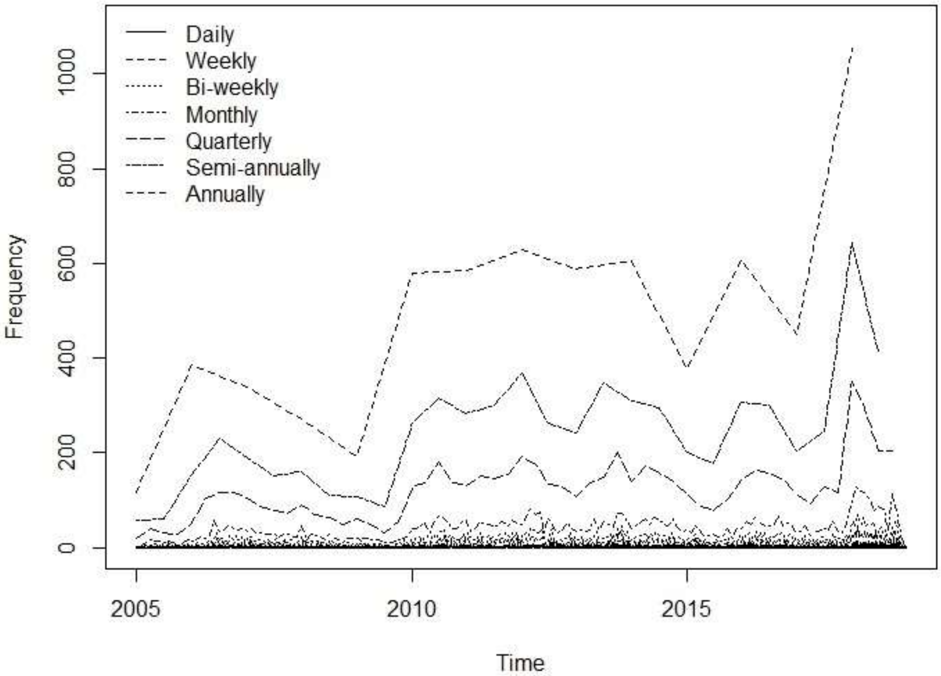


Figure B2: The plot of frequency in different time blocks.

Appendix C. Time-series analysis

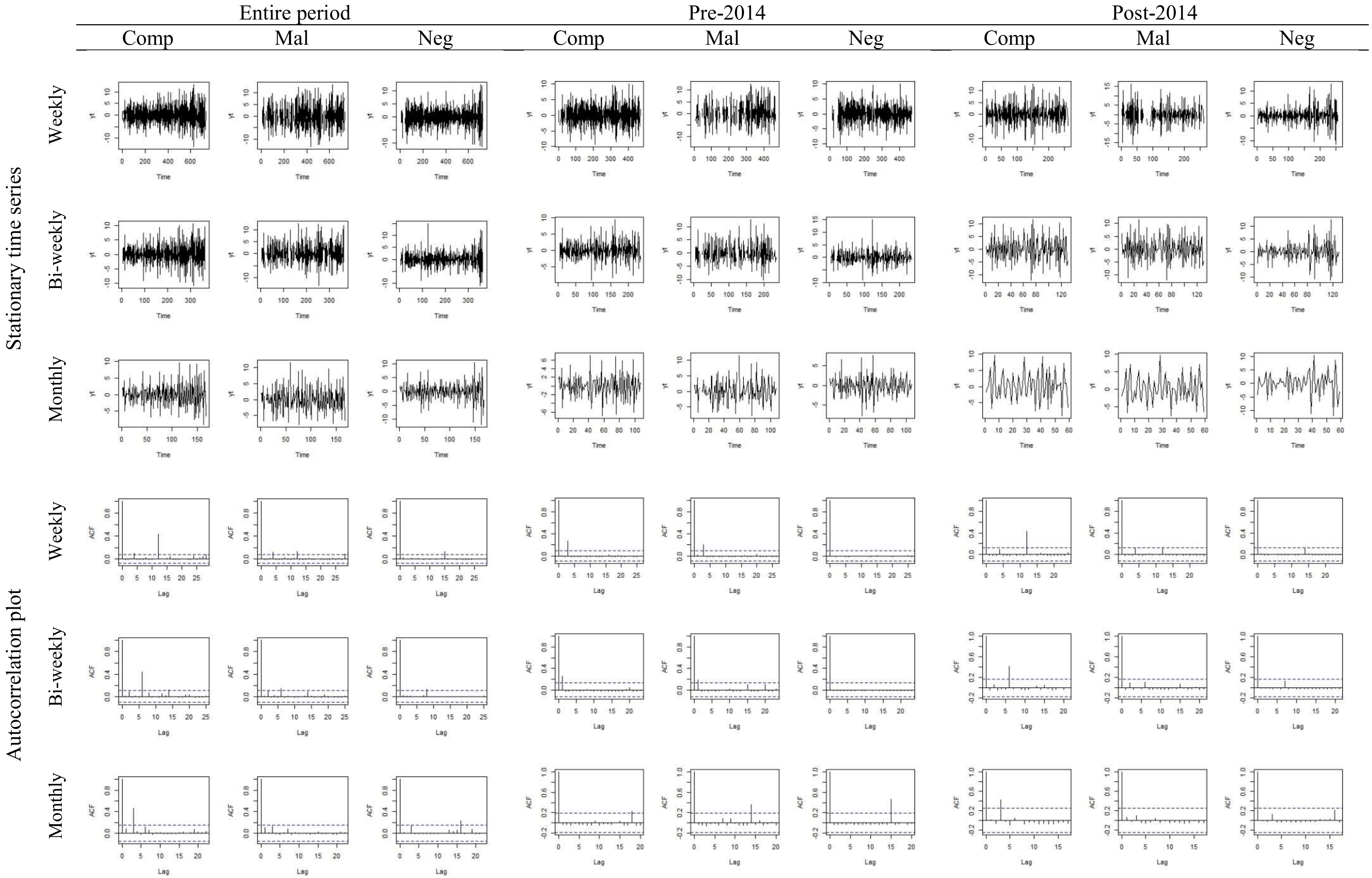


Figure C1: Time-series and autocorrelation plots for the entire period.

Table C1. Testing Results for Stationarity

		Complete		
		Weekly	Biweekly	Monthly
Entire	ADF	-8.76***	-4.86***	-4.89***
	PP	-749.85***	-377.90***	-184.09***
	KPSS	0.109	0.105	0.114
Pre-2014	ADF	-7.37***	-5.80***	-4.90***
	PP	-522.13***	-163.37***	-105.20***
	KPSS	0.071	0.067	0.068
Post-2014	ADF	-6.07***	-2.98	-2.74
	PP	-264.88***	-137.82***	-62.39***
	KPSS	0.113	0.110	0.109
		Malicious		
		Weekly	Biweekly	Monthly
Entire	ADF	-8.36***	-6.29***	-5.32***
	PP	-764.81***	-385.40***	-157.59***
	KPSS	0.055	0.054	0.056
Pre-2014	ADF	-7.46***	-6.00***	-5.12***
	PP	-499.24***	-181.72***	-103.67***
	KPSS	0.058	0.054	0.050
Post-2014	ADF	-5.80***	-3.94**	-3.52**
	PP	-269.03***	-139.82***	-55.48***
	KPSS	0.089	0.084	0.081
		Negligent		
		Weekly	Biweekly	Monthly
Entire	ADF	-8.73***	-6.05***	-5.34***
	PP	-715.13***	-355.81***	-174.45***
	KPSS	0.111	0.111	0.116
Pre-2014	ADF	-7.85***	-6.03***	-4.86***
	PP	-465.72***	-232.53***	-106.10***
	KPSS	0.082	0.084	0.084
Post-2014	ADF	-6.16***	-4.85***	-3.62**
	PP	-253.06***	-130.16***	-61.39***
	KPSS	0.051	0.052	0.053

Note: *,** and *** indicate that the *p*-value is less than the significance levels, 10%, 5% and 1% respectively.

Table C2. Results of Fitting Autoregressive Model

Panel A: Fitted autoregressive model (Full series)

		Entire period			
Data	Block	Model	AIC	BIC	AICc
Complete	Weekly	AR(12)	9086.33	9150.65	9086.84
	Biweekly	AR(6)	4797.92	4829.14	4798.23
	Monthly	AR(3)	2328.69	2344.31	2328.94
Malicious	Weekly	AR(4)	9094.00	9121.57	9094.08
	Biweekly	AR(2)	4806.25	4821.86	4806.32
	Monthly	AR(0)	2336.41	2342.66	2336.44
Negligent	Weekly	AR(0)	7887.98	7897.17	7887.99
	Biweekly	AR(0)	4202.95	4210.76	4202.96
	Monthly	AR(0)	2060.41	2066.66	2060.44

Panel B: Fitted autoregressive model (Pre- and Post-2014 series)

		Pre-2014				Post-2014			
Data	Block	Model	AIC	BIC	AICc	Model	AIC	BIC	AICc
Complete	Weekly	AR(3)	3777.47	3798.23	3777.55	AR(12)	3527.60	3577.50	3529.07
	Biweekly	AR(1)	2049.85	2060.23	2049.90	AR(6)	1861.52	1884.52	1862.43
	Monthly	AR(0)	1019.63	1025.00	1019.67	AR(3)	898.25	908.72	898.98
Malicious	Weekly	AR(3)	3919.96	3940.73	3920.05	AR(0)	3513.88	3521.01	3513.90
	Biweekly	AR(1)	2120.70	2131.08	2120.75	AR(0)	1853.88	1859.63	1853.91
	Monthly	AR(0)	1051.01	1056.38	1051.05	AR(0)	895.84	900.03	895.91
Negligent	Weekly	AR(0)	2894.64	2902.94	2894.65	AR(0)	3084.56	3091.69	3084.58
	Biweekly	AR(0)	1609.92	1616.84	1609.94	AR(0)	1639.14	1644.89	1639.17
	Monthly	AR(0)	823.60	828.97	823.64	AR(0)	798.36	802.55	798.43

Panel C: Test for conditional volatility (only for autoregressive series)

	Entire period				
	Complete			Malicious	
# Lags	Weekly	Biweekly	Monthly	Weekly	Biweekly
Lag=4	0.712	0.307	0.172	0.685	0.283
Lag=8	0.726	0.656	0.266	0.695	0.488
Lag=12	1.554	1.432	0.329	1.176	0.511
Lag=16	1.578	1.554	0.394	1.186	0.559
Lag=20	1.594	1.574	0.445	1.199	0.580
Lag=24	3.031	1.609	0.483	1.209	0.605

	Pre-2014				Post-2014		
	Complete		Malicious		Complete		
# Lags	Weekly	Biweekly	Weekly	Biweekly	Weekly	Biweekly	Monthly
Lag=4	0.137	0.076	0.659	0.317	0.200	0.097	0.066
Lag=8	0.196	0.123	0.705	0.437	0.244	0.211	0.178
Lag=12	0.223	0.178	0.768	0.552	0.484	0.409	0.353
Lag=16	0.244	0.233	0.824	0.651	0.526	0.470	0.552
Lag=20	0.272	0.275	0.876	0.758	0.565	0.559	0.749
Lag=24	0.300	0.328	0.941	0.864	0.985	0.654	0.951

Note: AR stands for autoregressive model, and AIC, BIC, and AICc for Akaike information criterion, Bayesian information criterion, and the corrected AIC, respectively. In panel B, the null hypothesis of the test is the presence of homoscedasticity in the error term. Panel C shows the results of Portmanteau-Q test for heteroscedasticity. *, ** and *** indicate that the *p*-value is less than the significance levels, 10%, 5% and 1%, respectively.

Table C3. Diagnostic Tests on the Residuals of the Fitted Models

		Entire period				
		Complete			Malicious	
		Weekly	Biweekly	Monthly	Weekly	Biweekly
Box-Pierce	Statistic	142.84***	73.41***	35.99***	10.03**	4.46
	P-value	<0.001	<0.001	<0.001	0.040	0.108
Ljung-Box	Statistic	145.54***	75.01***	37.07***	10.11**	4.51
	P-value	<0.001	<0.001	<0.001	0.039	0.105

		Pre-2014				**Post-2014**		
		Complete		Malicious		Complete		
		Weekly	Biweekly	Weekly	Biweekly	Weekly	Biweekly	Monthly
Box-Pierce	Statistic	34.06***	16.05***	18.99***	8.45***	50.54***	23.69***	11.08**
	P-value	<0.001	<0.001	<0.001	0.004	<0.001	<0.001	0.011
Ljung-Box	Statistic	34.43***	16.26***	19.19***	8.55***	53.29***	25.17***	12.05***
	P-value	<0.001	<0.001	<0.001	0.003	<0.001	<0.001	0.007

Note: *,** and *** indicate that the *p*-value is less than the significance levels, 10%, 5% and 1% respectively.

Appendix D. GEV fitting results

Table D1. Distribution Fitting Results (Entire period)

Panel A: GEV fitting results

		Statistics				Estimated parameter		
Data	Block	Loglik	AIC	K-S	A-D	Loc	Scale	Shape
Comp	Weekly	-9,714.6	19,435.2	0.030	0.802	14,527.2	33,853.7	2.272
	Biweekly	-5,378.3	10,762.7	0.035	0.567	66,749.2	143,848.1	2.115
	Monthly	-2,729.3	5,464.7	0.058	0.667	318,979.0	305,943.1	1.661
Mal	Weekly	-8,444.1	16,894.2	0.457***	296.64***	9.874	39.804	4.025
	Biweekly	-5,088.0	10,182.0	0.103***	7.785***	10,817.1	39,807.2	3.670
	Monthly	-2,652.3	5,310.7	0.043	0.459	132,562.1	354,270.3	2.636
Neg	Weekly	-8,681.8	17,369.6	0.035	1.626	4,970.6	10,172.7	1.922
	Bi-weekly	-4,777.6	9,561.3	0.036	0.429	17,671.6	33,863.1	1.817
	Monthly	-2,416.6	4,839.3	0.036	0.182	64,312.6	123,339.0	1.865

Panel B: Comparison with other distributions (AIC)

Data	Block	**GEV**	L-norm	Gamma	Cauchy	IG	Burr	GPD
Comp	Weekly	**19,435.2**	19,458.9	20,353.8	20,730.2	22,679.2	31,379.5	19,456.8
	Biweekly	**10,762.7**	10,814.0	11,229.7	11,395.9	12,625.8	17,987.2	10,784.4
	Monthly	**5,464.7**	5,503.7	5,665.2	5,776.0	5,473.7	5,920.2	5,480.1
Mal	Weekly	16,894.2	16,819.3	17,453.1	19,599.5	19,402.2	24,842.2	**15,665.4**
	Biweekly	10,182.0	10,063.5	10,343.0	11,033.2	11,965.8	15,957.8	**9,780.3**
	Monthly	**5,310.7**	5,398.2	5,415.3	5,665.8	6,344.0	8,812.9	5,359.8
Neg	Weekly	**17,369.6**	17,434.9	18,285.4	18,391.1	20,242.0	27,201.5	17,410.0
	Biweekly	**9,561.3**	9,602.3	10,086.9	10,046.9	11,264.1	15,632.5	9,625.1
	Monthly	**4,839.3**	4,859.4	5,078.6	5,080.8	4,902.9	5,251.7	4,848.3

Panel C: GPD fitting results for weekly and bi-weekly malicious series

Block	Loglik	AIC	K-S	Theshold	Scale	Shape
Weekly	-7,830.6	15,665.4	0.000	0.99	6,661.3	3.033
Bi-weekly	-4,888.1	9,780.3	0.016	0.99	26,831.8	2.971

Note: *, ** and *** indicate that the p-value is less than the significance levels, 10%, 5% and 1%, respectively. K-S and A-D stand for Kolmogorov-Smirnov test and Anderson-Darling test, respectively, and Loc stands for the location parameter. In panel B, L-norm, IG, and GPD stand for lognormal, inverse Gaussian, and generalized Pareto distribution, respectively. Note that there is no location parameter in panel C for GPD fitting because the mean of GPD distribution in the tail is theoretically infinite (Maillart and Sornette, 2010).

Table D2. Distribution Fitting Results (Pre- and Post-2014 series)

Panel A: GEV fitting results

		Pre-2014							Post-2014						
		Statistics				Estimated parameter			Statistics				Estimated parameter		
Data	Block	Loglik	AIC	K-S	A-D	Loc	Scale	Shape	Loglik	AIC	K-S	A-D	Loc	Scale	Shape
Comp	Weekly	-6,015.5	12,037.0	0.052	1.834	13,333.1	26,773.1	1.891	-3,660.3	7,326.7	0.060	1.422	21,737.7	55,029.1	2.517
	Biweekly	-3,263.0	6,532.0	0.044	0.384	55,127.6	91,119.2	1.547	-2,059.7	4,125.5	0.063	0.755	94,397.7	268,725.4	2.875
	Monthly	-1,624.6	3,255.2	0.044	0.185	199,985.4	285,713.4	1.576	-1,114.2	2,234.5	0.452***	20.18***	22,369.3	24,672.2	5.647
Mal	Weekly	-4,725.4	9,456.8	0.413***	258.32***	18,051.3	161,019.8	8.920	-3,406.9	6,819.8	0.280***	41.10***	18,532.7	131,906.4	7.117
	Biweekly	-3,077.4	6,160.7	0.147***	10.98***	4,022.4	16,328.0	4.054	-1,988.4	3,982.9	0.071	0.802	46,214.5	127,354.8	2.736
	Monthly	-1,585.6	3,177.2	0.080	0.591	60,882.8	143,610.6	2.287	-1,050.2	2,106.4	0.079	0.567	640,346.7	1,828,094.4	2.915
Neg	Weekly	-5,612.4	11,230.8	0.070**	3.695**	5,507.9	11,446.0	1.933	-3,070.7	6,147.3	0.071	1.899	4,805.4	8,715.6	1.745
	Biweekly	-3,051.2	6,108.5	0.043	0.351	23,078.8	38,682.9	1.504	-1,699.0	3,403.9	0.064	0.698	11,190.7	23,419.5	2.190
	Monthly	-1,528.3	3,062.6	0.047	0.186	82,423.4	127,506.6	1.438	-878.9	1,763.9	0.088	0.454	41,810.5	103,181.8	2.487

Panel B: Comparison with other distributions (AIC)

Data	Block	**GEV**	L-norm	Gamma	Cauchy	IG	Burr	GPD	**GEV**	L-norm	Gamma	Cauchy	IG	Burr	GPD
Comp	Weekly	**12,037.0**	12,122.0	12,443.7	12,658.1	14,142.8	19,418.8	12,187.1	**7,326.7**	7,371.5	7,669.8	7,973.3	8,280.7	11,978.1	7,336.0
	Biweekly	**6,532.0**	6,559.9	6,809.7	6,784.2	7,663.6	11,035.5	6,546.7	**4,125.5**	4,140.3	4,248.1	4,483.5	4,195.8	4,381.5	4,131.9
	Monthly	**3,255.2**	3,281.9	3,390.4	3,381.8	3,254.0	3,613.0	3,274.1	2,234.5	2,169.9	2,192.4	2,331.6	2,220.0	2,310.1	**2,168.6**
Mal	Weekly	9,456.8	10,002.5	10,342.9	11,870.8	11,390.3	14,390.3	**9,130.6**	6,819.8	6,754.1	6,975.7	7,581.6	7,938.0	10,429.1	**6,506.8**
	Biweekly	6,160.7	6,067.6	6,207.3	6,630.3	7,176.4	9,497.9	**5,845.2**	**3,982.9**	3,983.3	4,098.0	4,320.5	4,653.2	6,557.0	3,990.0
	Monthly	**3,177.2**	3,266.2	3,227.2	3,349.3	3,788.5	5,214.9	3,225.1	**2,106.4**	2,114.6	2,139.5	2,241.4	2,141.2	2,247.9	2,109.3
Neg	Weekly	11,230.8	11,132.4	11,424.8	11,834.4	13,087.3	17,409.2	**10,647.9**	**6,147.3**	6,229.7	6,677.9	6,501.2	7,009.4	9,902.6	6,195.2
	Biweekly	**6,108.5**	6,130.1	6,304.5	6,353.9	7,296.2	10,040.6	6,187.7	**3,403.9**	3,452.3	3,647.6	3,642.2	3,409.5	3,682.3	3,412.8
	Monthly	**3,062.6**	3,066.7	3,168.7	3,177.6	3,117.3	3,372.1	3,065.2	**1,763.9**	1,775.0	1,839.2	1,898.1	1,781.4	1,883.6	1,772.9

Panel C: GPD fitting results for malicious loss series

Data	Block	Loglik	AIC	K-S	Threshold	Scale	Shape
Mal	Weekly	-4,563.2	9,130.6	0.000	0.99	4,813.9	2.755
	Biweekly	-2,920.6	5,845.2	0.035	0.99	21,057.8	2.442
Neg	Weekly	-5,321.9	10,647.9	0.008	0.99	13,802.5	1.535

Data	Block	Loglik	AIC	K-S	Theshold	Scale	Shape
Comp	Month	1,087.3	2,179.4	0.138	0.99	859,777.4	3.460
Mal	Week	3,251.4	6,642.8	0.012	0.99	11,668.7	3.240

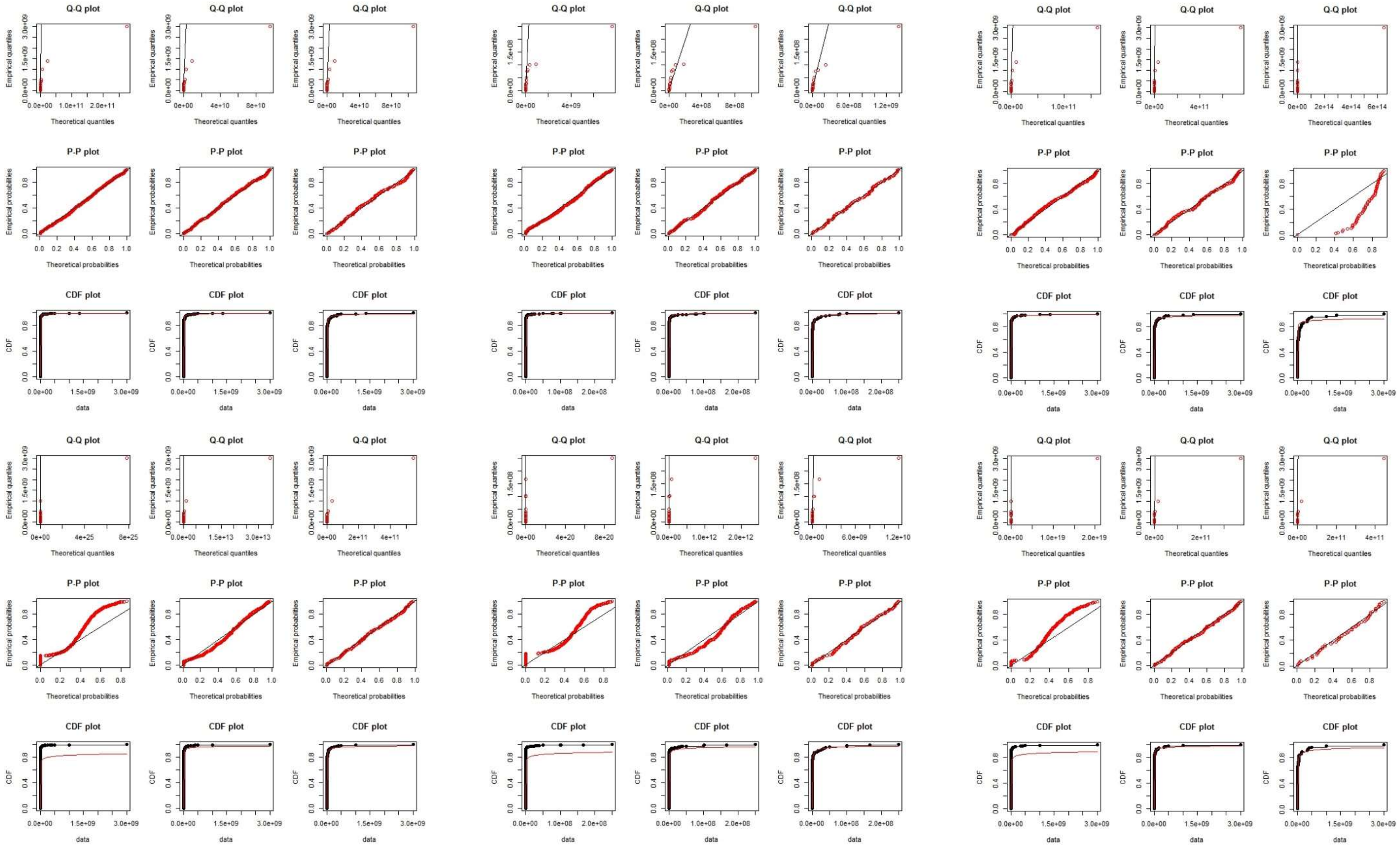
Note: *, ** and *** indicate that the *p*-value is less than the significance levels, 10%, 5% and 1%, respectively. K-S and A-D stand for Kolmogorov-Smirnov test and Anderson-Darling test, respectively, and Loc stands for the location parameter. In panel B, L-norm, IG, and GPD stand for lognormal, inverse Gaussian, and generalized Pareto distribution, respectively. Note that there is no location parameter in panel C for GPD fitting because the mean of GPD distribution in the tail is theoretically infinite (Maillart and Sornette, 2010).

Graphical diagnosis on fitting

	Entire period			Pre-2014			Post-2014		
	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly

Complete

Malicious



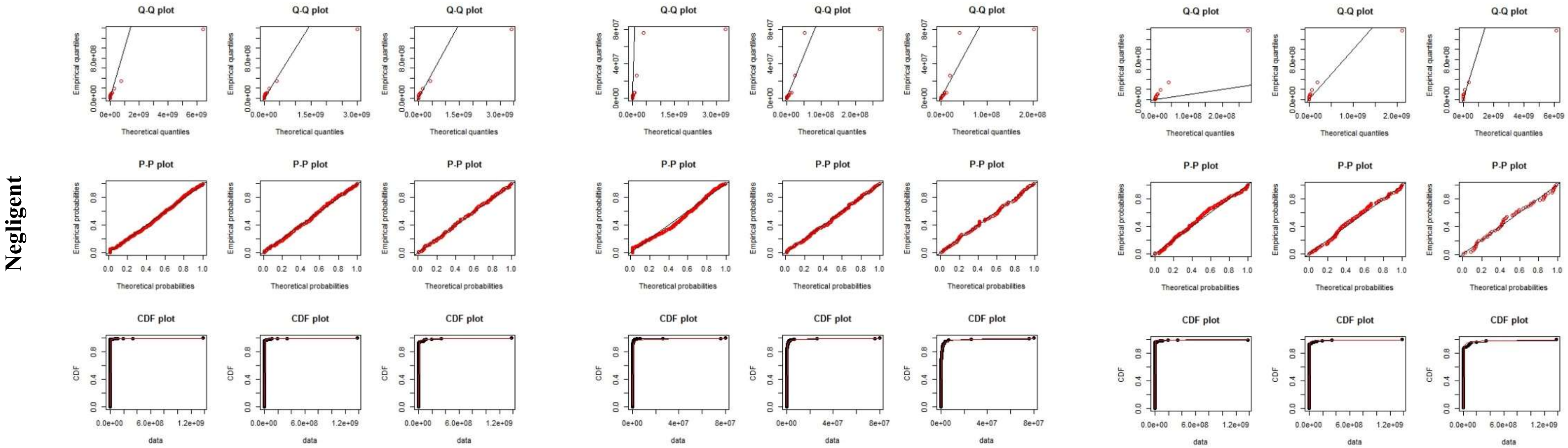


Figure D1: Graphical diagnosis of GEV fitting results. Each data type incorporates three different plots to check whether the fitting is good: QQ plot, probability density plot and cumulative density plot.

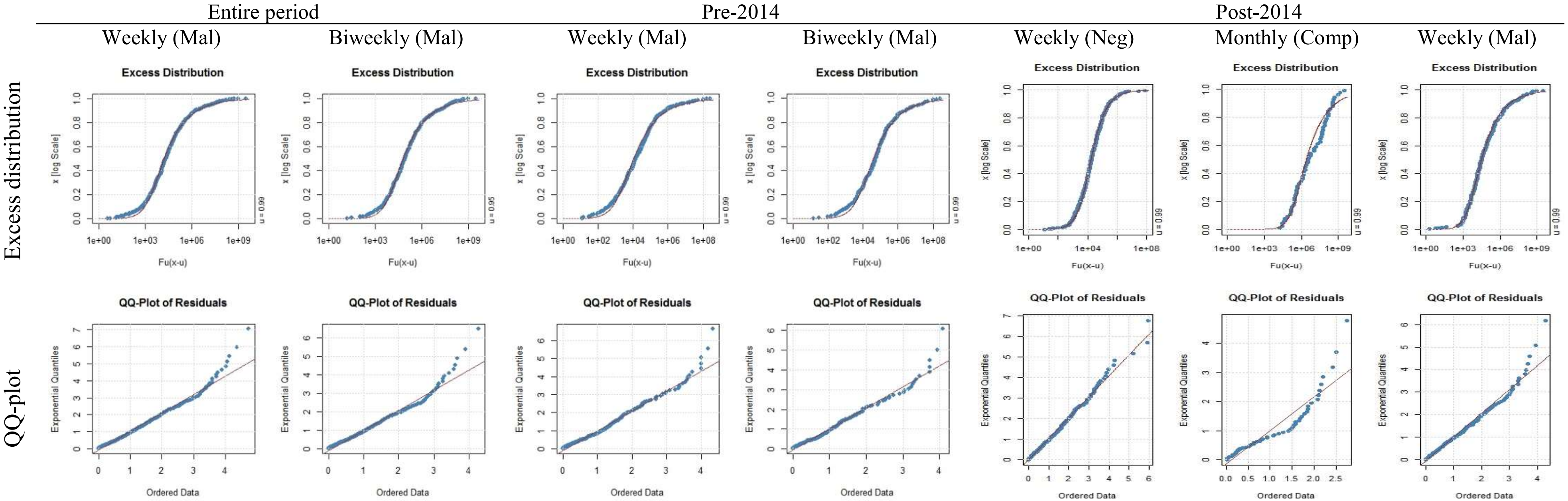


Figure D2: Graphical diagnosis of GPD fitting for seven series that fail to fit GEV.

Appendix E. Numerical results of dependence modeling

Table E1. Dependence Model Fitting Results

Panel A: Test for extreme dependency with Pickands function

	Entire period			Pre-2014			Post-2014		
	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly
Pickands test	0.338***	0.027*	0.025	0.381***	0.032**	0.027	0.042**	0.024	0.024

Panel B: Bivariate extreme value copulas

Family	Copula	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly
Extreme Value	GH	-	-	1.702 (0.192)	-	-	**-0.381** **(0.086)**	-	-0.865 (0.142)	1.989 (0.327**)
	Galambos	-	-	1.646 (0.189)	-	-	-0.102 (0.086)	-	-0.971 (0.136)	**1.890** **(0.242)**
	Tawn	-	-	**1.593** **(0.189)**	-	-	-0.362 (0.098)	-	-0.484 (0.205)	2.000 (0.375**)
	Husler-Reiss	-	-	1.617 (0.186)	-	-	0.054 (0.103)	-	**-1.032** **(0.131)**	1.870 (0.227**)
Elliptical	Gauss	1.733 (0.043**)	1.995 (0.026)	-	1.996 (0.029)	1.921 (0.031)	-	0.739 (0.028)	-	-
	t	3.790 (0.415***)	4.014 (0.040)	-	3.665 (0.492***)	3.675 (0.045**)	-	2.808 (0.045)	-	-
Archi-mean	Clayton	**-2.410** **(0.054**)**	1.968 (0.027)	-	**0.593** **(0.027)**	0.865 (0.028)	-	1.315 (0.024)	-	-
	Frank	1.973 (0.043**)	1.709 (0.024)	-	1.999 (0.029)	1.212 (0.030)	-	**-0.013** **(0.023)**	-	-
	Joe	2.000 (0.044*)	**1.632** **(0.030)**	-	2.000 (0.029)	**0.440** **(0.040)**	-	1.957 (0.055**)	-	-

Panel C: Parameter estimation

	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly	Weekly	Biweekly	Monthly
Size of parameter	0.105	1.024	0.090	0.080	1.051	1.109	0.524	0.675	0.221

Note: *, ** and *** indicate that the *p*-value is less than the significance levels, 10%, 5% and 1%, respectively. In panel B, the figures without parentheses indicate Akaike information criteria and the figures in parentheses account for the statistics of the goodness-of-fit test for extreme value copulas. Bold indicates the best fit model for the corresponding series.

Appendix F. Additional description on reinsurance design

Prior to pricing an insurance policy using the premium principles in Section 5.2, we need to model the loss process, X, which is a compound process using the collective risk model. Specifically, we use the combination of the Monte Carlo method with POT for severity reflecting the heavy-tailedness of cyber risk and the bootstrapping method for frequency to generate a compound loss process. The compound process is described as

$$\lambda_t = \sum_{i=1}^{N_t} X_{i,t}, \tag{F.1}$$

where $t = 1,2, \ldots$ is discrete time on a monthly basis, $X_{i,t}$ is a monthly severity process, and N_t is a monthly count process. We assume that the monthly severity process and the monthly count process are independent and each process is independent and identically distributed (i.i.d.).

One of the key issues in cyber risk (particularly, data breach risk) is how to measure the economic cost of the breach (Eling and Jung, 2018). A standard measurement method has not yet been developed in either academia or industry as a number of factors can affect the translation from a data breach to its economic cost (e.g., company size, industry, revenue, type of information breached, and other company-specific factors). In this application, we apply the 2018 estimate of the average cost of data breach loss across industry by the Ponemon Institute (2018), which is \$148 per record, to define the loss in monetary units.

We then use equations (23) and (24) to derive the premium estimates for the reinsurer and the insurer in the portfolio. We assume the quota level to be 25%, as used in some examples (see, e.g., Federal Insurance Office, 2014), and the assumed quota is applied to the premium estimation by equation (23). The contract duration is assumed to be one year for the portfolio; hence, the coverage limit for CatXL is set by the PMCL estimates likely to happen in the next year, as shown in Table 3. We diversify the limit for different risk types (malicious vs. negligent) and data period (entire vs. pre-2014 vs. post-2014) as done in Table 3. We also assume the fixed level of deductible for individual exposure, \$500,000, which is around the mean value in the current cyber-insurance market, where the deductible usually varies between \$5,000 and \$1 million with the asset value of an insured (Romanosky et al., 2017). Thus, we only consider the loss amount above the deductible per exposure in the reinsurance portfolio.

Appendix G. Fitting Peaks-over-Threshold

In this section, we illustrate the numerical result of POT fitting and the graphical diagnosis on the fitting result. POT in this practice consists of lognormal distribution in the body below the 99% threshold and the generalized Pareto distribution in the tail above the threshold. We check two other models with gamma distribution and Weibull distribution in the body for the sake of comparison and find that the model with lognormal distribution in the body generates the lowest AIC. Although it fails to show a perfect fit in the graphical diagnosis (see Figure G1), we use the POT model with lognormal distribution in the body to compare with our GEV approach.

The results in Table G1 show that the loss processes for the post-2014 period generally have a higher tail index (shape parameter) than those for the pre-2014 period do (particularly for the complete and malicious risk). This result is reflected in panel C of Table 3 in Section 5.1 by the estimated loss size, which indicates that the predicted amounts based on the post-2014 database are on average 9.3 times larger than those based on the pre-2014 database across three quantiles. The difference between malicious and negligent risks turns out to be significant as the estimates of malicious risk are on average 10.2 times bigger than those of negligent risk across three data periods. However, the extreme loss estimates by POT are not comparable in size with PMCL estimates in Table 3, showing that POT estimations might underestimate the size of dragon king loss for the future prediction.

Table G1. POT (Lognormal-GPD) Estimation

		Statistics		Parameters (body-Lognormal)		Parameters (tail-GPD, 99% threshold)	
Period	Data-type	Loglik	AIC	Mean	SD	Scale-Tail	Shape-Tail
Entire	Composite	62,621.8	-125,239.6	7.923	2.689	335.5 m	0.178
	Malicious	22,040.2	-44,076.4	8.166	3.344	503.3 m	0.190
	Negligent	40,421.8	-80,839.6	7.804	2.297	170.7 m	0.249
Pre-2014	Composite	36,362.9	-72,721.8	7.913	2.522	42.0 m	0.184
	Malicious	10,438.7	-20,873.5	7.873	3.180	55.8 m	0.149
	Negligent	25,846.7	-51,689.3	7.939	2.254	16.3 m	0.288
Post-2014	Composite	26,239.0	-52,474.1	7.948	2.948	484.0 m	0.188
	Malicious	11,581.8	-23,159.5	8.443	3.472	661.8 m	0.162
	Negligent	14,528.8	-29,053.5	7.602	2.479	282.5 m	0.153

Note: SD and m stand for standard deviation and million respectively.

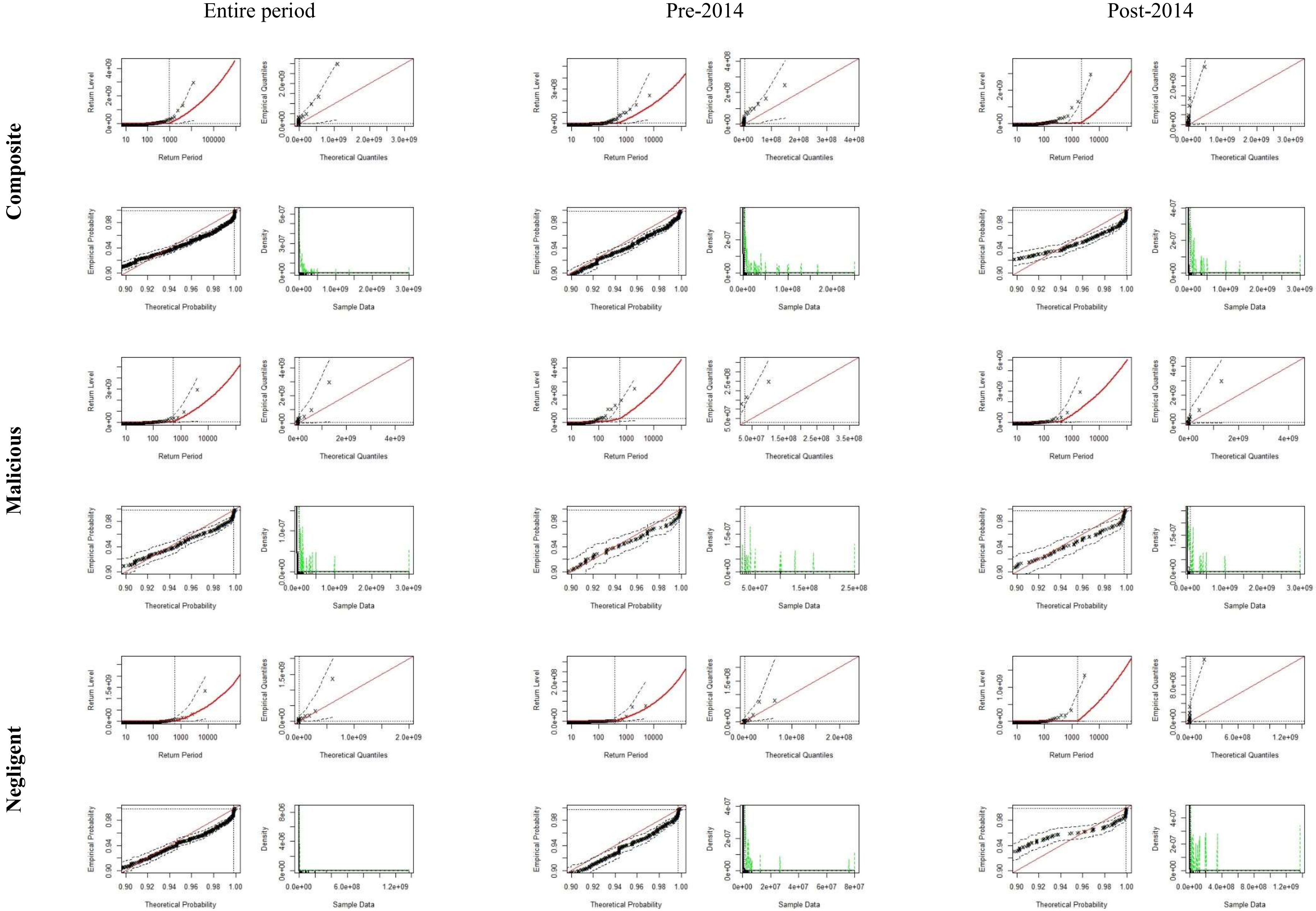


Figure G1: Graphical diagnosis on POT (lognormal-GPD) fitting results for the entire, pre-2014 and post-2014 periods. Each data type incorporates four different plots to check whether the fitting is good: Return level plot (upper-left), QQ plot (upper-right), P-P plot (lower-left) and density plot (lower-right).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.

Aven, T. (2013). On the meaning of a black swan in a risk context. *Safety Science, 57*, 44-51.

Bali, T. (2007). A generalized extreme value approach to financial risk measurement. *Journal of Money, Credit and Banking, 39*(7), 1613-1649.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and applications.* New York: Wiley.

Berr, J. (2017 May 16). "WannaCry" ransomware attack losses could reach $4 billion. *CBS News*. Retrieved from https://www.cbsnews.com/news/wannacry-ransomware-attacks-wannacry-virus-losses/.

Box, G., and Jenkins, G. (1976). *Time series analysis, control, and forecasting* (2nd ed.). San Francisco: Holden-Day.

Box, G., and Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association, 65*(332), 1509-1526.

Brown, J., Cummins, D., Lewis, C., and Wei, R. (2004). An empirical analysis of the economic impact of federal terrorism reinsurance. *Journal of Monetary Economics, 51*(5), 861-898.

Bücher, A., and Segers, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes, 17*(3), 495-528.

Chang, C., and Chang, J. (2017). An integrated approach to pricing catastrophe reinsurance. *Risks, 5*(3), 51.

Charpentier, A., and Le Maux, B. (2014). Natural catastrophe insurance: How should the government intervene? *Journal of Public Economics, 115*, 1-17.

Chavez-Demoulin, V., and Davison, A. (2012). Modelling time series extremes. *REVSTAT-Statistical Journal, 10*, 109-133.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* London: Springer-Verlag.

D'Agostino, R., and Stephens, M. (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker, Inc.

De Haan, L., and Ferreira, A. (2006). *Extreme value theory: an introduction.* New York: Springer.

Diebold, F., Schuermann, T., and Stroughair, J. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance, 1*(2), 30-35.

Dumm, R., Johnson, M., and Watson, C. (2015). An examination of the geographic aggregation of catastrophic risk. *The Geneva Papers on Risk and Insurance - Issues and Practice, 40*(1), 159-177.

Edmonds, T. (2018). *Terrorism insurance & Pool Re.* London: The U.K. Parliament.

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

Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics, 51*(2), 239-248.

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

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

Eling, M., and Toplek, D. (2009). Risk and return of reinsurance contracts under copula models. *European Journal of Finance, 15*(7-8), 751-775.

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

Embrechts, P. (2000). Actuarial versus financial pricing of insurance. *Journal of Risk Finance, 1*(4), 17-26.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance.* Berlin: Springer-Verlag.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica, 50*(4), 987-1007.

Federal Insurance Office. (2014). *The breadth and scope of the global reinsurance market and the critical role such market plays in supporting insurance in the United States.* Washington, DC: U.S. Department of the Treasury.

Ferreira, A., and De Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics, 43*(1), 276-298.

Fisher, R. A., and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180-190. Cambridge University Press.

Frachot, A., Georges, P., and Roncalli, T. (2001). Loss distribution approach for operational Risk. *Working paper*. Retrieved from https://ssrn.com/abstract=1032523.

Frees, E., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks, 4*(1), 4.

Fu, L., and Moncher, R. (2004). Severity distributions for GLMs: Gamma or lognormal? Evidence from Monte Carlo simulations. *Casualty Actuarial Society Discussion Paper Program*, 149-230.

Gatzert, N., and Schmeiser, H. (2012). Industry loss warranties: Contract features, pricing, and central demand factors. *Journal of Risk Finance, 13*(1), 13-31.

Genest, C., Kojadinovic, I., Neslehova, J., and Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli, 17*(1), 253-275.

Ghorbal, B. N., Genest, C., and Neslehova, J. (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *Canadian Journal of Statistics, 37*(4), 534-552.

Gilli, M. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics, 27*(2-3), 207-228.

Guerra, M., and Centeno, M. D. L. (2008). Optimal reinsurance policy: The adjustment coefficient and the expected utility criteria. *Insurance: Mathematics and Economics, 42*(2), 529-539.

Guerra, M., and Centeno, M. D. L. (2010). Optimal reinsurance for variance related premium calculation principles. *ASTIN Bulletin, 40*(1), 97-121.

Hofmann, A., Wheatley, S., and Sornette, D. (2019). Heavy-tailed data breaches in the Nat-Cat framework and the challenge of insuring cyber risks. *Working paper*.

Hurvich, C., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Jaffee, D., and Russell, T. (1997). Catastrophe insurance, capital markets, and uninsurable risks. *Journal of Risk and Insurance, 64*(2), 205-230.

Kaluszka, M. (2001). Optimal reinsurance under mean-variance premium principles. *Insurance: Mathematics and Economics, 28*(1), 61-67.

Kaluszka, M. (2005). Optimal reinsurance under convex principles of premium calculation. *Insurance: Mathematics and Economics, 36*(3), 375-398.

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., and Stulz, R. (2018). *What is the impact of successful cyberattacks on target firms?* Cambridge, MA: National Bureau of Economic Research.

Kleiber, C., Hornik, K., Leisch, F., and Zeileis, A. (2002). Strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software, 7*(2), 1-38.

Kojadinovic, I., and Yan, J. (2010). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis, 101*(9), 2234-2249.

Kojadinovic, I., Segers, J., and Yan, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics, 39*(4), 703-720.

Kunreuther, H. (2002). The role of insurance in managing extreme events: Implications for terrorism coverage. *Risk Analysis, 22*(3), 427-437.

Kunreuther, H. (2015). The role of insurance in reducing losses from extreme events: The need for public–private partnerships. *The Geneva Papers on Risk and Insurance - Issues and Practice, 40*(4), 741-762.

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1-3), 159-178.

Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields, 65*(2), 291-306.

Ljung, G., and Box, G. (1978). On a measure of a lack of fit in time series models. *Biometrika, 65*(2), 297-303.

Lloyd’s. (2017). *Counting the cost: Cyber exposure decoded.* London: Lloyd’s.

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

Moscadelli, M. (2004). The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee. *Working paper*. Retrieved from https://ssrn.com/abstract=557214.

Orcutt, M. (2017, April 6). Insurers scramble to put a price on a cyber catastrophe. MIT Technology Review. Retrieved from https://www.technologyreview.com/s/603937/insurers-scramble-to-put-a-price-on-a-cyber-catastrophe/.

Paté-Cornell, M.-E., Kuypers, M., Smith, M., and Keller, P. (2018). Cyber risk management for critical infrastructure: A risk analysis model and three case studies. *Risk Analysis, 38*(2), 226-241.

Phillips, P., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*(2), 335-346.

Pickands, J. (1981). Multivariate extreme value distributions. *Proceedings of the 43rd Session of the International Statistical Institute*, *49(2)*, 859-878.

Ponemon Institute. (2018). *2018 cost of a data breach study.* Traverse City, MI: Ponemon Institute.

PricewaterhouseCoopers (PwC). (2016). *Global state of information security survey 2016.*

Privacy Rights Clearinghouse. (2019). *Data breaches*. Retrieved from https://www.privacyrights.org/data-breaches.

Ralph, O. (2017, July 17). Extreme cyber attack could cause $120bn in economic damage. *The Financial Times*. Retrieved from https://www.ft.com/content/dff0649c-671a-11e7-8526-7b38dcaef614.

Rasoulian, S., Gregoire, Y., Legoux, R., and Senecal, S. (2017). Service crisis recovery and firm performance: insights from information breach announcements. *Journal of the Academy of Marketing Science, 45*(6), 789-806.

Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. (2017). Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk? *Working paper*.

Said, S., and Dickey, D. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika, 71*(3), 599-607.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Shevchenko, P. (2011). *Modelling operational risk using Bayesian inference.* Heidelberg: Springer.

Sklar, A. (1959). Fonctions de repartition à n dimensions et leurs marges. *Inst. Statist. Univ Paris 8*, 229-231.

Smetters, K. (2005). *Insuring against terrorism: The policy challenge.* Cambridge, MA: National Bureau of Economic Research.

Sornette, D., and Ouillon, G. (2012). Dragon-kings: Mechanisms, statistical methods and empirical evidence. *The European Physical Journal Special Topics, 205*(1), 1-26.

Taleb, N. N. (2007). *The Black Swan: the impact of the highly improbable* (2nd ed.). London: Penguin.

Wang, S. (2017). Integrated framework for information security investment and cyber insurance. *Working paper*. Insurance Risk and Finance Research Centre (IRFRC), Nanyang Technological University.

Wheatley, S., Maillart, T., and Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B, 89*, 7.

Wilkinson, M. (1982). Estimating probable maximum loss with order statistics. *Proceedings of the Casualty Actuarial Society, 69*, 195-209.

World Economic Forum (WEF). (2016). *Understanding systemic cyber risk.* Geneva: World Economic Forum.

Zheng, K., and Albert, L. (2019). A robust approach for mitigating risks in cyber supply chains. *Risk Analysis*, Forthcoming.

Essay III

Decision-making on cyber risk management: Interaction between market insurance and risk control measures under prospect theory

Abstract

This paper studies decision-making on cyber risk management in the presence of interdependent risk, comparing market insurance, self-protection and self-insurance. Our economic decision model reflects interdependent risk and loss aversion, a combination that—to the best of our knowledge—has not been examined in the literature, but is increasingly relevant in an interconnected world. We find that an agent with self-protection as the reference point is likely to not invest in other risk management activities (market insurance and self-insurance), providing support for the anecdotal evidence of a fatalistic behavior with respect to cyber risks (i.e., a perceived underinvestment in cyber risk management, which becomes rational in our decision model). However, we empirically show that the demand for additional risk management activities might increase as agents are exposed to increasing frequency rate year-by-year. The focus of our paper is the increasingly relevant field of cyber risk, but the results can be generalized to any other interdependent risk.

Keywords: Cyber risk, Decision theory, Loss aversion, Interdependent risk, Self-protection, Self-insurance

---

Martin Eling, Kwangmin Jung (2019).

This paper was presented at the 2019 annual seminar of European Group of Risk and Insurance Economists (EGRIE) in Rome, the 2019 annual meeting of American Risk and Insurance Association (ARIA) in San Francisco, the 2019 German Insurance Congress (DVfVW) in Berlin, the 2018 annual meeting of the Asia Pacific Risk and Insurance Association (APRIA) in Singapore, the 2018 Frankfurt Insurance Research Workshop of International Center for Insurance Regulation (ICIR) at Goethe University Frankfurt, the PEF/PiF poster seminar at the University of St. Gallen and the 7th research seminar of the Institute of Insurance Economics at the University of St. Gallen.

1 Introduction

The "WannaCry" ransomware attack in 2017 affected more than 200,000 computers running the Microsoft Windows operating system in 150 countries within one day; the estimated total monetary loss reached almost $4 billion (Berr, 2017). Deep Root Analytics, a data analysis company in the U.S., exposed a database including the personal information of 200 million American voters online in 2017 (Grenoble, 2017), a disclosure very different from the "WannaCry" attack, but a realistic threat and concern for firms, policy-makers and individuals.[83] Such extreme events are less likely to occur than the smaller cyber events of daily life (Eling and Wirfs, 2019), but can trigger huge, catastrophic consequences.

One of the key specifics in cyber risk to trigger a systemic loss is the interconnection in the network environment. For instance, anonymous attackers successfully paralyzed a number of web services operated by Domain Name System and hijacked millions of internet-connected household appliances in October 2016 (Perlroth, 2016). This distributed denial of service attack was possible due to the spread of networked home electronics. The success rate of this kind of attack and the size of loss are expected to grow as the Internet of Things expands and the interconnection becomes more complicated (Ernst & Young, 2015).

Cyber events in the interconnected network environment might significantly disrupt businesses in a range of fields, leading to an increase in the awareness of cyber risk among policy-makers, regulators and decision-makers in firms.[84] This should motivate decision-makers to invest substantial resources in cyber risk management, including the option to buy insurance to transfer this risk. However, several surveys reveal a perception that decision-makers regard cyber events (e.g., hacking, unintended disclosure or intentional breach by an insider) as rather unrealistic to themselves, until they finally experience a loss.[85]

[83] Other examples are the data breaches resulting from the hacking of Yahoo in 2013 (three billion accounts) and FriendFinder in 2016 (412 million users); see Armerding (2018) for the biggest data breach events in history.

[84] For example, PricewaterhouseCoopers (2017) conducts a survey with 9,500 decision-makers across 122 countries and documents the increasing awareness of potential consequences by cyber events. The estimated global loss of cybercrime is in an area of $600 billion (see, e.g., CSIS, 2018).

[85] For instance, de Smidt and Botzen (2018) conducted a survey with 172 professional decision-makers in the Netherlands and call this type of perception "not-in-my-organization effect" (p. 247). Jalali et al. (2019) also argue that many organizations constantly ignore or underestimate cyber risk. This sort of perception can be called "optimism bias" against a catastrophe event (or a negative event), as defined by Weinstein (1980). The author finds evidence of unrealistic optimism—namely agents believe that it is less likely for them to experience negative events than for others. This bias has also been studied in the insurance context, where, for example, people are more likely to insure against modest risk with high probability than against extreme risk with low probability (see, e.g., Schanz, 2019, for an overview of behavioral biases in the insurance research).

Some decision-makers also seem to have a fatalistic behavior toward cyber risk. Even if a threat becomes more realistic, no single party can be fully secure anyway (no matter what cyber risk management strategy is implemented); thus, decision-makers might perceive that the threat will not happen to them. Two recent surveys by the U.K. Government (DCMS, 2018) and EIOPA (2018) support this observation by showing that the most common reason not to take up cyber-insurance is that the decision-makers do not see themselves at high risk.

Extant literature about how agents make decisions facing (cyber) risk has already documented several results with respect to self-protection and the purchase of (cyber-) insurance.[86] However, all the above studies assume decision-makers under risk aversion, which—as indicated above—might not very well describe the actual behavior of market participants. Furthermore, no study has investigated a behavioral bias like the unrealistic optimism in the presence of interdependent risk.

To fill this gap in the literature, we study decision-making on interdependent cyber risk in the context of prospect theory (Kahneman and Tversky, 1979). We assume that agents are more averse against a loss than against a gain by comparing possible outcomes with a reference point.[87] The reference points considered in this paper rely on two possible instruments: self-protection and self-insurance. Self-protection is the key to describe the status quo of business parties in the network environment to build a realistic set-up in the cyber risk context. It also plays the role of a public good (Lohse, Robledo and Schmidt, 2012), which can affect the loss probability of other agents in the interconnected network environment. In contrast, self-insurance does not affect the risk of others, but rather reduce the loss amount that the agent can face with as a private good (Ehrlich and Becker, 1972).

We develop a conceptual model by comparing potential decisions on market insurance, self-insurance, and both options in the presence of self-protection with the reference points under

[86] With respect to insurance in general see Ehrlich and Becker, 1972; Hofmann, 2007; Mürmann and Kunreuther, 2008; Alary, Gollier and Treich, 2013. With respect to cyber insurance, the level of investment in the security system and the demand of insurance are modeled (see, e.g., Gordon and Loeb, 2002; Lelarge and Bolot, 2009; Shetty et al., 2010; Öğüt, Raghunathan and Menon, 2011; Hofmann and Ramaj, 2011; Wang, 2017). Several studies analyze the impact of interdependency between network systems on loss probability (see, e.g., Lelarge and Bolot, 2009; Shetty et al., 2010; Öğüt, Raghunathan and Menon, 2011; Hofmann and Ramaj, 2011). Some studies incorporate the lack of proving the loss occurrence by an insurer, which can be linked with the presence of information asymmetry (see, e.g., Lelarge and Bolot, 2009; Öğüt, Raghunathan and Menon, 2011), which is not the focus of this paper. One relevant study to our focus is Verendel (2008), who uses prospect theory to explain decisions on cybersecurity control. However, the author only considers decisions on self-protection measures, not market insurance and self-insurance. In addition, the possibility of interdependency in risks is not taken into account.

[87] Mersinas et al. (2016) carried out an experiment with 117 participants, including 59 IT security professionals and 58 students, and observed a behavioral bias in line with Kahneman and Tversky's (1979) finding that IT risk managers switch their risk preferences from risk aversion to risk seeking when they experience a loss.

interdependent risk.[88] Based on the reference points and the model framework, we find that an agent with the reference point of self-protection as an essential effort against cyber risk is likely to not invest in additional risk management measures (market insurance and self-insurance).[89] A fatalism with respect to cyber losses (thus, a demand side anomaly) is that agents stay uninsured despite increasing awareness, so the presence of cyber risk becomes rational in our decision model.[90] We also provide an empirical explanation of our theoretical findings considering the historical frequency rate of cyber losses.

Our findings are important for better understanding the nature of cyber risks and their consequences for decision-making on cyber risk management. The results are useful for risk managers and IT professionals in firms and insurance companies that are developing cyber-insurance policies. For firms, the results are important not only for decisions on internal risk management, but also for new requirements of reporting cyber incidents because the requirement emphasizes the importance of risk control measures (Eling and Wirfs, 2019).[91] Furthermore, this paper contributes not only to the growing literature on cybersecurity in the business and economics domain (e.g., Nagurney and Shukla, 2017), but also to the general business and economics literature in that it applies a descriptive decision model (prospect theory) in the presence of interdependent risk.[92]

---

88 Table A1 in Appendix A categorizes the models for decision-making under risk into whether they consider risk control measures and loss aversion under the presence of interdependent risk and illustrates the positioning of this study. As observed in the table, no study has looked at loss aversion (prospect theory) under interdependent risk. We further clarify our contribution in Table A2 by comparing our paper with two studies on insurance decision under loss aversion.

89 Eckles and Volkman-Wise (2019) attempt to explain, using prospect theory-type preferences, several phenomena that cannot be explained by the expected utility theory. In particular, the phenomena studied in the paper are the preference for low deductibles, the lack of demand for catastrophe insurance and the over-demand to insure small losses. Furthermore, they find that agents under loss aversion ($\lambda > 1$, where λ is the loss aversion parameter) tend to demand less insurance than agents under the expected utility theory. Our finding is exactly in line with their finding that agents with prospect theory-type preferences (loss aversion) are willing to take on more risk to avoid losses.

90 To argue that the behavior described in our decision model is rational, one might have to question whether loss aversion is rational. Tversky and Kahneman (1991) clarify it by concluding that the reference-dependent decision under loss aversion can be justified as a rational behavior when the consequences by the corresponding effect of the reference point on the decision are indeed experienced. This idea has also been employed in other contexts, such as decision-making about pension/old-age provision by Binswanger (2007), where the study evaluates normative decisions using loss aversion assumption (specifically, the lexicographic loss aversion model).

91 In the U.S., reporting requirements for data breaches have been introduced in many states since 2002 (NCSL, 2019). In the European Union such reporting requirements will apply from 2018 (European Union, 2016). After their enforcement, more data and information will be available. This has already happened with data samples in the U.S. as the Privacy Rights Clearinghouse "Chronology of Data Breaches" (PRC, 2019).

92 Apart from cyber risk context, Kunreuther and Pauly (2018) find a perception that many uninsured individuals decide to take up insurance only after suffering a loss. In particular, this perception can be observed from the fact that, for the low-probability, high-consequence (LP-HC) events, individuals rarely experience a loss event either personally or socially by observing first-hand other people's losses from disasters. This perception can be explained by our framework because interdependent risk is also a matter for disaster risk or disease risk

The remainder of this paper is organized as follows: We provide a brief illustration on our descriptive decision model (prospect theory) in Section 2 and develop our model framework by describing interdependent risk, market insurance and reference points in Section 3. This framework is used to show how agents make decisions on market insurance and risk control measures with different reference points in Section 4. Finally, Section 5 presents the conclusion.

2 Theoretical background: Prospect theory

The expected utility theory is based on the assumption that an individual makes a decision under risk by maximizing her expected utility over some final states of wealth (von Neumann and Morgenstern, 1953); the two economic states typically considered in the insurance domain are determined by a loss occurrence. A pre-defined (concave) utility function depending on the risk preferences is used to determine the utility level; however, the theory assumes dependence on the final wealth for the utility level, which largely simplifies the analysis on individual choice over risky prospects (Tversky and Kahneman, 1991).

Prospect theory developed by Kahneman and Tversky (1979) is based on the assumption that an individual dislikes a loss more than she likes a gain of a comparable size; hence, the level of utility in an economic situation is determined by the gain-loss domain. An individual makes a decision on the purchase of insurance based not on the final wealth, but on gains and losses compared to a reference point. This implies that gains and losses count the opportunity cost by comparing a potential decision with the reference point. Tversky and Kahneman (1992) complement their initial model (Kahneman and Tversky, 1979) by taking into account the cumulative functional form separately to gains and losses, thereby resolving two possible problems of the first model: lack of the explanation on stochastic dominance and lack of ability to accommodate a large number of outcomes.

We consider an individual facing two possible outcomes in an economic situation, x_1 and x_2. The individual has the loss probability, p, and makes decisions using a value function (hereafter, KT value function) and a probability weighting function (hereafter, KT weighting function). The probability weighting function is a non-linear form reflecting overweight on less likely events and underweight on more likely events. The function features rank-dependent weighting, and the violation that the sum of the subjective probability by the function is not 1. The

(e.g., cancer) so that vulnerability of an individual (or property) can affect the neighborhood; hence, our framework can be generalized to this context.

mathematical expressions of the value function and non-linear weighting functions are (Tversky and Kahneman, 1992):

$$v(x) = \begin{cases} x^{\alpha} & if\ x \geq 0 \\ -\lambda(-x)^{\alpha} & if\ x < 0 \end{cases}, \tag{1}$$

$$\theta^{+}(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}, \tag{2}$$

$$\theta^{-}(p) = \frac{p^{\xi}}{(p^{\xi} + (1-p)^{\xi})^{1/\xi}}, \tag{3}$$

where $\alpha \in [0,1]$ is a parameter to determine the diminishing sensitivity for concavity in the gain domain and for convexity in the loss domain, λ is a loss aversion parameter, and γ and ξ are probability weighting parameters differentiated depending on the gain or loss state. Tversky and Kahneman (1992) propose the values of α (= 0.88) and λ (= 2.25) with the median parameters found from their experimental study. In this value function (see Figure 1), a loss is defined by a negative value of x, whereas a gain is defined by a positive value. The diminishing sensitivity, α, indicates that an individual is risk averse in the gain state ($x \geq 0$) and risk seeking (i.e., loss averse) in the loss state ($x < 0$)—that is, $v''_{-} > 0, v''_{+} < 0,$ where v''_{-} is the second order condition in the loss domain and v''_{+} is the second order condition in the gain domain. The reference point at the origin in Figure 1 is not differentiable (inflection point), and the higher curvature appears in the loss state than in the gain state.

The expected value of the "prospect" (an economic situation) is determined by the combination of the value function and the weighted probability as follows:

$$V = \theta^{+}(p) \cdot v(x_1) + \theta^{-}(q) \cdot v(x_2), \tag{4}$$

where x_1 and x_2 indicate outcomes of two states (gain and loss) and p and q are the corresponding probabilities. The expected value of the "prospect" can be defined as subjective expected value for the decision-making, which is conceptually equivalent to the expected utility theory. The probability applied to the states is the subjective probability measure analytically calculated using the probability weighting function. According to the theory, very small probabilities are either overweighted or rounded to 0 (Kunreuther and Michel-Kerjan, 2014; Schmidt, 2016).

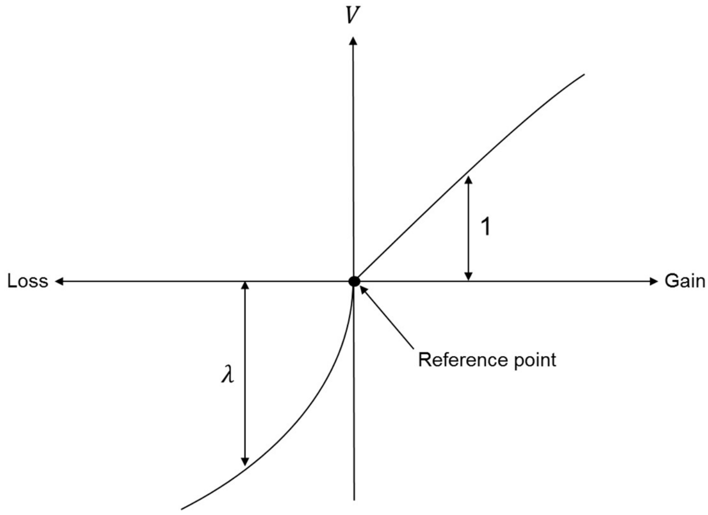


Figure 1. The shape of the value function. The curvature of the gain state is concave, whereas that of the loss state is convex. λ is the parameter of the loss aversion, showing how convex the curvature is.

3 Model framework

3.1 Modeling interdependency with self-protection

Economic agents[93] nowadays operate their business in an interconnected network environment, represented in Figure 2 with nodes as atomic elements controlled by agents (Böhme and Schwartz, 2010). If one agent were breached in an attack, there would be a high likelihood that the remaining agents in the environment would be breached as well.[94] This indirect risk to the other agents results in externality and is of importance in the network environment, because a systemic risk by a global hacking attack or virus can take place.

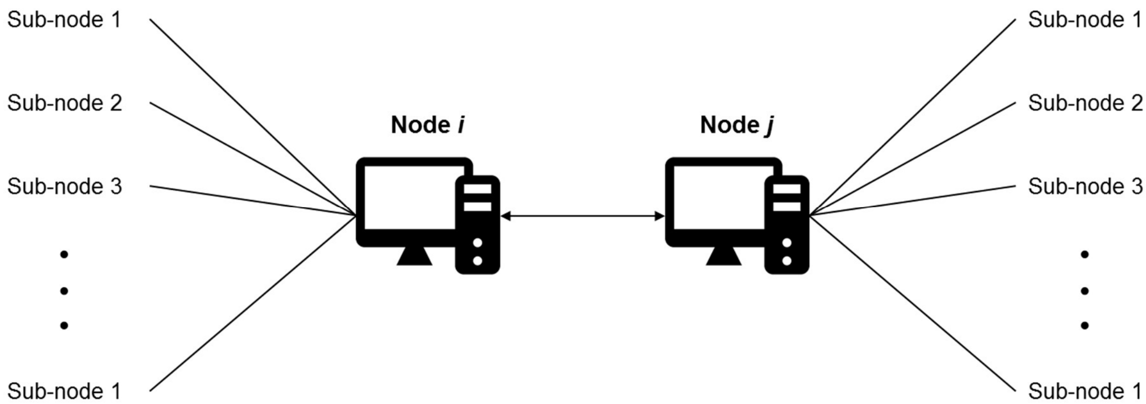


Figure 2. Interconnected network environment between agent *i* and agent *j*.

[93] Following Böhme and Schwartz (2010), we use "agent" to describe possible economic entities looking for cyber-insurance. Potential agents of cyber-insurance include firms, individual clients and governmental and non-governmental institutions, according to Böhme and Schwartz (2010).

[94] Hofmann (2007) and Mürmann and Kunreuther (2008) also parameterize both direct and indirect risks into the model to reflect the possibility of contagion risk from other sources. However, neither study considers the degree of interdependency, counting only the presence of interdependency.

For simplicity, suppose that there are two agents (i and j) in the network. Each agent makes a monetary effort on enhancing the security system denoted by s_i and s_j. These efforts represent all possible management activities by the agents to improve their security levels, such as the financial investment in updating the security software, cybersecurity training for employees, the production of a manual for possible scenarios and regular investigations into security. In the insurance context, this kind of effort is called self-protection, which can reduce the loss probability (Ehrlich and Becker, 1972). The loss probability of each agent is characterized by a defense function with the self-protection effort defined by Böhme and Schwartz (2010).[95]

$$p = D(s, G), \tag{5}$$

where G is a function of network topology indicating 0 when the network is connected and 1 when it is disconnected and $D(s, G) \in [0,1]$. In this paper, we consider the interconnected network environment, leading to $D(s) = D(s, G)$. The defense function $D(\cdot)$ satisfies the following assumption:

Assumption 1:

(i) The loss probability decreases with the increasing level of the security investment:

$$\frac{\partial D(s)}{\partial s} < 0. \tag{6}$$

(ii) Increasing marginal loss probability over the marginal self-protection level is present:

$$\frac{\partial^2 D(s)}{\partial s^2} > 0. \tag{7}$$

(iii) The loss probability cannot reach zero despite the substantial capacity of the agent to invest in security:

$$\lim_{s \to \infty} D(s) > 0. \tag{8}$$

Assumption 1 describes that the defense function is convex and assumes positive externality; however, the risk will not disappear even if the agent spends infinite amounts for self-protection. In other words, the rate of risk reduction slows down as the level of self-protection is higher; hence, no security measure can fully eliminate the risk.

---

[95] The loss probability in the cyber risk context is also called *vulnerability*, which indicates the probability that a breach attack is successful (Gordon and Loeb, 2002; Wang, 2017). We stay with the term loss probability, meaning that the loss occurs when a system is breached.

At the core of the model is interdependent risk. Mürmann and Kunreuther (2008) provide a theoretical background on the optimal level of self-protection and insurance in the presence of interdependent risk.[96] Here, following Hofmann (2007) and Mürmann and Kunreuther (2008), we consider the presence of interdependent risk, but all agents invest in the security system for self-protection. The probability of cyber loss for agent *i* in this setting can be defined as

$$p_i = D_i(s_i) + \big(1 - D_i(s_i)\big) \times D_j\big(s_j\big) \in [0,1]. \tag{9}$$

The loss probability of agent i (= p_i) is affected by the externality from interconnected agent *j* with its defense function, $D_j\big(s_j\big)$. This implies that the size of self-protection by agent j influences the loss probability of agent *i*. Following Mürmann and Kunreuther (2008), we also assume that the externality is perfect in our model, which means that the incurred loss in one policyholder spreads to the other policyholder with certainty. Given Assumption 1, the following assumption also holds:

Assumption 2:[97]

(i) Diminishing loss probability with interdependent risk over increasing self-protection level:

$$\frac{\partial p_i}{\partial s_i} < 0. \tag{10}$$

(ii) Increasing marginal loss probability with interdependent risk over increasing marginal self-protection level:

$$\frac{\partial^2 p_i}{\partial s_i^2} > 0. \tag{11}$$

(iii) The loss probability with interdependent risk does not disappear despite the agent's substantial capacity to invest in self-protection:

$$\lim_{s_i \to \infty} p_i(s_i) > 0. \tag{12}$$

---

96 Mürmann and Kunreuther (2008) call the interdependent risk contamination between agents and consider the case of no contamination in comparative statics. However, no contamination between agents is not the case for the cyber risk context, especially in the hyper-connected world with the Internet of Things. Therefore, we do not take into account the case of no contamination in our model.

97 Given that assumption 1 holds, assumptions (2-i) and (2-ii) can be derived in the following way:

$$\frac{\partial p_i}{\partial s_i} = D_i'(s_i) - D_j\big(s_j\big) \cdot D_i'(s_i) = \Big(1 - D_j\big(s_j\big)\Big) \cdot D_i'(s_i) < 0,$$

$$\frac{\partial^2 p_i}{\partial s_i^2} = \Big(1 - D_j\big(s_j\big)\Big) \cdot D_i''(s_i) > 0,$$

3.2 Market insurance design and self-insurance

The cyber-insurance market, particularly in the U.S. where the market is currently most developed, typically provides policies with deductibles, premiums between \$10,000 and \$100,000, and cover limits between \$10 million and \$50 million (Romanosky et al., 2017). The deductibles are, generally speaking, as low as \$5,000 and reach amounts between \$500,000 and \$1 million for insureds, whose asset values are around \$1 billion or more.[98] We denote the deductible with d and additionally consider a loading factor ($= \delta$) proportional to the loss probability to account for cost in underwriting cyber-insurance. The price of cyber-insurance relies on the level of self-protection. Note that the presence of information asymmetry possibly leading to adverse selection and moral hazard is not taken into account in this study. The indemnity I and the premium π with the deductible are defined as (Zweifel and Eisen, 2012, Chapter 3)

$$I = \max[L - d, 0] \cdot \beta = (L - d)^{+} \cdot \beta, \tag{13}$$

$$\pi = (1 + \delta) \times p_i \times I, \tag{14}$$

where L is a loss and β is the coverage of a proportional insurance with the loading factor δ.

Self-insurance does not influence loss probability (Ehrlich and Becker, 1972), meaning that loss probability in the presence of interdependent risk is not affected by the implementation of self-insurance.[99] Following Böhme and Schwartz (2010), we embed self-insurance into our model by denoting the level of self-insurance for agent i with $g_i \in [0,1]$ and a cost function of self-insurance by $K(g_i)$. We assume that the cost function of self-insurance is concave with respect to g_i ($K'(g_i) > 0, K''(g_i) < 0$).

3.3 Reference points

A key approach to analyze decision-making under loss aversion is to set the reference point (Tversky and Kahneman, 1991). A decision-maker makes use of the reference point by comparing it with a potential decision, thereby framing the gain and loss of the decision. For example, for insurance decision-making, the reference point is compared with an insurance plan

---

[98] Romanosky et al. (2017) analyzed 100 cyber-insurance policies and documented that cyber-insurance carriers are mainly concerned about the correlated risk (interdependent risk between network systems), leading to a systemic risk but not so much the possible information asymmetry.

[99] Self-insurance in this paper can be described by a loss reduction activity as part of the risk control plan (Hofmann and Peter, 2016). Several forms of self-insurance have been implemented in practice. For example, Ehrlich and Becker (1972) regard sprinkler systems as a self-insurance measure to reduce the size of a loss from fires. In the context of cyber risk, Grossklags et al. (2008) exemplify self-insurance activity with having good backups or conducting regular backups.

that could be bought. We consider two possible reference points under prospect theory: status quo with self-protection and status quo with both self-protection and self-insurance. The reference point based on self-protection is cyber-specific because all business parties are connected via computerized operations and today's network systems require regular security updates.

Status quo with self-protection as a public good

As a reference point, we assume that all agents make investments in regular security updates to protect against possible cyber-attacks. This self-protection effort reduces loss probability. It could also be that the government requires all economic parties to enhance their security systems to a certain level. This definition of the reference point is material in that the positive externality from concurrent efforts in the interconnected network environment matters to social welfare. Therefore, this positive externality plays a role in promoting the public good.[100]

We compare two scenarios with this reference point: decisions on market insurance with self-protection and decisions on self-insurance with self-protection. All scenarios assume that the self-protection measure is implemented in any decision circumstance to build a realistic setting.

Status quo with both self-protection and self-insurance

Here the agent makes an investment in enhancing its security system to protect itself against a possible cyber-attack and conducts a self-insurance plan simultaneously with self-protection. The self-insurance plan can reduce the magnitude of a loss, particularly for a potential extreme cyber case, but it does not serve as a public good because it reduces the size of a loss only for those who invest in self-insurance. It can thus be regarded as a private good (Grossklags et al., 2008), which provides a benefit only for the agent implementing the plan. Simultaneous implementations on risk control measures might impose a significant cost to the agent, but reduce the loss probability and size. This definition of the reference point is material in that it can show interaction between market insurance and both risk control measures.

In addition, a safe option with the maximum effort as the reference point is plausible for a prospective cyber-insured, which could offer a boundary value to the insured in decision-making. Schmidt (2016) defines the safe option as the purchase of insurance with full coverage; however, full insurance is not a realistic option in the current cyber-insurance market, as

[100] To the best of our knowledge, there has been no literature to investigate decision-making on self-protection under prospect theory. However, several studies identify the equilibrium under expected utility theory that the interdependent nature of risk leads one to underinvest in self-protection when there exists a positive externality from another's self-protection measure and a limiting insurance coverage can improve individual or social welfare (Hofmann, 2007; Mürmann and Kunreuther, 2008).

discussed in Section 3.2. (because of cover limits and deductibles). Thus, we set the safe option as the possible maximum effort on risk controls by agents.[101] We compare one possible scenario with this reference point: decisions on market insurance only with self-protection. In this case, we assume that market insurance is the substitute for self-insurance and the agent continues to maintain self-protection efforts as the basic risk management tool. Table 1 summarizes the parameters of the model. We focus on a model with two states of the world, either a loss state or a no-loss state, whose components are illustrated in Table 2.[102]

Table 1. Model Parameters

Notation	Explanation
$\boldsymbol{v}(\cdot)$	Value function
$\boldsymbol{W}$	Initial wealth
$\boldsymbol{L}$	Loss amount
$\boldsymbol{D_i}(\boldsymbol{s_i})$	The defense function for agent *i* as a function of the self-protection level ($s_i \in [0, W]$)
$\boldsymbol{D_j}(\boldsymbol{s_j})$	The defense function for agent *j* as a function of the self-protection level ($s_j \in [0, W]$)
$\boldsymbol{p_i}$	The loss probability of agent *i* in the presence of interdependent risk
$\boldsymbol{d}$	Deductible ($0 \leq d$)
$\boldsymbol{\delta}$	Proportional loading factor ($\delta \in [0,1]$)
$\boldsymbol{\beta}$	Insurance coverage ($\beta \in [0,1]$)
$\boldsymbol{I}$	Indemnity with deductible
$\boldsymbol{\pi}$	Cyber-insurance premium
$\boldsymbol{g_i}$	The level of self-insurance ($g_i \in [0,1]$)
$\boldsymbol{K}(\boldsymbol{g_i})$	A cost function of self-insurance

In summary, the following points reflect the features embedded in our framework, which affect the decision-making on cyber risk management and the current cyber-insurance market:

1. *The presence of interdependent risk in the network environment.*
2. *Realistic design for insurance in the current cyber-insurance market: no available full coverage (deductible and partial coverage).*
3. *Self-protection (essential risk control measure in the interconnected business environment) as the reference point.*

However, our model can also be generalized to any decision-making problem in the presence of interdependent risk under loss aversion.

[101] As supporting evidence, Baillon et al. (2016) empirically tested with 139 participants what reference points they would choose for decision-making and found that the status quo and the maximum outcome that they can achieve are most often selected as the reference point.

[102] We apply the structure of tables in Schmidt (2016) to Table 2 for a clear overview of the model framework.

Table 2. Model Components

Reference point 1: Status quo with self-protection (public good)

	Case 1: Comparison with market insurance (+ self-protection)	
State	No-loss	Loss
Probability	$1 - p_i$	p_i
Reference point	$W - s_i$	$W - s_i - L$
Final wealth	$W - s_i - \pi$	$W - s_i - \pi - L + I$
Gain/Loss	$-\pi$	$I - \pi$
	Case 2: Comparison with self-insurance (+ self-protection)	
State	No-loss	Loss
Probability	$1 - p_i$	p_i
Reference point	$W - s_i$	$W - s_i - L$
Final wealth	$W - K(g_i) - s_i$	$W - K(g_i) - (1 - g_i) \cdot L - s_i$
Gain/Loss	$-K(g_i)$	$g_i \cdot L - K(g_i)$

Reference point 2: Status quo with self-protection and self-insurance (public good + private good)

	Case 3: Comparison with market insurance (+ self-protection)	
State	No-loss	Loss
Probability	$1 - p_i$	p_i
Reference point	$W - s_i - K(g_i)$	$W - s_i - K(g_i) - (1 - g_i) \cdot L$
Final wealth	$W - s_i - \pi$	$W - s_i - \pi - L + I$
Gain/Loss	$K(g_i) - \pi$	$I - \pi + K(g_i) - L \cdot g_i$

4 Decision-making on cyber risk management under loss aversion

In the expected utility theory, the utility level in any state is determined by the final wealth, which mostly provides a positive value of utility for the state. On the contrary, the value function in prospect theory consists of a positive value from the gain domain and a negative value from the loss domain in accordance with a reference point. Thus, a decision-making problem can be solved by determining the sign of the value function. In other words, if the value function turns out to be positive, a potential decision is preferred over a reference point, whereas the reference point is preferred over the decision if the value function is negative.[103]

1. $V > 0$ *A decision is preferred.*
2. $V \leq 0$ *A reference point is preferred.*

We assume that the decision-maker is a value maximizer with the KT value function to find the optimal decision on risk controls and market insurance, thereby solving the following optimization problem:

$$\max_{z \in \mathcal{H}} V(z), \tag{15}$$

where $z = (p, 1 - p, x_1, x_2)$ in equation (4) and $\mathcal{H}$ is a decision space under prospect theory.

[103] Andalib et al. (2018) also use the sign of the value function to figure out the desirability of the choice. In other words, if the value function is negative, the choice (the scenario in our case) is undesirable. Finding the sign of the value function with the reference point of the average project costs, they evaluate an investment project upon the framework of the real options valuation.

As a benchmark case, the optimal demand for market insurance under the expected utility theory can be examined. Eckles and Volkman-Wise (2019) compare the expected utility theory and prospect theory by taking a power law utility function similar to CRRA utility for the former. This function is comparable with the value function in the gain domain with risk aversion (equation 1), and the diminishing sensitivity parameter, α, can be regarded as the risk aversion parameter in [0,1]. Following Eckles and Volkman-Wise (2019), we derive the optimal decision on market insurance under the expected utility theory with our model framework. The model set-up in Eckles and Volkman-Wise (2019) takes into account the deductible with a loading factor as in this paper, but does not count a self-protection measure.

Our finding is consistent with their finding that agents with self-protection effort under the expected utility theory will purchase full insurance given if a fair premium is provided, whereas partial insurance is preferred in the presence of a loading factor (see Appendix B.1). However, it provides no evidence of agents' behavioral bias toward extreme cyber events, considering them to be unrealistic so they tend to underinvest in risk control measures. In what follows, we establish a conceptual framework under loss aversion assumption in contrast with the classical decision model by evaluating loss and gain compared to the reference point considering the current market circumstance to rationalize this market behavior.

4.1 Case 1: Comparison with the decision on market insurance

The reference point is set as the status quo with self-protection in this case; hence, positive externality works in the interconnected network environment. A public good potentially functions for all relevant parties in the network, and agent i intends to evaluate the optimal decision on self-protection and the purchase of market insurance by comparing two options. In this case, agent i recognizes the no loss state as a loss due to the premium payment, whereas she recognizes the loss state as a gain because of the indemnity payment. With the gain and loss values in Table 2, the expected value under prospect theory is defined as

$$V = (1 - p_i) \cdot v(-\pi) + p_i \cdot v(I - \pi). \tag{16}$$

Clearly, it needs to hold that the indemnity is larger than the premium to derive a gain and a loss of the decision on market insurance ($I > \pi$) where the loss is bigger than the deductible level, leading to the condition $p_i < \frac{1}{1+\delta}$.[104] If purchasing market insurance is preferred, the value

[104] To recognize the loss state as a gain, the outcome of the loss state ($= I - \pi$) must be positive; thus, $I - \pi = (L - d)^+ \cdot \beta - (1 + \delta) \cdot p_i \cdot (L - d)^+ \cdot \beta > 0$, leading to $1 - (1 + \delta) \cdot p_i > 0$.

function must be positive. Applying the KT value function to equation (16), the preferred decision on market insurance is defined as

$$V = -\lambda \cdot (1 - p_i) \cdot \pi^{\alpha} + p_i \cdot (I - \pi)^{\alpha} > 0. \tag{17}$$

Equation (17) can be rewritten as (see Appendix B.2):

$$\left[\frac{1-(1+\delta)\cdot p_i}{(1+\delta)\cdot p_i}\right]^{\alpha} > \lambda \cdot \frac{1-p_i}{p_i}. \tag{18}$$

Inequality (18) holds only if the loading factor is small enough, KT parameters (= λ, α) are also small enough, and the loss probability is large enough.[105] This implies that market insurance with self-protection effort is only preferred for agents, who have less loss aversion (lower λ) and more diminishing sensitivity to the reference point (lower α) when a loss is more likely to occur (higher p).[106] This result is in line with Schmidt's (2016) finding, where the author sets no insurance (the status quo) as the reference point. To provide a clearer idea on these conditions, we derive a numerical example with particular values of parameters, shown in Table 3. We set the loading factor in the range between 10% and 30%[107] and consider different KT parameters to identify the criteria to satisfy inequality (18).[108]

We find that an agent with higher loss aversion (higher λ) and less diminishing sensitivity to the reference point (higher α) tends to not make decisions on purchasing cyber-insurance. This tendency is more prominent as the loading factor is higher. Even for an agent with the opposite propensity (lower loss aversion and more diminishing sensitivity), market insurance is an attractive option for risk transfer only when the loss probability is high enough. This finding demonstrates a behavior that the agent is inclined to stay with self-protection, which is an essential measure in the interconnected business environment. The agent only considers market

105 In addition, it is observed that although inequality (18) depends on the size of the loading factor potentially associated with the coverage level, it is indifferent from the size of deductible, implicitly derived by cancelling out the indemnity from the loss and gain of the state.

106 The parameter for diminishing sensitivity (= α) determines the size of curvature of the value function, where the lower level of the parameter leads to a higher curvature around the reference point in both gain and loss states as described in the left panel of Figure 3. The lower level of the parameter addresses more diminishing sensitivity to the reference point, meaning that the marginal increase in utility (value) is much larger nearer the reference point.

107 Romanosky et al. (2017) observe that the safety loading is usually set between 25% and 35% in the U.S. cyber insurance market, the size of loading that varies between 10% and 30% in other property and casualty lines as evidenced in, for example, Cummins et al. (2006) for property insurance (18% to 26% between 1997 and 2004) in the U.S. and Insurance Europe (2015) for motor insurance (23.2% on average between 2005 and 2013) in Europe.

108 Inequality (18) with no loading factor is identical to Proposition 1 in Schmidt (2016), where the value of the decision is positive if $p_i > \lambda^{1/(1-\alpha)}/(1+\lambda^{1/(1-\alpha)})$. As Schmidt (2016) concludes, purchasing market insurance is not preferred (negative value of the decision) for most realistic cases that have low loss probability.

insurance unfairly priced in the current market if a loss occurrence is nearly certain with the amount above the deductible level. The following proposition summarizes this finding.

Proposition 1: Suppose an agent with the KT value function ($\alpha \in [0,1]$ and $\lambda \geq 1$) makes a decision about purchasing market insurance with continuous effort on self-protection by comparing it to its status quo only with self-protection. The agent is willing to purchase market insurance if both conditions are satisfied; thus, inequality (18) holds, and the loading factor is lower than $\frac{1-\alpha}{\alpha}$.

Proof. See Appendix B.2. ■

Table 3. Criteria for Market Insurance in Case 1

		α				
$\delta = 0.1$		0.1	0.3	0.5	0.7	0.88
	1.25	T ($p_i \geq 0.57$)	T ($p_i \geq 0.61$)	T ($p_i \geq 0.69$)	F	F
	1.75	T ($p_i \geq 0.66$)	T ($p_i \geq 0.73$)	F	F	F
λ	2.25	T ($p_i \geq 0.72$)	T ($p_i \geq 0.81$)	F	F	F
	2.75	T ($p_i \geq 0.77$)	F	F	F	F
	3.00	T ($p_i \geq 0.78$)	F	F	F	F
$\delta = 0.2$		0.1	0.3	0.5	0.7	0.88
	1.25	T ($p_i \geq 0.58$)	T ($p_i \geq 0.64$)	F	F	F
	1.75	T ($p_i \geq 0.67$)	F	F	F	F
λ	2.25	T ($p_i \geq 0.73$)	F	F	F	F
	2.75	T ($p_i \geq 0.78$)	F	F	F	F
	3.00	F	F	F	F	F
$\delta = 0.3$		0.1	0.3	0.5	0.7	0.88
	1.25	T ($p_i \geq 0.58$)	F	F	F	F
	1.75	T ($p_i \geq 0.68$)	F	F	F	F
λ	2.25	F	F	F	F	F
	2.75	F	F	F	F	F
	3.00	F	F	F	F	F

Note: The table shows the results in three cases (10%, 20% and 30%) of loading factor as an example. As the loading factor is bigger, criteria for market insurance shrink further. T indicates that inequality (18) holds with the level of the loss probability in parentheses, whereas F indicates the nonsatisfaction of inequality (18). The loss probability with the validity of inequality (18) should be lower than 91% (=1/(1+δ)).

We next investigate how the self-protection measure affects the value of the decision. The optimization problem for the agent with equation (16) can be resolved with respect to the level of self-protection. The following proposition reveals the decision on self-protection effort when the agent is considering an insurance option in the decision-making process.

Proposition 2: Suppose that an agent in the interdependent network environment is a value maximizer with the expected value in equation (16). The marginal value of the agent decreases over the level of self-protection if the following condition is satisfied:

$$\frac{\alpha}{1+\alpha} < p_i < \frac{1}{(1+\alpha)(1+\delta)}, \tag{19}$$

where α indicates the parameter of diminishing sensitivity homogeneous in both gain and loss spaces, as shown by Tversky and Kahneman (1992). Conversely, increasing marginal value is observed over the level of self-protection if the following condition is satisfied:

$$\frac{1}{(1+\alpha)(1+\delta)} < p_i < \frac{\alpha}{1+\alpha}. \tag{20}$$

Proof. See Appendix B.3. ■

Proposition 2 leads to the link between the level of self-protection and insurance premium that the decreasing marginal value with the level of self-protection holds when the loading factor is smaller than $\frac{1-\alpha}{\alpha}$, whereas the increasing marginal value holds with a higher loading factor (see Appendix B.3). For instance, applying the KT parameter ($\alpha = 0.88$), we identify that increasing the self-protection effort results in declining marginal value if the loading on the insurance premium is lower than 13.6%. Intuitively, it describes a relationship that, by comparing the insurance decision with the reference point of self-protection, the agent in the KT prospect theory world is more incentivized to increase self-protection if a highly loaded premium is given. This supports the finding in Table 3 in that inequality (18) does not hold as the loading factor is higher, showing that the value of the decision on market insurance becomes negative.

As illustrated in the right panel of Figure 3, the area above the line indicates a positive marginal value reflecting an incentive to increase self-protection with highly loaded insurance scenarios. Proposition 2 also holds when the first order condition is investigated with respect to the level of self-protection by interconnect agent j.[109] Thus, the self-protection measure and the insurance decision are substitutes for an agent with the reference point of self-protection, which is counter to Ehrlich and Becker (1972) under expected utility theory. In addition, positive externality from self-protection as a public good is more effective, particularly in a situation where agents cannot find affordable insurance policy in the market.

[109] The optimization problem is symmetrical between agent *i* and interconnected agent *j*, meaning that the result of the first order condition with respect to the level of self-protection by the interconnected agent is identical to that with respect to agent *i*'s self-protection level.

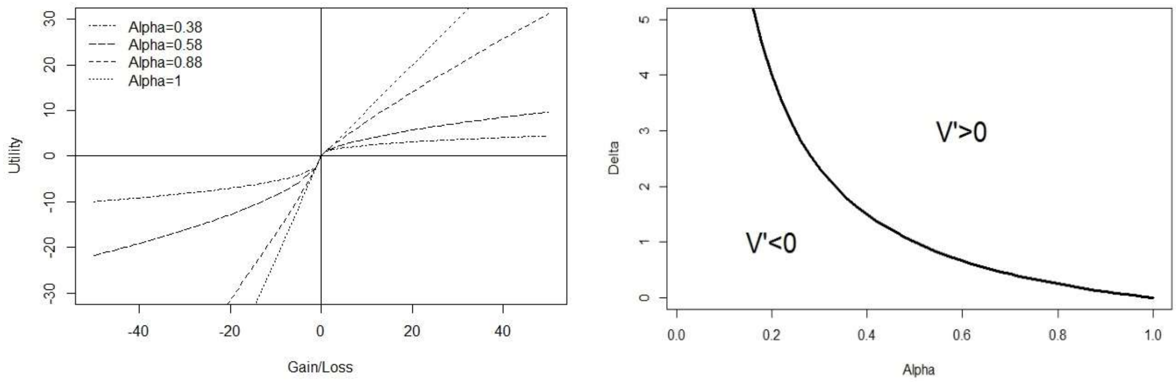


Figure 3. The description on varying diminishing sensitivity and the marginal productivity of self-protection measure when considering market insurance. The left panel displays the KT value function with different sizes of the parameter for diminishing sensitivity, and the right panel shows the sign of the marginal value with respect to the effort on security to describe the incentive for self-protection.

Following Schmidt (2016), we graphically describe indifference curves of both decision outcomes: maintaining the status quo and purchasing market insurance. We find the marginal productivity of the self-insurance coverage by solving the optimization problem with regard to β as follows:

$$\frac{\partial V}{\partial \beta} = (1 - p_i) \cdot v_0' \cdot (-\pi') + p_i \cdot v_1' \cdot [(1 - (1 + \delta) \cdot p_i) \cdot I'] = 0, \tag{21}$$

where v_0' and v_1' indicate the marginal values in loss and gain domains, respectively.

The optimization problem in equation (21) can be solved by satisfying the following first order condition:

$$-\frac{(1 - p_i) \cdot v_0'}{p_i \cdot v_1'} = \frac{(1 - (1 + \delta) \cdot p_i) \cdot I'}{-\pi'} = \frac{\mathrm{d}x_1}{\mathrm{d}x_0} < 0, \tag{22}$$

where $\mathrm{d}x_0$ and $\mathrm{d}x_1$ are marginal values in the loss and gain domains, respectively, and $I', \pi' > 0$.

In both loss and gain domains, the marginal utilities, v_0' and v_1', are positive as observed in Figure 1. The right hand side of equation (22) addresses the slope of the indifference curve maximizing the expected value; thus, the above condition implies a negative slope of the indifference curve in a (x_0, x_1)-space when considering the decision on market insurance in the existence of self-protection. According to Lemma 1 in Schmidt (2016), the curvature of the KT value function can be determined by the sign of the value, V, showing the opposite sign between the value function and the second order condition. Thus, assuming that $p_i < \frac{1}{1+\delta}$ holds, the indifference curve is concave (linear or convex, respectively) if $V > 0$ (= or < 0, respectively), meaning that inequality (18) holds. We depict the indifference curves of both cases, where staying with self-protection only is optimal and purchasing market insurance is preferred, as shown in Figure 4.

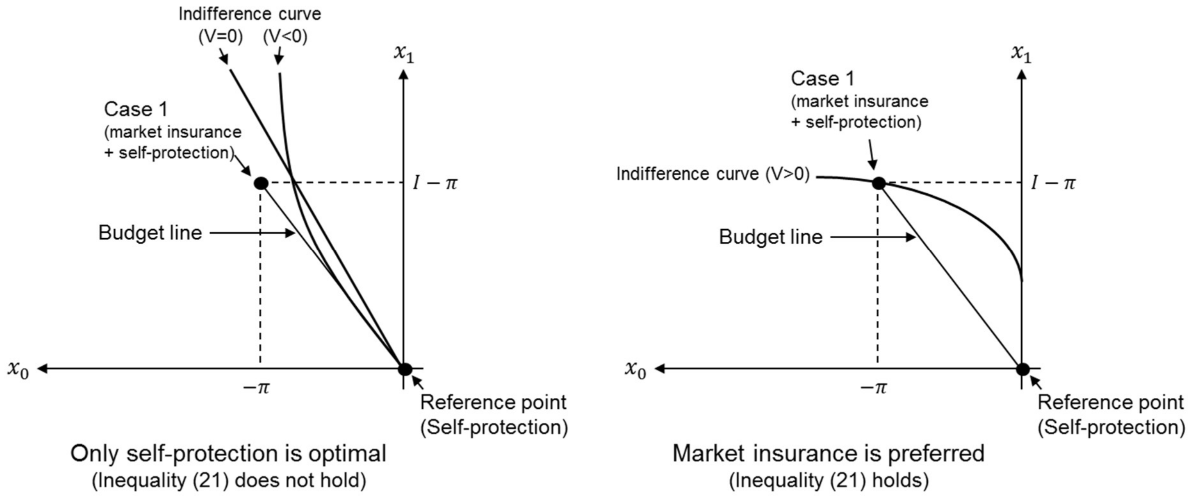


Figure 4. Indifference curves based on Proposition 1. x_0 is the outcome in the loss state, and x_1 is the outcome in the gain state. The left panel displays indifference curves describing the optimality of staying on the reference point, whereas the right panel shows the indifference curve of the preference for market insurance. We follow Schmidt (2016) for these plots.

The (x_0, x_1)-space lies in the second quadrant because the loss state (x_0) is negative and the gain state (x_1) is positive to achieve Proposition 1. In the left panel, staying with the reference point is optimal so that the indifference curve through the origin is either linear (V = 0) or convex (V < 0) above the point indicating the preference for market insurance. The right panel explains the optimality of the preference for market insurance with the concave indifference curve (V > 0), which lies above the origin.

4.2 Case 2: Comparison with the decision on self-insurance as a private good

In this case, we compare the decision for self-insurance on top of self-protection measure with the reference point. Agent *i* recognizes the no loss state as the occurrence of a loss due to the cost for self-insurance. The expected value of this case is defined as

$$V = (1-p_i)\cdot v\big(-K(g_i)\big) + p_i \cdot v\big(g_i \cdot L - K(g_i)\big). \tag{23}$$

For the self-insurance plan to be effective in this decision-making, $g_i \geq \frac{K(g_i)}{L}$ needs to hold so that the expected value is determined in both loss and gain states (see Appendix B.4). In other words, this condition implies that the self-insurance coverage above a certain level is required to achieve the validity of the decision-making problem in self-insurance under loss aversion.

We evaluate the necessary condition on the positive value addressing the tendency for self-insurance. The preferred decision on self-insurance is defined as

$$V = -\lambda \cdot (1 - p_i) \cdot K(g_i)^\alpha + p_i \cdot \left(g_i \cdot L - K(g_i)\right)^\alpha > 0. \tag{24}$$

We rewrite equation (24) as follows (see Appendix B.4):

$$p_i > \frac{1}{\left(\frac{g_i \cdot L}{K(g_i)} - 1\right)^\alpha / \lambda + 1}, \tag{25}$$

where $\frac{g_i \cdot L}{K(g_i)}$ accounts for the ratio between the self-insurance endowment and its cost; thus, $\frac{g_i \cdot L}{K(g_i)} - 1$ can be said to be the net margin of self-insurance. Under the assumption that self-insurance is effective in this case (i.e., $g_i \geq \frac{K(g_i)}{L}$), net margin should be positive.

Net margin of self-insurance depends on the magnitude of a loss so that a larger loss results in a bigger net margin.[110] As can be observed in Table 4, a more loss averse agent (higher λ) with less diminishing sensitivity to the reference point (higher α) is more likely to not invest in self-insurance, and this propensity is stronger as the net margin of self-insurance becomes lower (i.e., the criterion on the loss probability in inequality (25) is higher). However, the criterion on the loss probability to prefer self-insurance is still high in any case (at least above 0.5). This finding is clarified in Proposition 3.

Proposition 3: Suppose an agent with the KT value function ($\alpha \in [0,1]$ and $\lambda \geq 1$) decides to implement self-insurance with continuous effort on self-protection by comparing it to its status quo only with self-protection. The agent is willing to implement self-insurance if both conditions are satisfied—namely, inequality (25) holds, and $g_i \geq \frac{K(g_i)}{L}$.

Proof. See Appendix B.4. ■

In order to find the marginal productivity of the self-insurance coverage, an optimization problem with regard to g_i reveals

---

[110] Let us denote the net margin function by $A(L, g_i) = \frac{g_i \cdot L}{K(g_i)} - 1$. The first order condition of net margin with respect to the loss variable is

$$\frac{\partial A(L, g_i)}{\partial L} = \frac{g_i}{K(g_i)} > 0.$$

It shows that self-insurance can be more effective when the size of a loss is substantial.

$$\frac{\partial V}{\partial g_i} = (1-p_i)\cdot v_0' \cdot [-K(g_i)'] + p_i \cdot v_1' \cdot (L - K(g_i)') = 0, \tag{26}$$

where v_0' and v_1' indicate the marginal values in loss and gain domains, respectively.

The optimization problem in equation (26) can be solved by satisfying the following first order condition:

$$-\frac{(1-p_i)\cdot v_0'}{p_i \cdot v_1'} = \frac{L - K(g_i)'}{-K(g_i)'} = \frac{\mathrm{d}x_1}{\mathrm{d}x_0} < 0, \tag{27}$$

where $\mathrm{d}x_0$ and $\mathrm{d}x_1$ are marginal values in the loss and gain domains, respectively.

Table 4. Criteria for Self-insurance in Case 2

		α				
Net margin=10%		0.1	0.3	0.5	0.7	0.88
	1.25	0.611	0.714	0.798	0.862	0.905
	1.75	0.688	0.777	0.847	0.898	0.930
λ	2.25	0.739	0.818	0.877	0.919	0.945
	2.75	0.776	0.846	0.897	0.932	0.954
	3.00	0.791	0.857	0.905	0.938	0.958
Net margin=50%		0.1	0.3	0.5	0.7	0.88
	1.25	0.573	0.606	0.639	0.670	0.697
	1.75	0.652	0.683	0.712	0.740	0.763
λ	2.25	0.707	0.735	0.761	0.785	0.805
	2.75	0.747	0.772	0.795	0.817	0.835
	3.00	0.763	0.787	0.809	0.830	0.847
Net margin=100%		0.1	0.3	0.5	0.7	0.88
	1.25	0.556	0.556	0.556	0.556	0.556
	1.75	0.636	0.636	0.636	0.636	0.636
λ	2.25	0.692	0.692	0.692	0.692	0.692
	2.75	0.733	0.733	0.733	0.733	0.733
	3.00	0.750	0.750	0.750	0.750	0.750

Note: The table shows the critical values of the loss probability to achieve inequality (25) with varying possible values of net margin as examples (the condition holds if the loss probability is higher than the critical value for each case). The bigger the net margin is, the lower the loss probability level is in inequality (25). However, the level of the loss probability to satisfy a positive value of the decision on self-insurance is considerably high for all three cases.

In both loss and gain domains, the marginal values, v_0' and v_1', are positive as observed in Figure 1, and the marginal cost of self-insurance ($= K(g_i)'$) is also positive due to the concavity. Thus, based on the assumption that $g_i \geq \frac{K(g_i)}{L}$ holds, the indifference curve is concave (linear or convex respectively) if $V > 0$ (= or < 0 respectively), meaning that inequality (25) holds.

As in Figure 4, we again depict indifference curves of both cases, where staying only with self-protection is optimal and implementing self-insurance is preferred (see Figure 5). We only focus on the second quadrant, where the loss state (x_0) is negative and the gain state (x_1) is positive to achieve Proposition 3. The reference point is optimal in the left panel, with either the linear

(V = 0) or convex (V < 0) indifference curve above the point indicating the preference for self-insurance. In contrast, self-insurance is preferred in the right panel with the concave indifference curve (V > 0), which lies above the origin. The finding in this section also supports the evidence of a fatalistic behavior that the agent is inclined to stay with self-protection as a basic effort required in the interconnected business environment. The agent takes into consideration self-insurance coverage above a certain level only if the loss probability is substantial.

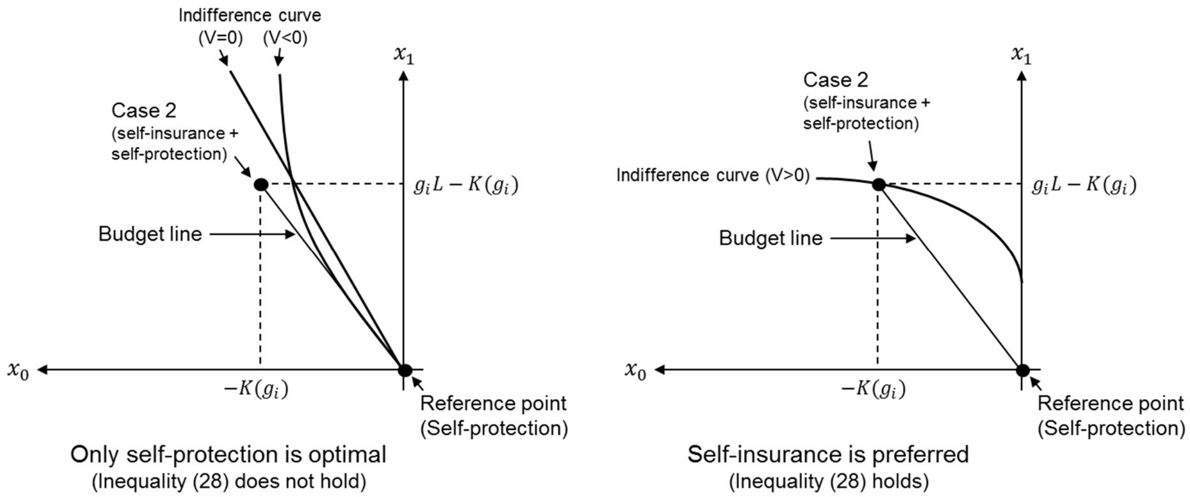


Figure 5. Indifference curves based on Proposition 3. The left panel displays indifference curves describing the optimality of staying on the reference point, whereas the right panel shows the indifference curve of the preference for self-insurance. We follow Schmidt (2016) for these plots.

Figure 6 illustrates inequality (25) by varying the diminishing sensitivity parameter (= α) and the self-insurance level (= g_i) under the assumption that $g_i \geq \frac{K(g_i)}{L}$ is satisfied. We isolate the effects of other parameters by setting examples—namely, panel A: $L = 100{,}000$, $K(g_i) = 10{,}000$, $\lambda = 2.25$; panel B: $L = 100{,}000$, $K(g_i) = 10{,}000$, $g_i = 0.2$; panel C: $L = 100{,}000$, $K(g_i) = 10{,}000$, $\alpha = 0.88$. For each example, we focus on the interaction of varying parameters with the loss probability. The result supports the finding that, below a certain level of the loss probability, an agent is not willing to implement a self-insurance plan, although the level varies with the parameters. In panels B and C, we vary the loss aversion parameter (= λ) above one to reflect a loss averse agent—that is, if $\lambda < 1$, she is less averse against a loss than against a gain.

For instance, as panel A indicates, an agent with an already high loss aversion (= 2.25) is more indifferent from the level of self-insurance coverage as she has more diminishing sensitivity to the reference point (lower α). In addition, panels B and C show that an agent with higher loss aversion (higher λ) tends to give up more self-insurance on controlling risk, thereby showing a more distinct behavioral bias. Here, higher loss aversion addresses higher risk seeking in the loss domain, as described in Section 2.

Panel A: (g_i, α)-space   Panel B: (α, λ)-space   Panel C: (g_i, λ)-space

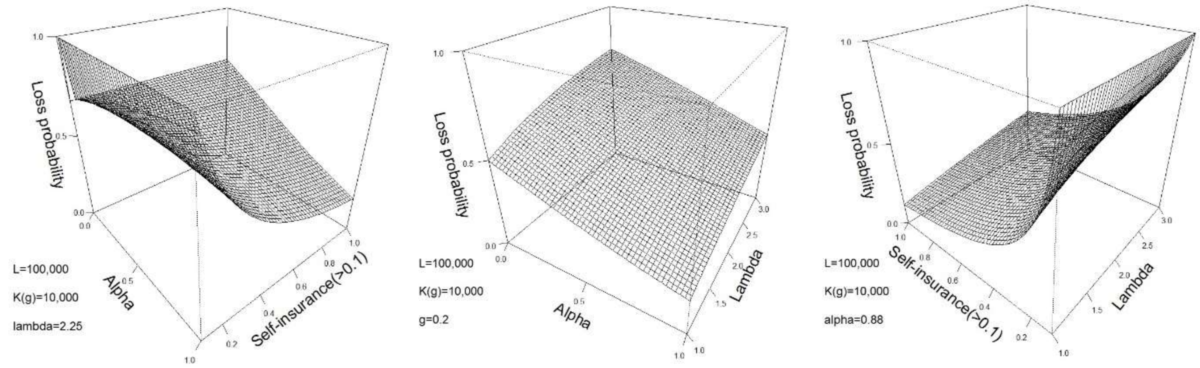


Figure 6. Three-dimensional plots representing Proposition 2.

4.3 Case 3: Status quo with self-protection and self-insurance

An agent could make the maximum effort by itself, which can provide a safe option without any transfer measure. Based on this possible maximum effort, the agent evaluates decision-making outcome by switching from self-insurance to market insurance when self-protection continues to be implemented as a basic management option. This evaluation is relevant under the assumption that market insurance and self-insurance are substitutes for each other; hence, the agent might take into account the difference in the cost and the value of the decision between two options. The expected value of this case is defined as

$$V = (1 - p_i) \cdot v(K(g_i) - \pi) + p_i \cdot v(I - g_i \cdot L + K(g_i) - \pi). \tag{28}$$

To see the indifference curve in the second quadrant, the signs of the gain and the loss should be opposite each other, which splits equation (28) into two cases: 1) $K(g_i) - \pi \leq 0, I - g_i \cdot L + K(g_i) - \pi > 0$ and 2) $K(g_i) - \pi > 0, I - g_i \cdot L + K(g_i) - \pi \leq 0$.[111]

The first case indicates that market insurance charges a higher premium to provide higher coverage for loss than the cost of self-insurance with lower coverage.[112] The second case is the

[111] As the size of the deductible gets closer to 0, the first case is more likely to appear, which can be shown as

$$1)\ K(g_i) - \pi = K(g_i) - (L-d)^+ \cdot \beta \cdot (1+\delta) \cdot p_i \leq 0,$$

$$2)\ I - g_i \cdot L + K(g_i) - \pi = I - \pi + [K(g_i) - g_i \cdot L] = (1 - (1+\delta) \cdot p_i) \cdot (L-d)^+ \cdot \beta > 0.$$

It can result in higher critical values of the loss probability, as in panel A of Table 5, which implies that an agent is less likely to switch its self-insurance plan to market insurance when the premium level and the size of indemnity become higher, but the transition to market insurance is not effective (low net margin).

[112] To prove this, the following inequality should hold

$$|I - g_i \cdot L| > |K(g_i) - \pi|,$$

where $I - g_i \cdot L > 0$ as $K(g_i) - \pi \leq 0$ and $I - g_i \cdot L + K(g_i) - \pi > 0$. Then, the right side of this inequality can be rewritten opposite instance, where self-insurance with more coverage costs higher than market insurance with lower coverage does. The following proposition summarizes decision-making outcomes of two possible cases in this scenario.

Proposition 4: Suppose an agent with the KT value function ($\alpha \in [0,1]$ and $\lambda \geq 1$) decides to purchase market insurance with continuous effort on self-protection by comparing it to the maximum effort (self-protection and self-insurance). The agent is willing to purchase market insurance instead of self-insurance if the following conditions are satisfied:

For market insurance

i) Positive net effect of the transition from self-insurance to market insurance:	$\left[\frac{I - g_i \cdot L}{\pi - K(g_i)} - 1\right]^\alpha > \lambda \cdot \frac{1 - p_i}{p_i}$
ii) Positive net effect of staying with self-insurance:	$\left[\frac{g_i \cdot L - I}{K(g_i) - \pi} - 1\right]^\alpha < \frac{1}{\lambda} \cdot \frac{p_i}{1 - p_i}$

Proof. See Appendix B.5. ■

The equations in the parentheses on the left hand side can explain the net effect of choosing a particular insurance type; the equation in the first condition addresses the net effect of the transition from self-insurance to market insurance, whereas the one in the second condition indicates the net effect of staying with self-insurance. These net effects are all positive with the inequality assumptions from equation (28). Table 5 shows the criteria for the loss probability that makes the conditions of Proposition 4 hold. For the first condition in panel A, the numerical results are identical to those in Table 4 (Case 2). This result implies that a more loss averse agent (higher λ) with less diminishing sensitivity to the reference point (higher α) is more likely to stay with self-insurance; this tendency is stronger as the net effect of the transition from self-insurance to market insurance decreases. The results in panel B of Table 5 from the second condition are exactly in line with this finding by showing the reverse outcomes in that a more loss averse agent (higher λ) with less diminishing sensitivity to the reference point (higher α) is more likely to take up market insurance instead of self-insurance; it tends to be stronger as the net effect of staying with self-insurance decreases.

$$I - g_i \cdot L = (\beta - g_i) \cdot L - d \cdot \beta > 0.$$

Thus, $\beta - g_i > 0$ should hold.

Table 5. Criteria for Replacing Self-insurance by Market Insurance in Case 3

Panel A: The critical values of the loss probability in the first condition of Proposition 4

		α				
Net effect=10%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.61$)	T($p_i \geq 0.71$)	T($p_i \geq 0.80$)	T($p_i \geq 0.86$)	T($p_i \geq 0.91$)
	1.75	T($p_i \geq 0.69$)	T($p_i \geq 0.78$)	T($p_i \geq 0.85$)	T($p_i \geq 0.90$)	T($p_i \geq 0.93$)
λ	2.25	T($p_i \geq 0.74$)	T($p_i \geq 0.82$)	T($p_i \geq 0.88$)	T($p_i \geq 0.92$)	T($p_i \geq 0.95$)
	2.75	T($p_i \geq 0.78$)	T($p_i \geq 0.85$)	T($p_i \geq 0.90$)	T($p_i \geq 0.93$)	T($p_i \geq 0.95$)
	3.00	T($p_i \geq 0.79$)	T($p_i \geq 0.86$)	T($p_i \geq 0.91$)	T($p_i \geq 0.94$)	T($p_i \geq 0.96$)
Net effect=50%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.57$)	T($p_i \geq 0.61$)	T($p_i \geq 0.64$)	T($p_i \geq 0.67$)	T($p_i \geq 0.70$)
	1.75	T($p_i \geq 0.65$)	T($p_i \geq 0.68$)	T($p_i \geq 0.71$)	T($p_i \geq 0.74$)	T($p_i \geq 0.76$)
λ	2.25	T($p_i \geq 0.71$)	T($p_i \geq 0.74$)	T($p_i \geq 0.76$)	T($p_i \geq 0.79$)	T($p_i \geq 0.81$)
	2.75	T($p_i \geq 0.75$)	T($p_i \geq 0.77$)	T($p_i \geq 0.80$)	T($p_i \geq 0.82$)	T($p_i \geq 0.84$)
	3.00	T($p_i \geq 0.76$)	T($p_i \geq 0.79$)	T($p_i \geq 0.81$)	T($p_i \geq 0.83$)	T($p_i \geq 0.85$)
Net effect=100%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)
	1.75	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)
λ	2.25	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)
	2.75	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)
	3.00	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)

Panel B: The critical values of the loss probability in the second condition of Proposition 4

		α				
Net effect=10%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.50$)	T($p_i \geq 0.39$)	T($p_i \geq 0.28$)	T($p_i \geq 0.20$)	T($p_i \geq 0.14$)
	1.75	T($p_i \geq 0.58$)	T($p_i \geq 0.47$)	T($p_i \geq 0.36$)	T($p_i \geq 0.26$)	T($p_i \geq 0.19$)
λ	2.25	T($p_i \geq 0.64$)	T($p_i \geq 0.53$)	T($p_i \geq 0.42$)	T($p_i \geq 0.31$)	T($p_i \geq 0.23$)
	2.75	T($p_i \geq 0.69$)	T($p_i \geq 0.58$)	T($p_i \geq 0.47$)	T($p_i \geq 0.35$)	T($p_i \geq 0.27$)
	3.00	T($p_i \geq 0.70$)	T($p_i \geq 0.60$)	T($p_i \geq 0.49$)	T($p_i \geq 0.37$)	T($p_i \geq 0.28$)
Net effect=50%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.54$)	T($p_i \geq 0.50$)	T($p_i \geq 0.47$)	T($p_i \geq 0.44$)	T($p_i \geq 0.40$)
	1.75	T($p_i \geq 0.62$)	T($p_i \geq 0.59$)	T($p_i \geq 0.55$)	T($p_i \geq 0.52$)	T($p_i \geq 0.49$)
λ	2.25	T($p_i \geq 0.68$)	T($p_i \geq 0.65$)	T($p_i \geq 0.61$)	T($p_i \geq 0.58$)	T($p_i \geq 0.55$)
	2.75	T($p_i \geq 0.72$)	T($p_i \geq 0.69$)	T($p_i \geq 0.66$)	T($p_i \geq 0.63$)	T($p_i \geq 0.60$)
	3.00	T($p_i \geq 0.74$)	T($p_i \geq 0.71$)	T($p_i \geq 0.68$)	T($p_i \geq 0.65$)	T($p_i \geq 0.62$)
Net effect=100%		0.1	0.3	0.5	0.7	0.88
	1.25	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)	T($p_i \geq 0.56$)
	1.75	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)	T($p_i \geq 0.64$)
λ	2.25	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)	T($p_i \geq 0.69$)
	2.75	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)	T($p_i \geq 0.73$)
	3.00	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)	T($p_i \geq 0.75$)

Note: The table shows the criteria for the loss probability to achieve the conditions of Proposition 4 with varying possible values of net effect as examples (the conditions hold if the loss probability is higher than the critical value for each case). T indicates that two conditions of Proposition 4 in panels A and B hold with the level of the loss probability in the corresponding parentheses. The net effect in panel A indicates the effect of the transition from self-insurance to market insurance, and the one in panel B accounts for the effect of staying with self-insurance.

In Figure 7, we again draw indifference curves of both decision outcomes—namely, maintaining self-insurance and replacing it by market insurance. The marginal productivity of market insurance can be obtained by solving the optimization problem in terms of β with two cases in loss and gain domains as follows:

1) $K(g_i) - \pi \leq 0$ and $I - g_i \cdot L + K(g_i) - \pi > 0$:

$$\frac{\partial V}{\partial \beta} = (1 - p_i) \cdot v_0' \cdot (-\pi') + p_i \cdot v_1' \cdot [(1 - (1 + \delta) \cdot p_i) \cdot I'] = 0, \tag{29}$$

2) $K(g_i) - \pi > 0$ and $I - g_i \cdot L + K(g_i) - \pi \leq 0$:

$$\frac{\partial V}{\partial \beta} = (1 - p_i) \cdot v_0' \cdot [(1 - (1 + \delta) \cdot p_i) \cdot I'] + p_i \cdot v_1' \cdot (-\pi') = 0, \tag{30}$$

where v_0' and v_1' indicate the marginal values in loss and gain domains respectively.

These optimization problems are solved by satisfying the following first order conditions:

1) $K(g_i) - \pi \leq 0$ and $I - g_i \cdot L + K(g_i) - \pi > 0$:

$$-\frac{(1 - p_i) \cdot v_0'}{p_i \cdot v_1'} = \frac{(1 - (1 + \delta) \cdot p_i) \cdot I'}{-\pi'} = \frac{\mathrm{d}x_1}{\mathrm{d}x_0} < 0, \tag{31}$$

2) $K(g_i) - \pi > 0$ and $I - g_i \cdot L + K(g_i) - \pi \leq 0$:

$$-\frac{(1 - p_i) \cdot v_0'}{p_i \cdot v_1'} = \frac{-\pi'}{(1 - (1 + \delta) \cdot p_i) \cdot I'} = \frac{\mathrm{d}x_1}{\mathrm{d}x_0} < 0, \tag{32}$$

where $\mathrm{d}x_0$ and $\mathrm{d}x_1$ are marginal values in the loss and gain domains, respectively, and $I', \pi' > 0$.

Based on Lemma 1 in Schmidt (2016), we determine the curvature of the KT value function in Figure 7 using the sign of the value, V, showing that the indifference curve is concave (linear or convex respectively) if $V > 0$ (= or < 0, respectively), satisfying Proposition 4.

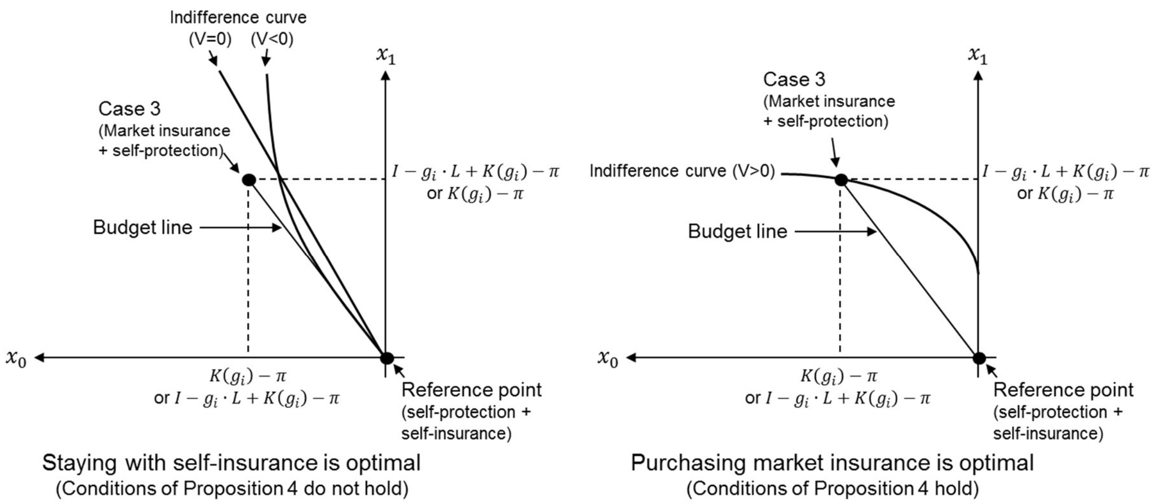


Figure 7. Indifference curves based on Proposition 4. The left panel displays indifference curves describing the optimality of staying on the reference point, whereas the right panel shows the indifference curve of the preference for market insurance instead of self-insurance. We follow Schmidt (2016) for these plots.

5 Empirical support and implication for policy

5.1 Empirical evidence on loss probability

In each of the three cases, we find the necessary conditions to show that a decision considered is preferred to the reference point. These conditions require a certain level of loss probability so that an agent decides to implement additional risk management activities (either market insurance or self-insurance). Although the level of probability leading the agent to carry out additional actions on risk management is relatively high in all cases, the recent literature identifies that the frequency of cyber loss events shows significantly faster growth, particularly for larger losses, over the last decade (Hofmann, Wheatley and Sornette, 2019).[113] Thus, the recent trend can make it more likely to achieve the conditions of propositions. To obtain some relevant empirical indications, we estimate appropriate loss probabilities using a publicly available database for cyber risk provided by Private Rights Clearinghouse (PRC).[114]

PRC collects information on data breach events and categorizes losses by eight types of breach (PRC, 2019), among which we only focus on four types of breach (hacking, insider breach, payment card fraud and unknown) as a malicious type of risk by following the categorization in Edwards et al. (2016). The malicious type of risk is more relevant to our model framework, particularly with regard to the presence of interdependent risk. The definition of cyber risk is generally broader than data breach risk (Eling and Wirfs, 2019), but here we restrict it to data breach loss, which is one of the main coverages in most cyber-insurance policies (Romanosky

---

[113] Hofmann, Wheatley and Sornette (2019) update the results of Wheatley et al. (2016) for statistical features of cyber risk with larger and longer datasets between 2005 and 2017 from the Private Rights Clearinghouse (PRC) database and Open Security Foundation Dataloss database formerly used in Wheatley et al. (2016). They find a significant increase in cyber loss frequency by using generalized linear model (GLM) with log-linear negative binomial function. They specifically show 8% annual growth for data breach losses over 10,000 and 19% growth for losses over 1 million during the data period.

[114] PRC is a non-profit organization collecting and updating information on data breach events published in the media. Each loss observation can at least be found by either a government agency or a verifiable media source. With this effort, PRC forms the largest public database for data breaches that incorporates information on the breach date, the event location, the entity level, the loss type and the total number of breached records. Despite the provision of the largest data pool for cyber risk, the PRC database has several limitations; there might be a possibility of a bias in the data completeness (backfilling bias) so that the data collected more recently might be relatively less complete than the data from an older period. This bias might underestimate the frequency rate for the recent years (e.g., 2017, 2018) in Table 6, although the impact of the bias might be small because the database is continuously updated on a daily basis with the reliable data source standard (see, e.g., Eling and Wirfs, 2019, for a similar argumentation with SAS OpRisk database). Second, some observations have a zero value, indicating that the corresponding loss events are not publicly acknowledged or are still under investigation (Eling and Jung, 2018). We exclude zero values in our application to minimize the self-selection bias, which—according to Kamiya et al. (2018) using the same dataset—does not meet cyberattack notification laws and it might not be compulsory for firms to disclose information. However, this exclusion could clearly underestimate our application; hence, we need to make sure that the actual frequency rate might be higher than what is estimated with the non-zero values.

et al., 2017). Over the entire data period from 2005 to 2018, we find 4,053 observations related to malicious risk (as of February 5, 2019), among which we use the 2,774 non-zero records.

In Table 6, we estimate the number of days experiencing at least one loss event per year over the data period between 2005 and 2018. From the insurer's perspective, a loss frequency can be estimated by the ratio between the number of claims and the number of risks in the pool. However, from the insured perspective, it is difficult to measure a possible rate of loss frequency without the availability of loss data and information on the risk pool. As a proxy for the rate of loss frequency from an insured perspective, we attempt to estimate the number of days when at least one breach loss is publicly announced over the period. The rate for each year is measured upon different deductible levels ($5,000 as lowest to $1 million, at $250,000 increments) subject to possible cyber-insurance coverages, using the deductibles based on the finding in Romanosky et al. (2017) as described in Section 3.2.

Importantly, the observations in the PRC database account for only the number of breach records; thus, we need to translate them to the monetary unit to apply the deductibles. Studies on the measurement of the economic cost for data breach are lacking, but the Ponemon Institute (2018) reports its analysis on the cost of data breaches on an annual basis. Thus, we adapt the average economic cost on a yearly basis as estimated by the Ponemon Institute to obtain approximate measures of economic loss from data breach events. Table 6 shows the frequency rate per year from 2005 to 2018 with five deductible levels. Although it does not exactly reflect the actual loss probability, we can observe that the rate is significantly increasing, showing, for example, a threefold increase over the last decade (from 2008 to 2018 across all deductible categories). It implies that the dynamic landscape of cyber risk with a dramatic increase in loss frequency can lift up the demand for additional risk management tools (market insurance or self-insurance), resulting in a demand that is higher by a less averse agent against a loss.

For further evidence[115] on the increasing frequency rate of cyber loss events that might heavily affect the subjective probability of potential insureds, we list anecdotal statements and empirical clues in Appendix C. The list demonstrates that cyber loss events and attacks are more likely to occur in our daily lives and business, particularly in recent years, which can result in a significant increase in the subjective probability of potential insureds against cyber risk. Overall,

[115] Apart from evidence provided by practical reports, several concrete estimates can be found in academic studies. For example, Edwards et al. (2016) estimate the probability of the breach size above 1, 5 and 10 million in the next one year to be 99.3%, 75.6% and 53.6%, respectively, under the assumption that the status quo is maintained. Eling and Schnell (2018) estimate the annual frequency rate of 7.4% by applying the data breach dataset on the S&P 500 companies.

our finding implies that although a behavioral bias against cyber risk still exists in the market, it becomes more inevitable to implement additional plans for risk management as agents are exposed to increasing frequency rate each year, thereby leading the cyber-insurance market to stay on track.

Table 6. Empirical Frequency Rate of Data Breach Loss (per year)

	Deductible				
Year	\$5,000	\$250,000	\$500,000	\$750,000	\$1,000,000
2005	14.0%	11.5%	11.0%	10.4%	9.3%
2006	19.7%	14.5%	13.2%	11.0%	9.6%
2007	18.1%	12.3%	11.5%	11.0%	9.9%
2008	18.9%	13.9%	12.3%	10.1%	9.8%
2009	13.7%	8.2%	7.4%	7.4%	7.4%
2010	28.2%	16.2%	12.9%	11.8%	11.8%
2011	37.5%	20.0%	15.1%	13.7%	12.9%
2012	42.1%	25.1%	19.7%	16.4%	15.3%
2013	32.9%	24.7%	21.1%	17.8%	15.6%
2014	32.1%	21.6%	17.8%	16.7%	15.9%
2015	20.0%	16.2%	14.5%	13.4%	12.3%
2016	36.6%	29.5%	24.6%	23.5%	22.4%
2017	36.4%	31.0%	28.5%	24.9%	22.7%
2018	59.7%	44.7%	38.9%	34.8%	32.3%

Note: This table shows a frequency rate per year of data breach loss over the period between 2005 and 2018. The data is derived from PRC and shows the percentage of days with a loss event larger than the deductible.

5.2 Implication for cyber risk management

Self-protection as the status quo for potential cyber-insureds not only reflects the current market situation, but also has some regulatory implications. Self-protection facilitates the reinforcement of the security system of all interconnected parties (positive externality); hence, it can function as a public good in the interconnected network environment. The government can utilize this characteristic to increase the overall security level against cyber risk in society by forcing all business agents to meet a certain level of security requirement. This regulation should be backed by an appropriate assessment of a security system by the governmental agency to evaluate whether business parties applying this regulation appropriately implement cyber security measures. As another form of public intervention, the government might be able to incentivize business parties to enhance the security system by providing a partial tax exemption or subsidy for the cost of protective measures, particularly for small companies.

For example, the Internal Revenue Service (IRS), a governmental agency in the U.S., offers a tax deduction for individuals and firms who have experienced enormous losses due to catastrophes. The Federal Emergency Management Agency (FEMA) in the U.S. financially supports individuals and businesses by subsidizing, for example, flood insurance premiums through thousands of insurance agents in the nation, which is called FEMA's National Flood Insurance Program (NFIP). This program also allows insureds to purchase more than the limit that the NFIP requires—an endorsement called excess flood protection. This type of regulation (public intervention) can also help cyber-insurers reduce information asymmetry and underwrite cyber-insurance with an optimal deductible and premium, possibly containing a comprehensive risk management offer.[116] Therefore, in the cyber risk context, we can base the status quo on the implementation of self-protection.

Another way to force cyber security enhancement for business parties could be a monopolistic insurer (Mürmann and Kunreuther, 2008) or a co-insurance scheme provided by multiple insurers. The compulsory implementation of protective measures could be required by a monopolistic insurer or possibly oligopolistic insurers with an economic incentive for premiums more easily than in the competitive market. An insurer in the competitive market might not be able to impose the requirement for the protective measures as their implementation will cost potential insureds, which might lead the insurer to lose its market share. For instance, von Ungern-Sternberg (1996) observe that monopolistic state insurers in Switzerland (19 cantons) with much lower damage rates offer almost 70% lower premiums for housing insurance than private insurers (seven cantons) in the competitive market do. An oligopoly market can also drive potential insureds to make more effort on preventive measures with the price reduction. The assumption on an oligopoly market for cyber-insurance might make sense in the current market situation (particularly in the U.S., as the largest market), where three insurers out of 500 cyber-carriers account for nearly 50% of the market share and almost 10 players address 90% (Romanosky et al., 2017).

In a similar context, co-insurance[117] can be another market-wise solution to enhance cyber risk management. Specifically, co-insurance can play the role of reducing the possibility of moral hazard by, for example, building up a homogenized risk assessment in the underwriting process (ex-ante) and ex-post joint assessment by insurers in the co-insurance pool, which can reduce the cost for insurers to monitor a possible moral hazard behavior. Markets such as Lloyd's can help insurers share catastrophe risks via a co-insurance scheme depending on, for example, the financial capacity or the extent of understanding (cyber) risk.

[116] Some cyber-insurance policies in the current market offer a so-called risk management package, which incorporates from building a detection system and a reporting process to the monetary coverage (e.g., Allianz Cyber Protect and HSB Cyber Insurance by Munich Re).

[117] Co-insurance scheme here indicates the sharing of risk by (re)insurers (sharing of the pool), not the scheme between an insurer and an insured. However, the latter can also reduce moral hazard as a similar form to the deductible insurance by assigning the portion of a loss to each party.

6 Conclusion

We propose a conceptual framework for decision-making in cyber risk management with a descriptive decision model under loss aversion. The framework is constructed under the assumption of interdependent risk in the interconnected network environment. We consider market insurance and risk control measures (self-protection and self-insurance) and investigate their interaction. It is vital for any decision-making problem under loss aversion to set a relevant, realistic reference point. In this regard, we postulate self-protection as the status quo of the reference point, as self-protection as a form of enhancing (or maintaining) a cyber security system is an essential management tool in the interconnected business world. We compare the reference point with different decisions under the assumption that self-protection continues to be implemented—specifically, the decisions are 1) market insurance and 2) self-insurance. We also set self-protection and self-insurance as the second reference point, which is interpreted as the maximum effort for controlling cyber risk (public good + private good), and compare it to the decision on market insurance only with the self-protection measure under the assumption that both insurance plans are substitutes for each other.

In the different scenarios, we come to the consistent conclusion that an agent with the reference point of self-protection as a basic management tool against cyber risk is more likely to avoid additional risk management measures (market insurance and self-insurance). In particular, a more loss averse agent is more likely to maintain the status quo (self-protection); the higher the net margin of insurance policies is for the agent, the less likely she is to give up additional measures. These findings address a fatalism observed in the current market that the agent tends to not choose any additional risk management tool, but stay only with self-protection.

We find conditions on the range of loss probability, the loading factor for market insurance, and the level of coverage for self-insurance, where an agent is willing to decide on market insurance and self-insurance. Based on the findings, it can be concluded that market insurance and self-insurance are substitutes of self-protection as the reference point under loss aversion, which is counter to Ehrlich and Becker's (1972) finding. We further estimate the frequency rate per year as a proxy of the loss probability from the insured perspective using an empirical dataset (PRC database). Although the anecdotal evidence of a fatalistic behavior against cyber risk is identified, we find that it is more likely to observe the increasing demand for additional risk management activities as agents are exposed to an increasing frequency rate year after year. A critical remark on our model framework is that it can not only be applied to the cyber risk context, but also generalized to any interdependent risk model under loss aversion, which particularly considers measures to control risk.

Although this study first opens up the discussion of a descriptive decision model on market insurance and risk control measures in the presence of interdependent risk, several limitations are still inherent in our model. In particular, several assumptions can be relaxed to adapt other possibilities in the model. For instance, a likelihood of successful contagion between nodes can be affected by other factors (e.g., heterogeneous security systems) so that the perfect contagion could not occur in the network system. Furthermore, we do not take into account the possibility of probability distortion due to the modeling of interdependent risk; however, it could be a possible avenue to combine both features in a model. Co-insurance or a reinsurance portfolio can be considered as an alternative insurance design possibly embedded in the model.

Appendix A. Review of literature and position of the paper

A.1. Literature on risk control measures and interdependent risk

In Table A1, we categorize the references dealing with risk control measures in a decision-making framework. The categorization is defined upon three criteria: interdependency in risks, propensity under risk, and relation to cyber risk. We consider 20 academic papers, including the present study, and find that the studies in the presence of interdependent risk tend to model the impact of self-protection, but not self-insurance. This might be due to the fact that self-protection affects the loss probability, which is important when modeling interdependent risk. In contrast, the studies examining the absence of interdependent risk take into account both self-protection and self-insurance to identify decision-making in risk control measures and market insurance. In addition, most cyber risk-related papers in our categorization analyze only self-protection in relation to cyber-insurance by accounting for cyber security enhancement; a noted exception is Johnson et al. (2011), who investigate the role of self-insurance. In this regard, our paper contributes to the literature by considering both self-protection and self-insurance in the decision-making framework against cyber risk under interdependent risk.

Table A1. The Classification of Risk Control and Loss Aversion under Interdependency in Risks

		Risk control measures			
		No interdependent risk		Interdependent risk	
Risk-aversion (EU)	Non-cyber	· EB72 (SP, SI) · DE85 (SP, SI) · S90 (SP, SI) · KS93 (SP, SI)	· S11 (SP, SI) · LRS12 (SP, SI) · AGT13 (SP, SI)	· KH03 (SP) · H07 (SP) · MK08 (SP)	
	Cyber-related	· KNL18 (SP)		· ÖMR05 (SP) · LB09 (SP) · SSFW10 (SP) · HR11 (SP) · JBG11 (SP, SI)	· ÖRM11 (SP) · ZXW13 (SP) · NL14 (SP) · HS18 (SP)
Loss-aversion (PT)	Non-cyber	-		-	
	Cyber-related	· V08 (SP)		**Present paper (SP, SI)**	

Note: SP and SI stand for self-protection and self-insurance, respectively. The references above are specified in the following; EB72: Ehrlich and Becker (1972); DE85: Dionne and Eeckhoudt (1985); S90: Shogren (1990); KS93: Konrad and Skaperdas (1993); KH03: Kunreuther and Heal (2003); ÖMR05: Öğüt, Menon and Raghunathan (2005); H07: Hofmann (2007); MK08: Mürmann and Kunreuther (2008); V08: Verendel (2008); LB09: Lelarge and Bolot (2009); SSFW10: Shetty et al. (2010); S11: Snow (2011); HR11: Hofmann and Ramaj (2011); JBG11: Johnson, Böhme and Grossklags (2011); ÖRM11: Öğüt, Raghunathan and Menon (2011); LRS12: Lohse, Robledo and Schmidt (2012); AGT13: Alary, Gollier and Treich (2013); ZXW13: Zhao, Xue and Whinston (2013); NL14: Naghizadeh and Liu (2014); HS18: Hota and Sundaram (2018); KNL18: Khalili, Naghizadeh and Liu (2018).

A.2. Comparison with the insurance literature under prospect theory

Although we can find a great number of studies on insurance demand under the classical decision theory, there has been very little literature on insurance demand under prospect theory. One possible explanation for the lack of study in this context is the difficulty in choosing a reference point and the vagueness of how to examine the decision-making under the descriptive model. However, we attempt to overcome these obstacles by focusing on a specific context (cyber risk management). Nevertheless, our framework can be also generalized to the context about risk control measures and interdependent risk under loss aversion. In this regard, we compare this paper with two relevant studies under loss aversion in Table A2. All studies in the table fall within the insurance context by commonly investigating the optimal decision on insurance.

Table A2. Comparison to Two Relevant References under Loss Aversion

	Schmidt (2016)	Eckles and Volkman-Wise (2019)	Present paper
Focus of study	Insurance demand under prospect theory	Insurance demand under prospect theory	Interaction between risk control measures and market insurance under prospect theory
Risk controls	No	No	Yes (self-protection and self-insurance)
Insurance design	Full insurance	Deductible and loaded premium	Deductible and loaded premium
Interdependent risk	No	No	Yes
Reference point	· Status quo with no insurance · Take up full insurance	· Take up insurance · Initial wealth · Status quo with deductible insurance	· Status quo with self-protection · Self-protection and self-insurance (full self-control)
Decision	· Optimal level of insurance coverage	· Optimal level of insurance coverage	· Optimal level of insurance coverage · Optimal level of self-insurance coverage
Main results	· An agent will demand either full coverage or no insurance, depending on the loss probability. · The indifference curve of the KT value function relies on the sign of the value function.	· Prospect theory explains 1) the preference for low deductibles for mandatory insurance, 2) the lack of demand for non-mandatory insurance and 3) the over-demand to insure small losses.	· An agent with the reference point of self-protection as a basic management tool against cyber risk is more likely to avoid additional risk management measures (market insurance and self-insurance), addressing a fatalism observed in the market.

Appendix B. Proofs

B.1. Optimal demand for market insurance under the expected utility theory

Following Eckles and Volkman-Wise (2019), we assume that a risk-averse agent has a comparable utility function with the KT value function as follows:

$$u(w) = w^{\alpha},$$

where $u(\cdot)$ is a utility function under the expected utility theory, w is the level of wealth for the agent and $\alpha \in [0,1]$ is the risk aversion parameter showing concavity, which is similar to the diminishing sensitivity parameter of prospect theory in the gain domain.

This function is consistent with the KT value function in the gain domain, where a power law function is used to represent risk aversion. In this section, we derive the optimal demand for market insurance under the expected utility theory when an agent implements self-protection.

In Section 3.3, we have two states of the world representing loss and no-loss cases. Referring to the final wealth of Case 1 in Table 2, we can write the expected utility for the agent as

$$E(U) = (1 - p_i) \cdot u(W - s_i - \pi) + p_i \cdot u(W - s_i - \pi - L + I).$$

The first order and second order conditions on the expected utility with respect to the insurance coverage, β, are respectively

$$\frac{\partial E(U)}{\partial \beta} = -p_i \cdot (1 - p_i) \cdot (1 + \delta) \cdot (L - d)^{+} \cdot u'(W - s_i - \pi) + p_i \cdot (1 - (1 + \delta) \cdot p_i) \cdot (L - d)^{+} \cdot u'(W - s_i - \pi - L + I),$$

$$\frac{\partial^2 E(U)}{\partial \beta^2} = (1 - p_i) \cdot p_i^2 \cdot (1 + \delta)^2 \cdot [(L - d)^{+}]^2 \cdot u''(W - s_i - \pi) + p_i \cdot (1 - (1 + \delta) \cdot p_i)^2 \cdot [(L - d)^{+}]^2 \cdot u''(W - s_i - \pi - L + I).$$

As shown in Eckles and Volkman-Wise (2019), the utility function $u(w) = w^{\alpha}$ is concave with the positive first derivative and the negative second derivative given $\alpha \in [0,1]$. Thus, the second order condition is negative, which leads the optimization problem to end up with a global maximum.

The optimization problem can be written as

$$\frac{\partial E(U)}{\partial \beta} = -p_i \cdot (1-p_i) \cdot (1+\delta) \cdot (L-d)^+ \cdot u'(W - s_i - \pi) + p_i \cdot (1-(1+\delta) \cdot p_i) \cdot (L-d)^+ \cdot u'(W - s_i - \pi - L + I) = 0.$$

By taking the utility function, it can be rewritten as

$$(1-p_i) \cdot (1+\delta) \cdot (W - s_i - \pi)^{\alpha-1} = (1-(1+\delta) \cdot p_i) \cdot (W - s_i - \pi - L + I)^{\alpha-1}.$$

This formula ends up with

$$\left(1 - \frac{\delta}{(1-p_i) \cdot (1+\delta)}\right) \cdot \left(\frac{W - s_i - \pi - L + I}{W - s_i - \pi}\right)^{\alpha-1} = 1.$$

This equation holds when full insurance is given—that is, $\delta = 0$ and $L = I$. Thus, it is concluded that the agent under the expected utility theory will purchase full insurance when the premium is fairly priced, which is consistent with Mossin (1968) and Eckles and Volkman-Wise (2019). If the loading factor is present ($\delta > 0$), then the last equation does not hold and full insurance is not optimal. Instead, partial insurance will be the optimal equilibrium.

B.2. Proof of Proposition 1

We derive inequality (18) with a simple mathematical process as follows; first, the KT value function is applied to the expected value in equation (16), and the expected value should be positive when market insurance is preferred.

$$V = -\lambda \cdot (1-p_i) \cdot \pi^\alpha + p_i \cdot (I - \pi)^\alpha > 0.$$

With $\pi = (1+\delta) \times p_i \times I$, we obtain

$$V = -\lambda \cdot (1-p_i) \cdot [(1+\delta) \cdot p_i \cdot I]^\alpha + p_i \cdot [(1-(1+\delta) \cdot p_i) \cdot I]^\alpha > 0.$$

By canceling out $I^\alpha (> 0)$ in both terms on the left side of the inequality and transferring the first term to the right hand side of the inequality, we can rewrite it as

$$p_i \cdot (1-(1+\delta) \cdot p_i)^\alpha > \lambda \cdot (1-p_i) \cdot [(1+\delta) \cdot p_i]^\alpha.$$

As $p_i \geq 0$ and $(1+\delta) \cdot p_i \geq 0$,

$$\left[\frac{1-(1+\delta) \cdot p_i}{(1+\delta) \cdot p_i}\right]^\alpha > \lambda \cdot \frac{1-p_i}{p_i}.$$

Appendix B.3 proves the condition $\delta \leq \frac{1-\alpha}{\alpha}$.

B.3. Proof of Proposition 2

The agent with the value function in equation (1) optimizes equation (16) with respect to the level of self-protection as follows:

$$\begin{aligned}\frac{\partial V}{\partial s_i} &= \frac{\partial[-\lambda\cdot(1-p_i)\cdot\pi^{\alpha}+p_i(I-\pi)^{\alpha}]}{\partial s_i}\\ &= -\lambda\cdot\alpha\cdot\pi^{\alpha-1}\cdot(1+\delta)\cdot I\cdot p_i'+\lambda\cdot(\alpha+1)\cdot\pi^{\alpha}\cdot p_i'+(I-\pi)^{\alpha}\cdot p_i'-\alpha\cdot\pi\cdot(I-\pi)^{\alpha-1}\cdot p_i'\\ &= \lambda\cdot\pi^{\alpha}\cdot p_i'\cdot\left[(\alpha+1)-\frac{\alpha}{p_i}\right]+(I-\pi)^{\alpha-1}\cdot p_i'\cdot(I-\pi-\alpha\pi)\end{aligned}$$

We know that the loss probability is a convex function with respect to the level of self-protection according to assumption 2. It leads the above equation to rely on $\left[(\alpha+1)-\frac{\alpha}{p_i}\right]$ and $(I-\pi-\alpha\pi)$ to determine the first order condition. If both cases are all positive, the marginal value for the agent decreases over the level of self-protection, leading to the following condition:

$$\frac{\alpha}{1+\alpha}<p_i<\frac{1}{(1+\alpha)(1+\delta)}.$$

Conversely, if both cases are all negative, the marginal value increases over the level of self-protection with the following condition:

$$\frac{1}{(1+\alpha)(1+\delta)}<p_i<\frac{\alpha}{1+\alpha}.$$

These conditions determine the loading factor by equating the upper and lower bounds such that

$$\delta=\frac{1-\alpha}{\alpha}.$$

Decreasing marginal value over the level of self-protection holds when the loading factor is smaller than $\frac{1-\alpha}{\alpha}$, whereas increasing marginal value holds with a higher loading factor.

B.4. Proof of Proposition 3

We clarify loss and gain domains by ensuring the positive value of the size of gain to understand the decision under prospect theory. The necessary condition to achieve this is $g_i\geq\frac{K(g_i)}{L}$, which is solved from $g_i\cdot L-K(g_i)\geq 0$ with regard to g_i.

Inequality (23) can be derived with a simple mathematical process as follows: As demonstrated in Appendix B.3, the KT value function is first applied to the expected value in equation (21), and the expected value should be positive when self-insurance is preferred

$$V = -\lambda \cdot (1 - p_i) \cdot K(g_i)^\alpha + p_i \cdot \left(g_i \cdot L - K(g_i)\right)^\alpha > 0.$$

This can be rewritten as

$$p_i \cdot \left(g_i \cdot L - K(g_i)\right)^\alpha > \lambda \cdot (1 - p_i) \cdot K(g_i)^\alpha.$$

As $p_i \geq 0$ and $K(g_i) \geq 0$,

$$\left(\frac{g_i \cdot L - K(g_i)}{K(g_i)}\right)^\alpha > \lambda \cdot \frac{1 - p_i}{p_i}.$$

Solving it with respect to p_i yields

$$p_i > \frac{1}{\left(\frac{g_i \cdot L}{K(g_i)} - 1\right)^\alpha / \lambda + 1}.$$

B.5. Proof of Proposition 4

To prove Proposition 4, we consider two cases in the following:

i) $\boldsymbol{K(g_i) - \pi \leq 0}$ ***and*** $\boldsymbol{I - g_i \cdot L + K(g_i) - \pi > 0}$ *(positive net effect of the transition from self-insurance to market insurance)*

The expected value with the KT value function in this case can be written as

$$V = -\lambda \cdot (1 - p_i) \cdot [-(K(g_i) - \pi)]^\alpha + p_i \cdot [I - g_i \cdot L + K(g_i) - \pi]^\alpha.$$

To determine the preference for market insurance instead of self-insurance, we postulate a positive expected value; we can then rewrite the preceding equation as

$$p_i \cdot [I - g_i \cdot L + K(g_i) - \pi]^\alpha > \lambda \cdot (1 - p_i) \cdot [\pi - K(g_i)]^\alpha.$$

As $p_i \geq 0$ and $K(g_i) - \pi \leq 0$,

$$\left[\frac{I - g_i \cdot L + K(g_i) - \pi}{\pi - K(g_i)}\right]^\alpha > \lambda \cdot \frac{1 - p_i}{p_i}.$$

This leads to

$$\left[\frac{I - g_i \cdot L}{\pi - K(g_i)} - 1\right]^\alpha > \lambda \cdot \frac{1 - p_i}{p_i}.$$

ii) $\boldsymbol{K(g_i) - \pi > 0}$ ***and*** $\boldsymbol{I - g_i \cdot L + K(g_i) - \pi \leq 0}$ *(positive net effect of staying with self-insurance)*

Now the signs of two outcomes are switched. The expected value with the KT value function in this case is

$$V = -\lambda \cdot (1 - p_i) \cdot [-(I - g_i \cdot L + K(g_i) - \pi)]^\alpha + p_i \cdot [K(g_i) - \pi]^\alpha.$$

To determine the preference for market insurance instead of self-insurance, we need to have a positive expected value, leading to the following equation:

$$p_i \cdot [K(g_i) - \pi]^\alpha > \lambda \cdot (1 - p_i) \cdot [g_i \cdot L - I - (K(g_i) - \pi)]^\alpha.$$

As $p_i \geq 0$ and $I - g_i \cdot L + K(g_i) - \pi \leq 0$,

$$\left[\frac{K(g_i) - \pi}{g_i \cdot L - I - (K(g_i) - \pi)}\right]^\alpha > \lambda \cdot \frac{1 - p_i}{p_i}.$$

This leads to

$$\left[\frac{g_i \cdot L - I}{K(g_i) - \pi} - 1\right]^\alpha < \frac{1}{\lambda} \cdot \frac{p_i}{1 - p_i}.$$

Appendix C. Additional evidence on frequency rate

To support our empirical estimation on frequency rate of cyber risk in Section 4.4, additional evidence is listed in Table C1. The figures in the evidence are estimated by a variety of organizations that regularly report on the status quo of cyber security worldwide. It is distinct from the anecdotal/statistical evidence that cyber loss events and attacks are more likely to occur in our daily lives and business, particularly in recent years, thereby possibly leading to a significant increase in the subjective probability of potential insureds against cyber risk.

Table C1. List of Anecdotal/Statistical Evidence

Evidence	**Source**	**Year**
The average global likelihood of a breach above a minimum of 10,000 records in the next 24 months is 27.9%. Brazil and South Africa have the highest probabilities, with 43.0% and 40.9%, respectively.	Ponemon study	2018
The number of cyber incidents has doubled to 159,700 globally in 2017 compared to 2016.	Online Trust Alliance	2018
Approximately 70% of organizations believe that their cyber security risk significantly increased in 2017.	Ponemon study	2017
IoT attacks increased by 600% from 2016 to 2017.	Symantec	2017
The average annual number of breaches increased 27.4% each year from 2015 to 2017.	Accenture	2017
Cyberattacks are the biggest threat to mankind—an even bigger threat than nuclear weapons.	Warren Buffett (in the 2017 Berkshire Hathaway's annual shareholder meeting)	2017
Financial losses related to cybercrime in Hong Kong have risen 680% over the five years (2012–2016).	Hong Kong Police Force	2017
A 300% increase in ransomware attacks in the U.S. occurred in 2016 compared to 2015, as reported by the FBI.	FBI	2016
China was the country with the largest number of infected computers (57.24%) in 2015, 30% more than in 2014, followed by Taiwan (49.15%) and Turkey (42.52%).	Panda Security	2015

Note: The evidence is listed in chronological order, from the most recent publication, with a variety of sources, including industrial reports to a government agency.

References

Alary, D., Gollier, C., and Treich, N. (2013). The effect of ambiguity aversion on insurance and self-protection. *The Economic Journal, 123*(573), 1188-1202.

Andalib, M. S., Tavakolan, M., and Gatmiri, B. (2018). Modeling managerial behavior in real options valuation for project-based environments. *International Journal of Project Management, 36*(4), 600-611.

Armerding, T. (2018, December 20). *The 18 biggest data breaches of the 21st century*. CSO. Retrieved from https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html.

Baillon, A., Bleichrodt, H., and Spinu, V. (2018). Searching for the reference point. *Working paper.* Erasmus University Rotterdam.

Berr, J. (2017 May 16). "WannaCry" ransomware attack losses could reach $4 billion. *CBS News*. Retrieved from https://www.cbsnews.com/news/wannacry-ransomware-attacks-wannacry-virus-losses/.

Binswanger, J. (2007). Risk management of pensions from the perspective of loss aversion. *Journal of Public Economics, 91*(3-4), 641-667.

Böhme, R., and Schwartz, G. (2010). Modeling cyber-insurance: Towards a unifying framework. *Workshop on the Economics and Insurance Security (WEIS)*.

Center for Strategic & International Studies (CSIS). (2018). *Economic impact of cybercrime—No slowing down.* Washington, DC: McAfee.

Cummins, J., Lin, Y., and Phillips, R. (2006). Capital allocation and the pricing of financially intermediated risks: An empirical investigation. *Working paper*.

de Smidt, G., and Botzen, W. (2018). Perceptions of corporate cyber risks and insurance decision-making. *The Geneva Papers on Risk and Insurance-Issues and Practice, 43*(2), 239-274.

Department for Digital, Culture, Media & Sport (DCMS). (2018). *Cyber security breaches survey 2018.* London: U.K. Government.

Dionne, G., and Eeckhoudt, L. (1985). Self-insurance, self-protection and increased risk aversion. *Economics Letters, 17*(1-2), 39-42.

Eckles, D., and Volkman-Wise, J. (2019). Loss aversion, probability weighting, and the demand for insurance. *Working paper*.

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

Ehrlich, I., and Becker, G. (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy, 80*(4), 623-648.

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

Eling, M., and Schnell, W. (2018). Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the US risk-based capital standards, and the Swiss solvency test. *Working paper*.

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

Ernst & Young (EY). (2015). *Cybersecurity and the Internet of Things.* London: Ernst & Young.

European Insurance and Occupational Pensions Authority (EIOPA). (2018). *Understanding cyber insurance—A structured dialogue with insurance companies.* Frankfurt: EIOPA.

European Union. (2016). *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC* (General Data Protection Regulation). Regulation (EU) 2016/679 of the European Parliament and of the Council.

Gordon, L., and Loeb, M. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC), 5*(4), 438-457.

Grenoble, R. (2017, June 19). Registered to vote? If so, a GOP firm probably exposed your personal data. *HuffPost*. Retrieved from https://www.huffpost.com/entry/deep-root-analytics-gop-data-breach-voters_n_59402d52e4b09ad4fbe396c5.

Grossklags, J., Christin, N., and Chuang, J. (2008). Secure or insure?: a game-theoretic analysis of information security games. *Proceedings of the 17th international conference on World Wide Web* (pp. 209-218). ACM.

Hofmann, A. (2007). Internalizing externalities of loss prevention through insurance monopoly: an analysis of interdependent risks. *The Geneva Risk and Insurance Review, 32*(1), 91-111.

Hofmann, A., and Peter, R. (2016). Self-insurance, self-protection, and saving: On consumption smoothing and risk management. *Journal of Risk and Insurance, 83*(3), 719-734.

Hofmann, A., and Ramaj, H. (2011). Interdependent risk networks: the threat of cyber attack. *International Journal of Management and Decision Making, 11*(5-6), 312-323.

Hofmann, A., Wheatley, S., and Sornette, D. (2019). Heavy-tailed data breaches in the nat-cat framework & the challenge of insuring cyber risks. *Working paper*.

Hota, A., and Sundaram, S. (2018). Interdependent security games on networks under behavioral probability weighting. *IEEE Transactions on Control of Network Systems, 5*(1), 262-273.

Insurance Europe. (2015). *European motor insurance markets.* Brussels.

Jalali, M., Siegel, M., and Madnick, S. (2019). Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. *Journal of Strategic Information Systems, 28*(1), 66-82.

Johnson, B., Böhme, R., and Grossklags, J. (2011). Security games with market insurance. In J. Baras, J. Katz, & E. Altman, *Decision and game theory for security* (pp. 117-130). College Park, MD: Springer-Verlag.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47*(2), 263-292.

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., and Stulz, R. (2018). *What is the impact of successful cyberattacks on target firms?* Cambridge, MA: National Bureau of Economic Research.

Khalili, M. M., Naghizadeh, P., and Liu, M. (2018). Designing cyber insurance policies: The role of pre-screening and security interdependence. *IEEE Transactions on Information Forensics and Security, 13*(9), 2226-2239.

Konrad, K., and Skaperdas, S. (1993). Self-insurance and self-protection: A nonexpected utility analysis. *The Geneva Papers on Risk and Insurance-Theory, 18*(2), 131-146.

Kunreuther, H., and Heal, G. (2003). Interdependent security. *Journal of Risk and Uncertainty, 26*(2-3), 231-249.

Kunreuther, H., and Michel-Kerjan, E. (2014). Economics of natural catastrophe risk insurance. In *Handbook of the Economics of Risk and Uncertainty* (Vol. 1, pp. 651-699). Amsterdam: North-Holland.

Kunreuther, H., and Pauly, M. (2018). Dynamic insurance decision-making for rare events: The role of emotions. *Geneva Papers on Risk and Insurance—Issues and Practice, 43*(2), 335-355.

Lelarge, M., and Bolot, J. (2009). Economic incentives to increase security in the internet: The case for insurance. *INFOCOM 2009, IEEE* (pp. 1494-1502).

Lloyd’s. (2017). *Counting the cost: Cyber exposure decoded.* London: Lloyd’s.

Lohse, T., Robledo, J., and Schmidt, U. (2012). Self-insurance and self-protection as public goods. *Journal of Risk and Insurance, 79*(1), 57-76.

Mersinas, K., Hartig, B., Martin, K., and Seltzer, A. (2016). Are information security professionals expected value maximizers?: An experiment and surveybased test. *Journal of Cybersecurity, 2*(1), 57-70.

Mossin, J. (1968). Aspects of rational insurance pricing. *Journal of Political Economy, 76*(4-1), 553-568.

Mürmann, A., and Kunreuther, H. (2008). Self-protection and insurance with interdependencies. *Journal of Risk and Uncertainty, 36*(2), 103-123.

Naghizadeh, P., and Liu, M. (2014). Voluntary participation in cyber-insurance markets. *Workshop on the Economics of Information Security (WEIS)*.

Nagurney, A., and Shukla, S. (2017). Multifirm models of cybersecurity investment competition vs. cooperation and network vulnerability. *European Journal of Operational Research, 260*(2), 588-600.

National Conference of State Legislative (NCSL). (2019). Security breach notifications laws. Retrieved from http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx.

Öğüt, H., Menon, N., and Raghunathan, S. (2005). Cyber insurance and IT security investment: Impact of interdependence risk. *Workshop on the Economics of Information Security (WEIS)*.

Öğüt, H., Raghunathan, S., and Menon, N. (2011). Cyber security risk management: Public policy implications of correlated risk, imperfect ability to prove loss and observability of self protection. *Risk Analysis, 31*(3), 497-512.

Outreville, F. (1998). *Theory and practice of insurance.* New York: Springer Science & Business Media.

Perlroth, N. (2016, October 21). Hackers used new weapons to disrupt major websites across U.S. *The New York Times*. Retrieved from https://www.nytimes.com/2016/10/22/business/internet-problems-attack.html.

Ponemon Institute. (2018). *2018 cost of a data breach study.* Traverse City, MI: Ponemon Institute.

PricewaterhouseCoopers (PwC). (2017). *Strengthening digital society against cyber shocks: Key findings from the Global State of Information Security Survey 2018.* PricewaterhouseCoopers.

Privacy Rights Clearinghouse (PRC). (2019). *Data breaches*. Retrieved from https://www.privacyrights.org/data-breaches.

Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. (2017). Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?. *Working paper.*

Schmidt, U. (2016). Insurance demand under prospect theory: a graphical analysis. *Journal of Risk and Insurance, 83*(1), 77-89.

Schanz, K. (2019). *Underinsurance in mature economies: Reasons and remedies.* Zurich: The Geneva Association.

Shetty, N., Schwartz, G., Felegyhazi, M., and Walrand, J. (2010). Competitive cyber-insurance and internet security. In *Economics of information security and privacy* (pp. 229-247). New York, NY: Springer.

Shogren, J. (1990). The impact of self-protection and self-insurance on individual response to risk. *Journal of Risk and Uncertainty, 3*(2), 191-204.

Snow, A. (2011). Ambiguity aversion and the propensities for self-insurance and self-protection. *Journal of Risk and Uncertainty, 42*(1), 27-43.

Swiss Re. (2016). *Cyber: in search of resilience in an interconnected world.* Zurich: Swiss Re.

Tversky, A., and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*(4), 1039-1061.

Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

Verendel, V. (2008). *A prospect theory approach to security.* Göteborg: Chalmers University of Technology.

von Neumann, J., and Morgenstern, O. (1953). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

von Ungern-Sternberg, T. (1996). The limits of competition: Housing insurance in Switzerland. *European Economic Review, 40*(3-5), 1111-1121.

Wang, S. (2017). Integrated framework for information security investment and cyber insurance. *Working paper*. Insurance Risk and Finance Research Centre (IRFRC), Nanyang Technological University.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806-820.

Wheatley, S., Maillart, T., and Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B, 89*, 7.

Zhao, X., Xue, L., and Whinston, A. (2013). Managing interdependent information security risks: Cyberinsurance, managed security services, and risk pooling arrangements. *Journal of Management Information Systems, 30*(1), 123-152.

Zweifel, P., and Eisen, R. (2012). *Insurance economics.* Berlin Heidelberg: Springer-Verlag.

Essay IV

Risk aggregation in non-life insurance: Standard models vs. internal models

Abstract

Standard models for capital requirements restrict the correlation between risk factors to the linear measure and disregard undertaking-specific parameters. We propose a comprehensive framework for risk aggregation in non-life insurance using vine copulas and two levels of aggregation: base level (within asset and underwriting modules) and top level (between asset and underwriting modules). Using empirical data from Korean and German insurance companies, we compare our internal risk model with three regulatory standard models (Korean risk-based capital, Solvency II, Swiss Solvency Test) and show that the standard models lead to more than 50% higher capital requirements on average. Half of the overestimation results from the uniform parameter selection imposed by regulations and the other half comes from the linear correlation assumption. The differences between standard models and internal models might distort competition when both approaches are used in a single market.

Keywords: Insurance regulation, Risk aggregation, Vine copula, Capital requirements

---

Martin Eling, Kwangmin Jung (2018).

This paper was presented at the 2018 annual meeting of the American Risk and Insurance Association (ARIA) in Chicago, the 2018 annual meeting of the Asia Pacific Risk and Insurance Association (APRIA) in Singapore, the 2018 conference "Mathematical and Statistical Methods for Actuarial Science and Finance (MAF)" in Madrid, and the Ph.D. in Finance Seminar at the University of St. Gallen. This paper has been resubmitted to *ASTIN Bulletin* after the invitation to minor revisions with conditional acceptance.

1 Introduction

In the aftermath of the 2008 financial crisis, meeting the capital requirement became one of the central management tasks in the banking and insurance industries. Solvency II (SII) as the regulatory standard for insurance companies in the European Union covers potential losses from the asset and liability side of an insurer by calculating solvency capital requirements (SCR) based on either a standard formula or an accredited internal risk model (EC, 2014). Other regulatory standards such as the U.S. risk-based capital (U.S. RBC), the Swiss Solvency Test (SST) or the Korean RBC (K-RBC) of interest in this study have recently been developed or revised (FINMA, 2006; FSS, 2017). Additionally, the International Association of Insurance Supervisors (IAIS) is currently discussing a new homogenized risk-based capital standard (Insurance Capital Standard; ICS) for globally active insurance groups (IAIS, 2018).[118]

The regulatory frameworks aggregate risks from the assets and liabilities to protect the insurer against simultaneous losses in a stressed situation. The distributions of such risks vary, especially between risk factors in the asset and underwriting portfolios, including significant tail risks. However, despite significant differences in risk distributions, regulatory standard models (e.g., SII and the K-RBC) require insurers to aggregate them under the linear dependence assumption with predefined correlation parameters and do not allow undertakings to replace the correlation parameters by undertaking-specific parameters (EC, 2014, p. 151; FSS, 2017, p. 199; ICS, 2018, p. 121).[119] Obviously, it is computationally convenient to aggregate risks under the linear dependence assumption. However, a clear drawback of this approach is that capital requirements might be misestimated by uniformly aggregating risks, which can display different distributional features and shapes (Tang and Valdez, 2009). The linear correlation assumption could not capture any symmetric or asymmetric tail risks, despite significant evidence of tail risks between assets (see, e.g., Aas and Berg, 2009) and liabilities (see, e.g., Diers, Eling and Marek, 2012).

[118] After conducting several field tests with 50 insurance groups, IAIS will monitor the adoption of the new ICS standard for five years and then plans to launch the final approach in 2024 (IAIS, 2018, p. 11). ICS is closer to Solvency II framework than RBC or SST in that it is a stress approach with similar categorization in the market risk and the insurance risk modules and requires a 99.5% Value-at-Risk estimation for the calculation (IAIS, 2018, p. 63). However, it also combines a factor-based approach, as done in RBC, particularly for non-life risks (premium and claims reserve risks), which is simpler to implement than the stress approach (IAIS, 2018, p. 60). Thus, this new standard attempts to combine existing regulatory frameworks.

[119] Devineau and Loisel (2009) also point out that risk aggregation in the standard formula of Solvency II is linear with margins in the elliptical family, which can impose significant restriction to undertakings. The authors attempt to build a non-linear internal model with the nested simulation technique to overcome this issue. Their focus is on market risk (stock and interest rate risk) and simulation-based analysis.

In addition to the linear correlation assumption, the regulatory standard models do not use undertaking-specific parameters on risk factors. Instead, they predefine either parameters (SII, K-RBC, ICS) or a certain distributional assumption on asset and underwriting risks (SST) for all undertakings. However, the parameters predefined by the regulations do not fully reflect the empirical data for an individual undertaking and distributional properties vary between assets and underwriting risks in different datasets. It might be inappropriate to apply homogeneous parameters into datasets with heterogeneous properties and to aggregate different risks with the same statistical tools. All these limitations motivate us to study alternatives to the regulatory standard models.

To overcome such limitations, insurers construct and use internal risk models (FINMA, 2006, p. 4; EC, 2014, Chapter VI; ICS, 2018, Chapter 9.2).[120] The approval of an internal model by the regulator for calculating capital requirements signals to investors and analysts that the company can optimize its economic capital. However, establishing an internal risk model requires resources that small and mid-sized insurance companies cannot afford. This might lead to significant differences in solvency models used in one market so that competition might be distorted (Eling, Schmeiser and Schmit, 2007). While both standard and internal models should ensure effectiveness (i.e., meeting the regulatory capital requirements), it might be that the standard model significantly overestimates the necessary capital. If the benefit (possible reduction of necessary capital) exceeds the costs of developing and using an internal model, it would be more efficient to use the internal model. A distortion in competition might then arise if only large companies have the option to use a more efficient internal model. We empirically show that this is the case. Developing appropriate internal risk models and quantifying potential deviations from regulatory standard models are thus not only of interest for academics in actuarial science, but also highly relevant for insurance managers, regulators and public policy.

To understand potential deviations between internal models and standard models, we construct internal risk models that aggregate different risk types of a non-life insurer using realistic asset and underwriting portfolios. To our knowledge, the literature has presented no comprehensive framework to aggregate asset and underwriting risk under the consideration of the correlation

[120] The regulators thus do not force undertakings to adapt the standard model. The European Commission also requires insurers to conduct an assessment process called the own-risk and solvency assessment (ORSA; EC, 2014, p. 187) critically reflecting the methods and main assumptions of the standard model. The internal model has not been institutionalized in the Korean regulation system, but is planned to be adopted (FSS, 2017, p. 6). The modeling procedure and risk aggregation in this study illustrates how important it is to adequately reflect the risk factors and their dependency. Our results also supports the conclusion of Bølviken and Guillen (2017) that the correlation assumption in the current Solvency II framework should be replaced by copulas.

assumption and predefined parameters. Studies in the insurance field have typically limited the risk aggregation to the liability side (e.g., Tang and Valdez, 2009; Diers et al., 2012).[121] We overcome the limitations of previous studies by using the recently developed vine copula[122] and suggest a comprehensive method of measuring the solvency of a non-life insurer on both asset and liability by building a two-step aggregation model: base level and top level. The former models the dependence structures of different assets and underwriting risks from various lines of business (marginal modeling with undertaking-specific parameters). The latter couples the estimated aggregate distributions of assets and liabilities.

This paper contributes to the academic literature and insurance practice in two ways. First, we develop an integrated framework for the economic capital calculation that combines asset and insurance portfolios with undertaking-specific parameters. Here, we focus on the market risk and underwriting risk module, which are the two main risk drivers for a non-life insurance company, as evidenced, for example, in the Solvency II field tests. Then, we consider all possible multivariate dependence models in the copula field. This allows us to identify the internal risk model that best fits the data using an up-to-date copula methodology. The intention is not to develop new theory or methods, but to comprehensively discuss and apply some recently developed methods to an important field of actuarial science. Second, we empirically compare our internal risk model with regulatory standard models and document significant differences (61.2% on average for Korea and 57.8% on average for Germany) between the two

[121] In Appendix A, the contribution of this study is highlighted by comparing methodological aspect and outcome of this paper with those of five related references in the non-life insurance context: Pfeifer and Strassburger, 2008; Eling and Toplek, 2009; Tang and Valdez, 2009; Savelli and Clemente, 2011; Diers, Eling and Marek, 2012. Diers et al. (2012) propose a risk aggregation method for multi-line non-life insurers using Bernstein copulas. Savelli and Clemente (2011) analyze the premium risk of a multi-line non-life insurer by estimating the capital requirement with hierarchical Archimedean copulas. Tang and Valdez (2009) implement risk aggregation for general insurance and show that copulas capturing tail risks generate higher diversification benefits and reduce capital requirements. Eling and Toplek (2009) integrate asset and underwriting portfolios for the bivariate case using hierarchical Archimedean copulas and apply the model in a dynamic financial analysis for non-life insurance companies (using simulation examples, but no real data).

[122] The choice of the vine copula is based on several seminal studies in the context of the high-dimensional dependence modeling that stress the validity and efficiency of using the vine copula for financial and insurance data, possibly improving regulatory frameworks. For example, Dissmann et al. (2013) apply the regular vine copula structure into 16 financial assets to accurately capture their tail risk. Brechmann, Czado and Paterlini (2014) make use of the regular vine copula to model the dependence structure of operational risk losses, which represent to some extent similar distributional properties to insurance claims. They estimate regulatory capital for operational risks with their modeling process and find 38% smaller estimates than what the banking regulatory framework requires. Côté and Genest (2015) apply a tree-based copula structure similar to vine copula structure (or hierarchical Archimedean copula structure) on the insurance portfolio of a large Canadian insurance company and argue that regulatory frameworks need a flexible, multivariate risk model using high-dimensional copulas. All the studies clearly support the use of the vine copula (or high-dimensional copula-based approach) for the assessment of regulatory capitals to fully describe market and underwriting risks.

approaches.[123] Revealing these massive differences has important policy implications for insurance managers and regulators, since they might create an uneven playing field that places small- and mid-sized insurers at a disadvantage. We illustrate this by applying our models to companies of different sizes.

The rest of the paper is structured as follows. In Section 2, we present the regulatory frameworks (K-RBC, SII and SST), followed by the methodological framework. We describe the data used in the empirical study in Section 3. Section 4 presents the marginal modeling and dependence modeling at the base and top levels. Applications to the economic capital calculation are given in Section 5. Finally, we conclude and discuss further research questions in Section 6.

2 Regulation, theoretical background and model framework

2.1 Regulatory frameworks

Korean risk-based capital (K-RBC)

The RBC system was launched in the early 1990s by the National Association of Insurance Commissioners (NAIC) in the U.S. Modified versions have been applied in many other countries (e.g., Australia, Japan, Korea, Singapore and the U.K.). The K-RBC system by the Korean financial supervisory service (FSS), which is of interest in our empirical study, came into effect in April 2011 and was revised in 2017 (FSS, 2017). It requires insurers to estimate the RBC ratio and its components as follows (FSS, 2017, p. 37):

$$RBC\ ratio = \frac{Available\ capital}{Required\ capital} \times 100 \geq 150, \tag{1}$$

$$Available\ capital = C - I + S, \tag{2}$$

$$Required\ capital = \sqrt{\sum_{i=1} \sum_{j=1} Corr_{i,j} \times Risk_i \times Risk_j}, \tag{3}$$

[123] The data we consider is from Korea and Germany, two comparable markets from different parts of the world that have already implemented "risk-based capital" regulations. According to Swiss Re (2018), the Korean insurance market is ranked seventh in the world and third in Asia with regard to the total premium volume ($181.2 bn); the German insurance market is sixth in the world and third ($223.0 bn) in Europe. We also consider two different markets, since most studies concentrate on data from one market or on simulated data. (see Appendix A).

where C is the summation of core capital and supplementary capital, I are intangible assets, S are any capital deficiency of subsidiaries, $Corr_{i,j}$ is a correlation coefficient between risk i and risk j, and $Risk_i$ and $Risk_j$ are the exposures[124] of the i-th and j-th risk.

The K-RBC is a factor-based system requiring insurers to use a risk coefficient for each risk factor given in the standard model.[125] The risks in equation (3) are estimated from individual modules consisting of insurance, market, credit, interest rate and operational risk. The K-RBC system classifies insurance risk into four categories based on the coverage (life, long-term non-life, automobile and general insurance); the market risk module consists of three asset classes (short-term trading securities, derivatives and foreign exchange).[126] The evaluation of insurance risk is determined by premium risk and reserve risk, which are estimated by the sum of multiplication of risk exposure and adjusted risk coefficient under a predefined correlation.

The key difference between the initial system in 2011 and the revised 2017 system is that an insurer needs to consider the correlation and the diversification effect between sub-risks in the insurance risk module and the confidence level to estimate the economic capital is increased from 95% to 99% (FSS, 2017, p. 6, p. 43). However, the new framework still does not require correlations between sub-risks in the market risk module, which is a difference between the K-RBC and SII. Table 1 gives the assumed correlation matrices between the risk modules and between the risk classes of insurance risk.

Table 1. Correlation Matrices under the K-RBC (FSS, 2017, p. 36)

Panel A: Correlation of risk modules

	Insur	Market	Credit	Interest
Insur	1			
Market	0.25	1		
Credit	0.25	0.5	1	
Interest	0.25	0.5	0.5	1

Panel B: Correlation of insurance risk

	Life	LT-NL	Auto	General
Life	1			
LT-NL	0.25	1		
Auto	0.25	0.25	1	
General	0.25	0.5	0.5	1

Note: LT-NL=Long-term non-life risk.

[124] The term “exposure” is used here to represent the size of capital for a single risk in the regulatory framework. Here we follow the official document for the K-RBC (FSS, 2017, p. 29) which also employs this terminology.

[125] The standard risk coefficient for a risk factor i (=γ_i) can be calculated as follows (FSS, 2017, p. 29):

$$\gamma_i = \frac{VaR_\alpha}{Expected\ loss} = \frac{\inf\{X_i \in \mathbb{R}: X_i \geq F_{X_i}^{-1}(1-\alpha)\}}{E[X_i]},$$

where α is a confidence level for Value-at-Risk of the risk factor i, X_i is a loss vector and F_{X_i} is a probability function of X_i. Here, VaR_α is a possible loss amount (positive value) at the determined confidence level α. Insurers have an opportunity to adjust the standard risk coefficient using 50% of the difference in the combined ratio between an insurer and the industry average (FSS, 2017, p. 48; explained in detail in Appendix E).

[126] General insurance contains fire, package, marine, accident, liability and others, whereas long-term non-life insurance includes death/disability, disease, property, medical expense and others (see FSS, 2017, p. 47 for long-term non-life insurance and p. 55 for general insurance). Short-term trading securities consist of equities and bonds (treasury and corporate) classified as the short-term trading assets from the corporate balance sheet in Korean currency and the foreign exchange class is the risk caused by the change in the exchange rate from the country in which the asset is invested (see FSS, 2017, p. 130).

Solvency II (SII)

SII, the new European Union regulatory framework that took effect in January 2016, is a scenario-based system that considers risk parameters in predefined stress situations, which are calibrated with data from a variety of insurers from different European countries. SII is comprised of market, default, life, non-life, health and operational risk modules (EC, 2014). Market risk incorporates six risk factors (interest rate, equity, property, spread, currency and concentration) providing calibrated parameters for potential stressed scenarios. The standard capital requirement for a sub-module of market risk is calculated by the change in the net asset value (asset–liability) from a normal state to a stressed state (see EC, 2014, Section 5 in Chapter V for detail). The non-life underwriting risk module consists of premium risk, reserve risk and catastrophe risk factors for different lines of business with predefined correlation assumptions (EC, 2014, Annex IV). The capital requirement by SII, which is called solvency capital requirement (SCR), is calculated using the square root formula (equation (3)) with replacing $Risk_i$ and $Risk_j$ by SCR_i and SCR_j respectively.

The SCR for each module is evaluated at 99.5% confidence level based on parameters calibrated under certain stress situations, where values of assets and liabilities are affected by a particular risk. SII specifies the correlation measures within and between risk modules. Table 2 shows the correlation matrices between risk modules and within the non-life and market risk module. SII differentiates the correlation assumptions of the market risk module into two possible scenarios: "up" and "down" shocks of interest rate (see Appendix E for more detail).

Table 2. Correlation Matrices under SII (EC, 2010, p. 96; EC, 2014, p. 105, p. 233)

Panel A: Correlation of risk modules

	Mkt	Def	Life	Hth	NL
Mkt	1				
Def	0.25	1			
Life	0.25	0.25	1		
Hth	0.25	0.25	0.25	1	
NL	0.25	0.5	0	0	1

Panel B: Correlation of non-life risk

	Mot	Fire	MAT	Liab	Mis
Mot	1				
Fire	0.25	1			
MAT	0.5	0.25	1		
Liab	0.5	0.25	0.25	1	
Mis	0.5	0.5	0.5	0.5	1

Panel C: Correlation of market risk

Correlation for the "up" shock

	Interest	Equity	Property	Spread
Interest	1			
Equity	0.5	1		
Property	0.5	0.75	1	
Spread	0.5	0.75	0.5	1

Correlation for the "down" shock

	Interest	Equity	Property	Spread
Interest	1			
Equity	0	1		
Property	0	0.75	1	
Spread	0	0.75	0.5	1

Note: Mkt=Market risk; Def=Default risk; Hth=Health risk; NL=Non-life risk; Mot=Motor; MAT=Marine, aviation and transport; Liab=3rd party liability; Mis=Miscellaneous. In panel B and C, the original matrices consist of more factors, but we only show the relevant factors to our empirical data.

Swiss Solvency Test (SST)

The SST, introduced in 2006, evaluates the amount of risks on a market-consistent basis. It specifies 74 risk factors that potentially influence the asset side of an insurer and 13 risk factors in the non-life underwriting module (FINMA, 2006). SST proposes a model-based approach to derive capital requirements for different risk modules; a multivariate normal distribution is assumed for the market risk module (FOPI, 2004, p. 19; see Appendix E for more detail). The claims distributions are split into normal, large and catastrophic claims by using mean and standard deviation of the lognormal distribution, a compound Poisson process and a scenario-based model, respectively (FOPI, 2004, p. 22).[127] SST requires insurers to estimate their potential risk levels using Tail Value-at-Risk at 99% confidence level and uses the convolution to aggregate normal and large claims and the asset portfolio and underwriting portfolio (FOPI, 2004, p. 24).[128]

2.2 Theoretical background and model framework

Non-life insurers typically operate several lines of business and form a risk pool to realize diversification effects. Most of the earned premiums are invested in capital markets and diversified into risky and safe investments. We thus need to aggregate different risk factors from the asset and liability side of the balance sheet when estimating the insurers' total risk and capital requirements. A risk factor is here a single asset or a single line of insurance business. The starting point of the risk aggregation process is to generate the aggregate distribution of individual risk factors from the asset and liability side. At this base level of aggregation, each marginal distribution needs to be parametrically specified to reflect the distributional feature of each risk factor. Then, we aggregate the estimated margins considering the dependence structure between risk factors. Copula methods are widely used for this aggregation by transforming heterogeneous margins to uniform distributions.[129] In the following, we assume that the reader is familiar with the concept of copula and multivariate modeling with copula, which originates from Sklar's theorem (see Nelsen (2006), Chapter 2.3., for an overview).

Several copula methods have been developed to estimate the dependence structure of risk factors. Gaussian and student-t copulas belonging to elliptical family are popular functions used

[127] Large claims are defined as those above either 1 million or 5 million CHF; catastrophic claims are those that significantly influence several lines of business at the same time or are not covered by large claims distribution (FOPI, 2004, p. 44).

[128] FINMA (2006) conducted a field test using monthly asset data from different firms and calibrates a variance-covariance matrix in its standard model.

[129] Traditionally, the risk aggregation is carried out using a variance-covariance matrix under normality assumption on the margins leading to the joint normality, especially for the asset portfolio management; the insurance risks are traditionally aggregated under independence (Embrechts, McNeil and Straumann, 2002).

in practice due to their statistical convenience. Clayton and Gumbel copulas belong to the Archimedean family and are widely used in modeling dependency, projecting asymmetric tail dependency (Nelsen, 2006). Recently, high-dimensional dependence models have been developed and applied to different fields of study: Hierarchical Archimedean Copula structure (HAC; also called nested Archimedean copula) and Pair Copula construction (PCC; also called vine copula model).[130]

HAC is a multivariate dependence structure constructed by generators of Archimedean family in a multi-level hierarchical setting, which can help to overcome exchangeability of simple Archimedean copula approach. Various papers have employed HAC to reduce the dimension and find a more accurate high-dimensional dependence structure than typical copula functions such as elliptical copulas and simple Archimedean copulas (e.g., Joe, 1997; Embrechts et al., 2001; Nelsen, 2006; Savu and Trede, 2010). Similarly, PCC is used to reduce the dimension by pairing the variable set and more accurately establish a flexible dependence structure (Bedford and Cooke, 2001). PCC was introduced by Joe (1996) and has been developed by Bedford and Cooke (2002), Kurowicka and Cooke (2006) and Aas et al. (2009).

PCC is more flexible than HAC in terms of the range of copula functions possibly used in the construction, whereas it has more parameters to be estimated. Aas and Berg (2009) compare HAC and PCC theoretically and empirically, concluding that PCC is computationally more efficient and less restrictive than HAC. Aas et al. (2009) build a theoretical frame for PCC, provide algorithms and apply the model to a financial portfolio. However, there is still a lack of empirical studies on insurance data, especially in combination with asset modeling in a high-dimensional setting.[131] The PCC method is comprised of three structures: D-Vine (drawable vine copula), C-Vine (canonical vine copula) and R-Vine (regular vine copula). The D-Vine is a structure in a row with the hierarchical construction, the C-Vine describes a dependence

[130] See, for instance, Savu and Trede (2010) with HAC for financial data, Schepsmeier and Czado (2016) with PCC for car crash simulation data and Eling and Jung (2018) with PCC for data breach records. Daul, De Giorgi, Lindskog and McNeil (2003) describe another high-dimensional dependence model, the grouped t-copula, which models heterogeneous levels of tail dependence by aggregating sub-vectors of the d-dimensional vector of variables. We focus on tree (or level)-based multidimensional copula models recently developed, but the grouped t-copula by Daul et al. (2003) provide an alternative approach to model a large set of risk factors from different classes using the t-copula.

[131] There are several studies of risk aggregation for both assets and liabilities (see, e.g., Rosenberg and Schuermann, 2006; Aas et al., 2007; Aas and Berg, 2009 for asset aggregation; Tang and Valdez, 2009; Diers, Eling and Marek, 2012; Bermudez, Ferri and Guillen (2013); Eling and Jung, 2018 for insurance aggregation), but no integrated consideration of assets and liabilities. One relevant study to the correlation issue in Solvency II between market risk and underwriting risk is by Filipovic (2009), who discusses the interplay between the top-level correlation and the base-level correlation. Bølviken and Guillen (2017) construct an internal model incorporating the market, credit and non-life risk modules with Monte Carlo simulation under the lognormal assumption updating skewness, which, however, does not provide an empirical analysis.

structure where a core risk factor connects the rest of factors and the R-Vine links the risk factors according to the level of dependency.[132] The R-Vine as a more general structure can incorporate the C-Vine and the D-Vine according to the dependence structure of random variables.

In addition to these parametric dependence models, we consider the Bernstein copula as an alternative base-level model, which is a type of empirical copula with Bernstein polynomial (Sancetta and Satchell, 2004). We implement the D-Vine structure with Bernstein copula recently developed by Kauermann and Schellhase (2014) to make it comparable with other pairwise models (HAC and PCC).[133] Table 3 illustrates the list of possible copula methods for d-dimensional setting used in our empirical study.

Table 3. List of Copulas (Nelsen, 2006; Aas and Berg, 2009; Kauermann and Schellhase, 2014)

Model	# of parameters	Pros	Cons
Gaussian	$\frac{d(d-1)}{2}$	• Easy to use and interpret. • Normally distributed margins.	• No tail dependence. • Limited to symmetric and linear dependency.
Student-t	$\frac{d(d-1)}{2}+1$	• Easy to use and interpret. • Tail dependence.	• Limited to symmetric dependency.
Archimedean	1	• Easy to construct. • A great variety of copula families.	• Difficult interpretation with a single parameter. • Exchangeability.
Hierarchical Archimedean Copulas (HAC)	$d-1$	• Easy to construct by using generating functions of Archimedean copulas. • A more accurate dependence structure.	• Limited to Archimedean family. • Difficult interpretation. • Complicated form.
Pair Copula Construction (PCC)	$\frac{d(d-1)}{2}$	• A more accurate high dimensional dependency. • A variety of copulas. • Any type of dependence model.	• Difficult interpretation. • No explicit distribution function available.
Bernstein copula (D-Vine)	$\frac{d(d-1)}{2}\cdot(m+1)^2$, where m is the degree of Bernstein polynomial	• Advantageous when parametric copula functions are misspecified. • Flexible by controlling the polynomial degree.	• Still unknown parameter of optimal polynomial degree m. • Not parsimonious.

In the empirical study, we construct two HAC structures with Gumbel and Clayton copula. Different copula functions could be used for different pairs in one structure (i.e., the mixture of different Archimedean copulas), but some copula functions are incompatible with each other; thus, the hierarchical structure cannot be formed by using the generating functions of such

[132] See Aas et al., 2009 for more detail in specific algorithms of D-Vine and C-Vine. See Bedford and Cooke, 2001, 2002; Czado, 2010; Dissmann et al., 2013 for more detail in the density estimation and specification of R-Vine.

[133] For more detail on mathematical formulations for Bernstein approximation and copula density, see Sancetta and Satchell (2004), Pfeifer, Strassburger and Philipps (2009), Diers, Eling and Marek (2012), Kauermann, Schellhase and Ruppert (2013) and Yang et al. (2015).

copulas. For example, Savu and Trede (2010) prove the incompatibility of different generating functions in a hierarchical structure using Gumbel and Clayton copulas. For this reason, we use one copula function for one HAC structure. In addition, we focus on the R-Vine model, which is less restrictive than the D- and C-Vine models and can represent them depending on the dependence structure of variables. To form the R-Vine density, the following steps are required (Dissmann et al., 2013):

Step 1: The tree structure (the level of the structure) needs to be determined to identify linked trees (Czado, 2010, Section 3; Dissmann et al., 2013, Section 2.2).

Step 2: Parametric bivariate copula for each node of each tree needs to be determined.

Step 3: Corresponding dependence parameters to the estimated copula in step 2 need to be estimated.

When a vine model is used, a two-step approach consisting of marginal estimation and copula estimation introduced by Joe and Xu (1996) is employed for the parameter estimation due to the high dimensional optimization of a vine model (Czado, Jeske and Hofmann, 2013). The approach that uses the maximum likelihood method is called the *inference function for margins (IFM)*.[134] The transformation to the uniform distribution ($u_i \in [0,1], i = 1, \dots, d$) in the first tree of a vine model is required and applies to the following trees given the estimation in the previous tree (conditional transformation). The transformation is carried out non-parametrically using the ranks of the observations, which are called pseudo-observations (Aas and Berg, 2009). The pseudo-observations are employed for bivariate conditional copula functions in the sequential estimation in the abovementioned steps, and the optimization of pairwise likelihood function

[134] A simple case of IFM can be illustrated as follows. The d-dimensional probability function for the random vector $\mathbf{X}$ based on Sklar's theorem (1959) is described in the following:

$$F(\mathbf{X};\, \tau_1, \dots, \tau_d, \boldsymbol{\theta}) = C(F_1(x_1; \tau_1), \dots, F_d(x_d; \tau_d); \boldsymbol{\theta}),$$

where $\tau_i, i = 1, \dots, d$ is a parameter of a marginal function F_i, $\boldsymbol{\theta}$ is a set of dependence parameters by the copula function C.

The joint density function of the continuous random variables (vector X) can be derived as:

$$f(\mathbf{X};\, \tau_1, \dots, \tau_d, \boldsymbol{\theta}) = c(F_1(x_1; \tau_1), \dots, F_d(x_d; \tau_d); \boldsymbol{\theta}) \prod_{j=1}^{d} f_j(x_j; \tau_j),$$

where f_j is the corresponding probability density to the marginal function F_j ($j = 1, \dots, d$) and $c(\cdot)$ is a copula density function.

The log-likelihood function for the joint density function derived above has the following form:

$$L(\boldsymbol{\theta}, \tau_1, \dots, \tau_d) = \sum_{i=1}^{d} \log f(x_i; \tau_1, \dots, \tau_d, \boldsymbol{\theta}).$$

The parameter estimation by IFM consists of optimizations for univariate margins and the optimization of the d-dimensional log-likelihood with the dependence parameter. The derivation of the parameters for vine models using IFM has been studied in Haff (2013) and Czado, Jeske and Hofmann (2013).

with pseudo-observations is implemented by maximizing the pseudo-likelihood (for more detail, see Aas et al., 2009; Dissmann et al., 2013).[135]

The selection of a parametric bivariate copula for each node in the second step is conducted by Akaike Information Criteria (AIC)[136] and the Vuong and Clarke test (Vuong, 1989; Clarke, 2007). We consider a range of parametric copula functions to fit for each pair: independence, Gaussian, student-t, Clayton, Gumbel, Frank, Joe, survival Archimedean copulas and rotated Archimedean copulas (90° and 270°). Figure 1 depicts HAC and three types of PCC in the four-dimensional setting, where the first tree of PCC distinguishes the vine models.

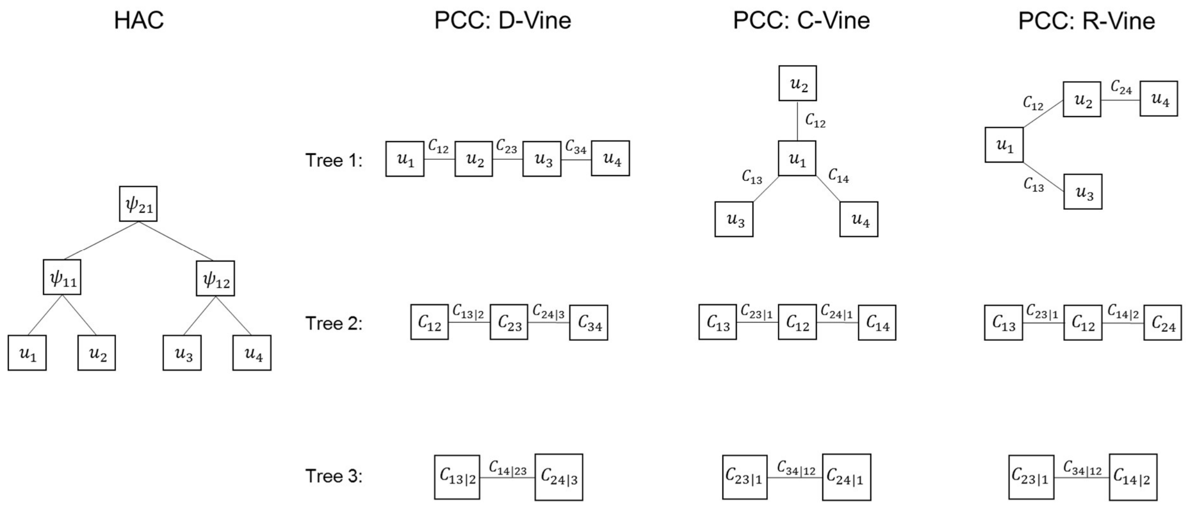


Figure 1. A graphical example of HAC and PCC (four-dimensional case). The mathematical definitions of HAC and PCC (R-Vine) are provided in Appendix B.

The copula models listed in Table 3 are used to aggregate at the base level with the marginal modeling in both asset and underwriting portfolios. Once the base-level aggregation is complete, one would have two aggregate distributions at hand from the two sides of the balance sheet. The estimated dependence structure from each portfolio is inherent in each aggregate distribution. The aggregate distribution for each portfolio can be described as:

$$R_a = \sum_{i=1}^{n_a} w_i X_{i1|C_{asset}} \; (assets) \; and \; R_l = \sum_{i=1}^{n_l} X_{i2|C_{\text{underwriting}}} \; (underwriting), \tag{4}$$

---

[135] The maximum pseudo-likelihood estimation (MPL) is introduced by Genest, Ghoudi and Rivest (1995) and has been developed in Chen and Fan (2006) for time-series copula modeling and Aas et al. (2009) for pair copula construction. The estimation using MPL is basically a semi-parametric approach, consisting of non-parametric marginal transformation and parametric estimation for dependence parameters (Genest et al., 1995).

[136] We use the information criterion defined by Akaike (1973) as follows:

$$AIC := -2Loglik_i(\Theta|\boldsymbol{u}) + 2k,$$

where $Loglik_i(\cdot)$ is the log-likelihood of i-th model, $\Theta = (\theta_1, \dots, \theta_k)$ is a set of parameters, $\boldsymbol{u} = (u_1, \dots, u_d)$ is a d-dimensional set of uniform margins and k is the number of parameters.

where $i = 1, \dots, n_a \; or \; n_l$ is the dimension of each portfolio, R_a is the return distribution of the asset portfolio, X_{i1} is the i-th risk factor of the asset portfolio from the estimated dependence structure (C_{asset}), w_i is the weight of the i-th asset[137], R_l is the loss distribution of the underwriting portfolio and X_{i2} is the i-th risk factor of the underwriting portfolio from the estimated dependence structure ($C_{insurance}$).

With the aggregate distributions, R_a and R_l, we model the dependence structure using pseudo-observations of R_a and R_l at the top level and aggregate two distributions with Sklar's theorem (1959).

$$F(x_a, x_l) = C(\hat{u}_a, \hat{u}_l), \tag{5}$$

where x_a is the aggregate asset distribution, x_l is the aggregate liability distribution, $\hat{u}_a = F_a(R_a)$ is the vector of the pseudo-observations for the asset portfolio and $\hat{u}_l = F_l(R_l)$ is the vector of the pseudo-observations for the underwriting portfolio. The probability functions of both aggregate distributions are conditional functions (copula distributions) given the sub-risks for both portfolios at the base level. A range of copula functions considered at the top level are enumerated in Section 4.3.

We note that the possible problems in multi-level risk aggregation under the current insurance regulatory frameworks are twofold. One problem results from the linear correlation assumption between risk factors, a problem that we address with our pairwise dependence model. A second problem is a possible interplay between risk factors from different classes, which Filipovic (2009) investigates by comparing the Solvency II two-level approach (i.e., a base correlation matrix within each risk class and a top-level correlation matrix between these risk classes) with an internal model using a bottom-up approach for a large correlation matrix of all risk types. Our modular two-level approach is consistent with the currently used standard models and thus only addresses the first problem, whereas Filipovic (2009) addresses the second problem. In this regard, our model complements the results by Filipovic (2009).[138] Figure 2 illustrates the

---

[137] In the application part, we employ the asset allocation based on the financial statement of the insurer for the aggregate asset distribution (see Table 4). Then, we also consider two asset allocation strategies to see the effect of each strategy on the estimation of the economic capital (see Table D2 in Appendix D).

[138] Filipovic (2009) shows that only correlation parameters set at the base level lead to unequivocally comparable solvency capital requirements across the industry. Beyond that, he also stresses that the correlation aggregation does not appropriately capture tails and tail dependence of risks and encourages the additional use of risk and dependence modeling, which is what we present in this paper. To compare our data with Filipovic (2009), we also measure the linear correlations across risk classes based on the QIS3 calibration method that Filipovic (2009) follows, given the asset allocation for each data. For the Korean data, the linear correlations between premium/reserve (non-life) risk and the equity factors are: KR_stock: 8.6%; KR2Y: 2.6%; KR5Y: 4.9%; KR10Y: 6.0%; KRcor: 4.3%; KRM3: 8.5%; Wrd_real: 8.6%; For the German data, Wrd_stock: 19.9%; EMU_stock: 20.1%; DE_stock: 11.0%; US2Y: 7.2%; DE2Y: 3.1%; EMU2Y: 2.1%; IBOXX_cor: 3.1%;

integrated structure we use to derive the total loss distribution from the asset and liability side. Figure 3 summarizes our methodology.

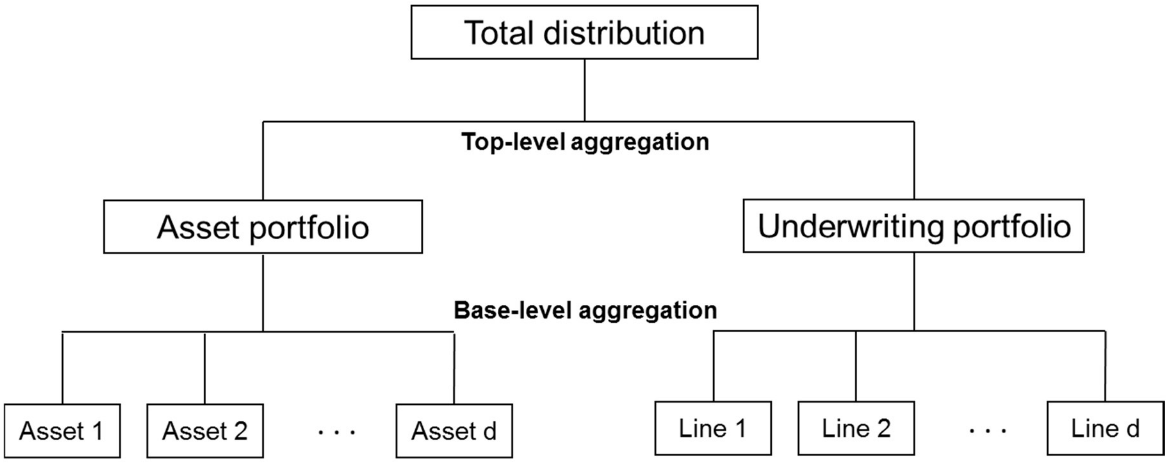


Figure 2. Integrated structure of risk aggregation for a non-life insurer.

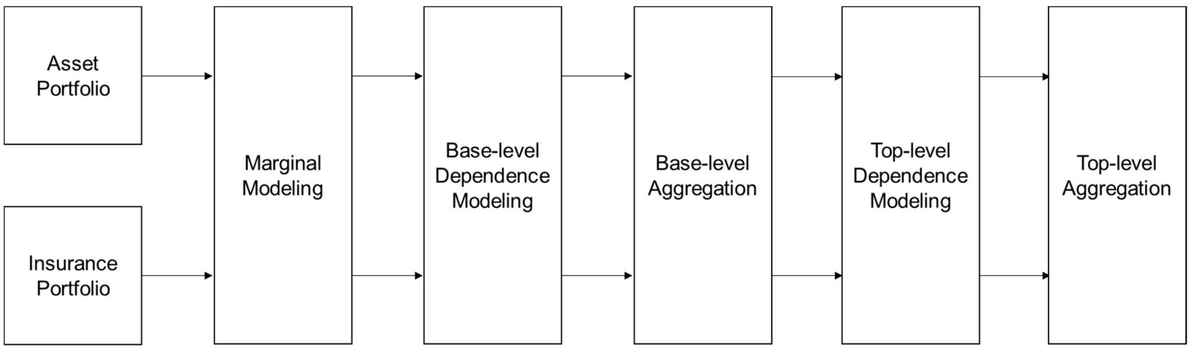


Figure 3. Process of the methodology implemented.

3 Data

In the empirical study, we consider a realistic portfolio selection for both assets and liabilities. Asset returns are measured with widely used benchmark indices. These indices consist of stocks, bonds, real estate and money market, usually accounting for most investments by non-life insurers (Eling, Gatzert and Schmeiser, 2009). A well-known benchmark asset portfolio for a large Korean non-life insurer is shown in Table 4, the portfolio that we consider to construct an internal model for market risk (Kim, 2015). The asset allocation in Table 4 is based on the financial statement of the insurer in 2016, allowing for adopting a realistic investment strategy. Monthly log returns from January 2002 to December 2016 (180 observations for each asset)

EURM3: 10.2%; Wrd_real: 3.6%; EU_real: 6.7%. These numbers are comparable with Figure 10 in Filipovic (2009), where the correlation between underwriting risk and equity risk is relatively high, whereas the correlation between underwriting risk and interest rate risk is relatively low.

are taken from Datastream. We use the total return index for the equity and the real estate securities, which consider the dividend reinvestment performance (Eling et al., 2009). For the insurance portfolio, we derive historical aggregate loss distributions of a Korean insurer from five lines of business (fire, motor, marine, liability and accident).[139] The loss distributions are also on a monthly basis and for the same time period. Table 5 presents the descriptive statistics for both portfolios. Appendix G gives corresponding information for the German dataset.

Table 4. Benchmark Indices (Kim, 2015)

Asset class	Index	Description	Asset allocation
Equity	KR_stock	MSCI Korea Index	5.0%
Fixed income	KR2Y	Korea 2-year Sovereign Bond Index	20.0%
	KR5Y	Korea 5-year Sovereign Bond Index	20.0%
	KR10Y	Korea 10-year Sovereign Bond Index	20.0%
	KRcor	AA 3-year Corporate Bond Index	20.0%
Money Market	KRM3	Korea standardized money supply M3	5.0%
Real estate[140]	Wrd_real	MSCI World Real Estate Index	10.0%

Note: MSCI stands for Morgan Stanley Capital International. The asset allocation is based on the 2016 annual report of the insurer.

Table 5. Descriptive Statistics

Panel A: Asset portfolio

	mean	sd	skewness	kurtosis	Max	median	min	JB-test
KR_stock	0.0081	0.0771	-0.3898	1.1737	0.2341	0.0087	-0.3028	15.81***
KR2Y	-0.0001	0.0118	0.0256	5.3170	0.0610	0.0019	-0.0444	219.55***
KR5Y	0.0001	0.0173	-0.5500	2.1224	0.0569	0.0030	-0.0615	44.87***
KR10Y	-0.0002	0.0252	-0.4863	2.8502	0.1039	0.0038	-0.0835	70.98***
KRcor	0.0039	0.0047	2.8663	16.6943	0.0385	0.0035	-0.0049	2,396.48***
KRM3	0.0071	0.0336	-0.2851	3.7564	0.1490	0.0100	-0.1289	112.62***
Wrd_real	0.0071	0.0564	-1.3226	6.7998	0.2050	0.0129	-0.3244	411.43***

Panel B: Insurance portfolio (₩ billion)

	mean	sd	skewness	kurtosis	Max	median	min
Fire	0.892	0.799	3.444	20.175	7.102	0.714	0.043
Motor	191.518	53.432	0.614	-0.443	356.089	174.658	97.938
Marine	2.655	2.280	2.198	6.731	14.122	2.031	0.041
Liability	2.870	1.928	1.193	1.724	11.594	2.221	0.253
Accident	14.289	8.826	0.094	-1.071	33.313	14.456	1.171

Note: The numbers in panel A are derived from geometric return series. *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. JB-test stands for Jarque-Bera test for normality assumption on the residual. ₩ stands for Korean Won.

139 The insurance loss data is derived from insurance statistics information services (INsis) operated by the Korea Insurance Development Institute (KIDI). The data is from the largest non-life insurer in the Korean non-life insurance market (with almost a quarter of the market share) in terms of total asset size. We construct our internal model with this dataset and check the robustness of our model with additional data from four companies in the Korean market, which are at quartiles with regard to total asset size and have the same lines of business in the underwriting portfolio (see Appendix F). In addition, we compare the economic capitals of four companies estimated by our internal model with the regulatory measures under the K-RBC to see the variation between the internal model and the standard model (see Table 9).

140 For the real estate index, local indices such as MSCI Korea or Asia-Pacific area are not available for the study period. The world index is also relevant to the possible investment of a Korean insurer, especially nowadays after the deregulation on real estate investment abroad in 2014 (FSC, 2014).

The Jarque-Bera test results on assets (panel A of Table 5) show that no return distribution satisfies the normality assumption. Figure C2 in Appendix C shows that most assets are either left- or right-skewed and leptokurtic. An asset return series is typically time-varying, so autocorrelation and heteroscedasticity need to be considered (Bollerslev, 1986). We fit ARMA-GARCH model with different innovations for the dataset to resolve this issue, as described in Section 4.1. The aggregate losses from the five insurance lines are right-skewed (Figure C3), among which motor insurance generates substantially larger losses than other lines.

4 Marginal modeling

4.1 Marginal modeling

The insurance modeling (individual risk model and collective risk model) is designed to estimate the total (aggregate) loss for a certain time period in the risk pool, where the losses are typically modeled by statistical tools for right-skewed frequency and severity (Cummins, 1991; Embrechts, 2002; Frees, 2015). In contrast, the financial return series are stationary (or linear) processes with the statistical features of mean-reversion and volatility clustering, which are modeled with different tools compared with the insurance claims modeling (Embrechts, Kluppelberg and Mikosch, 2013, Chapter 7).

Autocorrelation and heteroscedasticity for the asset portfolio are investigated before the data is used in the dependence modeling. We find autocorrelation and that the assumption of the equal variance of residuals is violated for most assets, showing the presence of volatility clustering and fluctuation (see Figure C1 in Appendix C). Based on the graphical identification of autocorrelation and heteroscedasticity, ARMA and GARCH models are specified (Bollerslev, 1986; Box, Jenkins and Reinsel, 1994).

Conditional mean model (ARMA[p,q]):

$$r_t = \mu + \varphi_1 r_{t-1} + \cdots + \varphi_p r_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \tag{6}$$

$$E[\varepsilon_t] = 0 \; and \; Var[\varepsilon_t] = \sigma_t^2,$$

Conditional variance model (GARCH[1,1]):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{7}$$

where r_t is a return at time t, μ is the drift term, φ_p is an auto-regressive coefficient, θ_q is a moving average coefficient and α_1 and β are GARCH coefficients.

For the conditional mean model, we test autocorrelation and moving average with different lags. For the conditional variance model, we fit GARCH(1,1) for each asset, but there is a need to apply heavy-tailed distributions due to the presence of extreme values observed in QQ-plots (see Figure C2).[141] Thus, the procedure to fit conditional variance is conducted in different parametric settings by specifying normal, skew normal, student-t and skew student-t distributions, all of which are widely used in the literature to analyze financial time-series data (see, e.g., Bollerslev, 1986; Bollerslev and Wooldridge, 1992; Jondeau and Rockinger, 2006; Eling, 2014). Many studies of conditional volatility modeling use the Gaussian distribution (Engle, 1982); we also consider skewed and leptokurtic distributions to find a better fit.

The ARMA-GARCH model for each return distribution is determined by the minimum Akaike information criteria (AIC) among different candidates. Table C1 shows that all assets have long-tailed residual distributions fitted by student-t, skew normal and skew student, consistent with the JB tests results. By specifying the ARMA-GARCH process for the marginal distributions of assets, intertemporal dependence is removed. The resulting standardized residuals are transformed to pseudo-observations (uniform margins for copula modeling) on the interval [0,1] using the probability integral transformation (Aas and Berg, 2009; Czado, Min and Schepsmeier, 2012). Pairwise scatter plots and Kendall's rank correlations of uniform margins in the left panel of Figure C4 (the right panel for the underwriting portfolio) show strong correlations between fixed income assets, generally above +0.2, whereas the correlation between different asset classes tends to be small (see, e.g., Avanzi et al., 2016).

With regard to the insurance portfolio, loss data consist of the monthly aggregate loss from each line of business. We fit different parametric distributions to the aggregate loss data to derive the copula margins under the individual risk model. Twelve continuous distributions frequently used in the actuarial and operational risk context (see, e.g., Frachot, Georges and Roncalli, 2001; Moscadelli, 2004; Fu and Moncher, 2004; Shevchenko, 2011; Eling, 2012; Frees, Lee and Yang, 2016; Eling and Jung, 2018) are considered: skew-normal, student-t, skew-student-t, lognormal, gamma, Weibull, inverse Gaussian, Cauchy, Burr, generalized Pareto distribution (GPD) and two peaks-over-threshold (POT) models with normal and lognormal in the body respectively and GPD in the tail. We determine the most appropriate distribution for each loss process based on AIC and goodness-of-fit (GoF) test (Table C2). We find that the skew student-t and gamma distributions best fit the historical losses in four cases and the lognormal is the best fit for the

[141] We fit GARCH(1,1) for the conditional volatility, which is widely used to model the volatility clustering of the financial time series data in the pair copula modeling context (see, e.g., Aas and Berg, 2009; Aas et al., 2009; Brechmann and Schepsmeier, 2013; Dissmann et al., 2013).

loss process of motor insurance. This finding is confirmed by QQ-plots in Figure C3, showing that the quantiles from fitted distributions are plotted along with the theoretical distributions. With the fitted distributions for insurance losses, we generate the copula margins and estimate the Kendall's rank correlations shown in Figure C4.

4.2 Dependence modeling and base-level aggregation

In order to estimate the dependence models for both portfolios, we first transform the variables to the estimated marginal distributions on [0,1]. Transformed uniform margins are applied to the copula functions listed in Table 3. We then compare the models to identify which model is a better fit for the dependence structures of both portfolios. For PCC modeling, we consider R-Vine model on a dependence structure. Table 6 provides three statistical measures to compare different models: log-likelihood, AIC and goodness-of-fit test result.

The R-Vine structure most accurately estimates the dependence structures of both portfolios, followed by the student-t model for the asset and by HAC-Gumbel for the underwriting portfolio. This finding is supported by literature in which vine models turn out to be superior to conventional copula models (see, e.g., Aas and Berg, 2009; Aas et al., 2009; Brechmann and Czado, 2013; Low et al., 2013; Brechmann et al., 2014; Schepsmeier, 2015; Eling and Jung, 2018).[142] In contrast, simple Archimedean copulas (Gumbel and Clayton) do not provide a good fit to both cases, which might result from the fact that estimating a high-dimensional dependence structure by Archimedean copulas is limited by one single parameter. This finding can be confirmed by the result of HAC as an extended model of simple Archimedean copulas for a high-dimensional setting in that HAC shows a better fit.

The Bernstein model[143] provides a comparable log-likelihood measure to that of R-Vine; however, as described in Table 3 and proven by AIC, the model is not parsimonious due to the polynomial degree. Elliptical models are also a good fit for the structure in both cases, especially the student-t model, which implies significant tail dependency in both portfolios. The

[142] Aas and Berg (2009) demonstrate the superiority of vine models to HAC, whereas Aas et al. (2009), Brechmann and Czado (2013), Brechmann et al. (2014) and Schepsmeier (2015) show a better fit of vine model than that of elliptical models. Low et al. (2013) and Eling and Jung (2018) find that vine models produce more outperforming statistical outcomes than do Archimedean copulas.

[143] The log-likelihood of Bernstein D-Vine is a penalized version of the log-likelihood, which can be estimated based on equation (16) in Kauermann and Schellhase (2014). AIC is also based on the penalized log-likelihood and corrected by using degree of freedom of the penalty parameter explained in equation (20) of Kauermann and Schellhase (2014). The goodness-of-fit test for Bernstein copula model (non-parametric case) has not been developed, since the p-value generation by the bootstrap method can generate inconsistent results for non-parametric case (Genest and Remillard, 2008). The determination of the optimal degree of the polynomial is still an open question as written in Table 3 (Pfeifer et al., 2009; Diers et al., 2012); thus, we implement four cases (K=5,10,15,20) and present the model (K=10) with the lowest AIC in Table 6.

tail dependency is numerically specified in the pairwise setting in Table C3 (Appendix C), where estimated dependence parameters in the first tree are also present. In the first tree, the parameters are derived from unconditional copula densities, which normally show a stronger dependency than those from conditional copula densities in the following trees. The mathematical proof of this statement is described in Dissmann et al. (2013). Savu and Trede (2010) and Okhrin, Okhrin and Schmid (2013) also explain this property in HAC structure.

Table 6. Dependence Modeling Results (Base level)

		Asset Portfolio			Insurance Portfolio		
Family	Copula	Log-lik	AIC	GoF [144]	Log-lik	AIC	GoF
Elliptical	Gaussian	189.73	-337.47	0.037**	270.62	-521.25	0.158***
	Student-t	209.31	-374.63	0.020	278.38	-534.77	0.063***
Archimedean	Gumbel	52.60	-103.20	0.046***	91.52	-181.04	0.275***
	Clayton	58.70	-115.39	0.034	46.23	-90.47	4.169***
HAC	Gumbel	152.04	-292.08	0.735	279.86	-551.71	0.584
	Clayton	162.14	-312.28	0.082*	159.15	-310.30	0.039**
PCC	**R-Vine**	**228.41**	**-416.81**	**118.71**	**294.51**	**-577.01**	**6.514**
Bernstein (D-Vine)		222.80	-87.29	-	278.27	-197.97	-

Note: Elliptical and Archimedean copulas are estimated via R package `copula`, HAC models are estimated via `copula` and `HAC`, R-Vine model is implemented via `VineCopula` and Bernstein D-Vine model is estimated via `penDvine`. The numbers in the goodness-of-fit results indicate the test statistics except for those of HAC models showing p-values (see footnote 142). *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit method for each portfolio.

Figure C5 displays the dependence structures of both portfolios in the first tree of R-Vine as the best fit. The graphical description of the estimated aggregate distributions at the base level with 100,000 samples is found in Figure 4 (based on R-Vine model as the best fit). The aggregate asset distribution is leptokurtic, which is in line with the stylized fact of financial returns, and a long-tail risk is identified from the insurance aggregate distribution.

144 The goodness-of-fit test for copula methods is based on the information matrix equality of White (1982), developed by Huang and Prokhorov (2014) and applied to vine models by Schepsmeier (2019). The test specification in a general case can be described as (Huang and Prokhorov, 2014; Schepsmeier, 2019):

$$H_0\colon \mathbb{H}(\theta) + \mathbb{C}(\theta) = 0,$$

where $\mathbb{H}(\theta) = E\left[\partial_\theta^2 \ln\left(c_\theta(u_1, \ldots, u_d)\right)\right]$ is the expected Hessian matrix of the score function (second-order derivative) and $\mathbb{C}(\theta) = E\left[\partial_\theta \ln\left(c_\theta(u_1, \ldots, u_d)\right)\left(\partial_\theta \ln\left(c_\theta(u_1, \ldots, u_d)\right)\right)^T\right]$ is the expected outer product of the corresponding score function.

When it comes to HAC model, the statistical closed-form of the goodness-of-fit test has not been developed (Savu and Trede, 2010; Okhrin and Ristig, 2014). Instead, we use the goodness-of-fit for copulas in higher dimensions introduced by Hofert and Mächler (2014), the test whose results are p-values (whereas other models show the test statistics) shown as the GoF result for HAC in Table 6. This test provides a global p-value based on pairwise Rosenblatt transform and enables to identify pairs that do not follow the null hypothesis as follows:

$$H_0\colon C_{\boldsymbol{\theta}}\ is\ HAC\ Gumbel\ (Clayton)\ structure\ with\ estimated\ parameters$$

For more detail on the test specification and the global p-value, see Hofert and Mächler (2014).

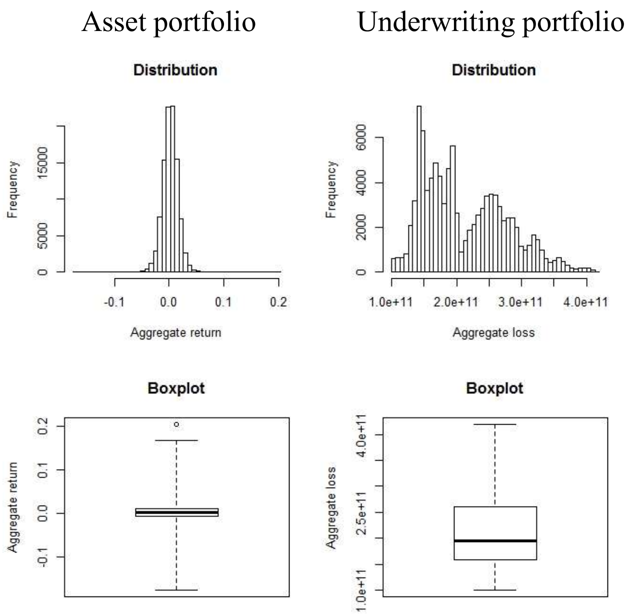


Figure 4. Base-level aggregate distributions.

4.3 Top-level aggregation

This section investigates the top-level dependency between the market risk module and the non-life risk module. Both the K-RBC and SII assume 0.25 correlation between the asset portfolio (market risk) and the insurance portfolio (non-life risk) (see Table 1, 2). However, this correlation does not adequately reflect the dependence structure in the empirical setting; hence, the standard model might lead to a significant gap between the required capital and the actual size of risk.

First, we conduct the independence test to check the validity of the independence assumption. The test result for the independence copula model is derived based on a bivariate asymptotic independence test of empirical Kendall's tau (see Genest and Remillard, 2004, for more detail on the test specification). Second, we conduct goodness-of-fit tests for other bivariate copulas under the top-level correlation assumption by regulations (= 0.25).[145] Four types of bivariate copula functions considered at the top level are mathematically described as follows (Nelsen, 2006; Eling and Toplek, 2009).

Independence copula (independence assumption):

$$C^{Ind}(\mathbf{u}) = \prod\left(C^{\text{Asset}}(u_1^a, \dots, u_k^a), C^{\text{Underwriting}}\left(u_1^l, \dots, u_d^l\right)\right), \tag{8}$$

[145] Here we do not consider Bernstein copula as done at the base level because 1) no statistical testing tool has been developed for this type as discussed in footnote 141; and 2) it is an empirical-type copula that does not allow the calibration of a dependence parameter.

Gaussian copula (linear correlation assumption):

$$C^{Gauss}(\mathbf{u}) = \mathbf{\Phi}_{\theta}\left(C^{\text{Asset}}(u_1^a, \dots, u_k^a), C^{\text{Underwriting}}\left(u_1^l, \dots, u_d^l\right)\right), \tag{9}$$

Student-t copula (symmetric tail correlation assumption):

$$C^{t}(\mathbf{u}) = \mathbf{t}_{\text{v},\theta}\left(C^{\text{Asset}}(u_1^a, \dots, u_k^a), C^{\text{Underwriting}}\left(u_1^l, \dots, u_d^l\right)\right), \tag{10}$$

Archimedean copula (asymmetric dependence assumption):

$$C^{Archi}(\mathbf{u}) = \boldsymbol{\psi}^{-\mathbf{1}}\left(C^{\text{Asset}}(u_1^a, \dots, u_k^a), C^{\text{Underwriting}}\left(u_1^l, \dots, u_d^l\right)\right), \tag{11}$$

where $\mathbf{\Phi}_{\theta}$ indicates a multivariate normal density with the correlation parameter of θ and $\mathbf{t}_{v,\theta}$ stands for a multivariate student-t density with degree of freedom, *v*, and the correlation parameter of θ.

Table 7 shows the test result for the presence of the independence and the GoF tests for other copulas, where the independence assumption holds for two marginal distributions. For other copulas with the correlation of 0.25, the tests reject the null hypothesis mostly at the 1% critical level. Therefore, the regulatory assumption of 0.25 is not adequate for modeling the top-level dependency of the empirical data. We also test goodness-of-fit for the top-level aggregation with the linear structure (Gaussian) at the base level, which we assume is closer to the regulatory framework. We find that the test leads to the same result as that with the best fit model at the base level, implying that the independence structure is a global fit for the top-level aggregation in this empirical case.

The independence assumption at the top level is reasonable in that the degree of dependency at a lower level of a hierarchical structure can be mathematically proven to be stronger than that at a higher level because the dependency at a higher level is conditional (see, e.g., Savu and Trede, 2010; Dissmann et al., 2013). Therefore, it can be inferred that the correlation at the top-level aggregation as assumed in the K-RBC and SII might not be mathematically justified under a hierarchical dependence structure. It also proves that the standard models aim at achieving conservative results, which do not adequately reflect the top-level dependency driven by the empirical case in our study. This finding is in line with the literature that for non-life insurance (property and casualty) a dependent feature between assets and liabilities rarely exists (Gründl, Dong and Gal, 2016).

There might still exist dependence between the assets and liabilities of a non-life insurance company, particularly due to interest rate risk. For example, non-life insurers with long-tail liabilities (e.g., workers' compensation insurance) might be more exposed to interest rate risk than non-life insurers with mainly short-tail liabilities (Gilbert, 2016). Also changes in discounting practice (new accounting standards requiring discounting of reserves) might influence the interest rate sensitivity and create dependence.

Table 7. Goodness-of-fit Results for the Top-Level Aggregation

		R-Vine (Base level)		Linear (Base level)	
		Statistics	P-value	Statistics	P-value
Independence test		**1.0813**	**0.280**	**0.3464**	**0.729**
Elliptical family	Gaussian	0.5438***	0.005	0.3327**	0.035
	Student-t	0.5826***	0.004	0.3767**	0.029
Archimedean family	Gumbel	0.5168**	0.011	0.3267**	0.049
	Clayton	0.6248***	0.002	0.3802**	0.018

Note: The independence test is implemented via R package `copula` and the tests for other copulas are conducted via `gofCopula`. *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively.

5 Applications

5.1 Application to Korean and German data

The purpose of estimating the economic capital is to identify the amount of capital necessary to prevent an insurer from becoming insolvent within a predefined time horizon and confidence level. The definition of one-period solvency in our model with the historical data is the excess of the assets over the liabilities expressed as:

$$A_1 - L_1 = A_0 \cdot (1 + r) - L_1 > RC_\alpha, \tag{12}$$

where A_1 is the market value of the asset portfolio consisting of equity, bonds, money market and real estate instruments at time 1. L_1 is the market value of the policyholders debt at time 1 represented by the aggregate loss from different lines of business in our model. A_0 is the market value of the asset portfolio at time 0 (present), which is the invested asset from the balance sheet in our model. We set A_0 at ₩18.6 trillion obtained from the same insurer for which we also take underwriting data.[146] r is a stochastic aggregate return on the asset portfolio estimated by our model. RC_α is the required capital at α confidence level, for example 99% in the K-RBC or 99.5% in SII. The time horizon defined by regulations is one year. Since the regulatory

[146] The statistical information on the assets of Korean insurers can be also obtained from INsis (see footnote 137). We derive the amount of the investment by downsizing the annual amount of the invested asset for the insurer into the annual amount with five business lines based on net earned premium (five lines account for 31.9%).

standards have an annual time horizon, we annualize the monthly asset return distribution and the monthly loss distribution.[147]

Based on the results from Section 4, we choose the independence copula to describe the top-level dependency and the R-Vine structure for the base-level structures. In Appendix D, we consider other dependence structures at the base and top levels to identify the impact of different dependencies on the economic capital. We estimate VaR at 99% and 99.5% and Tail Value at Risk (TVaR) at 99%, which are the current regulatory requirements for the K-RBC, SII and SST respectively. In Table 8, we compare our capital estimates to the required amounts of capital from the three regulatory standard models. Next to the Korean data, we present the results for our second, German dataset. Appendix E contains all conceptual details on the comparison between the standard and internal models; all implementation and estimation details for the German data are presented in Appendix G.

Table 8. Comparison of the Internal Model and Standard Model

Panel A: Korean Data	VaR 99% (K-RBC)	VaR 99.5% (SII)	TVaR 99% (SST)
Internal Model (₩ billion)	559.7	744.5	842.5
Standard Model (₩ billion)	846.3	1,177.5	1,471.5
Absolute difference (₩ billion)	286.6	433.0	629.0
Relative difference	51.2%	58.2%	74.7%
* from risk parameters	15.3 pp	24.6 pp	63.0 pp
* from correlation	35.9 pp	33.6 pp	11.7 pp
Panel B: German Data	VaR 99% (K-RBC)	VaR 99.5% (SII)	TVaR 99% (SST)
Internal Model (€ million)	31.3	51.2	55.3
Standard Model (€ million)	45.5	80.7	94.4
Absolute difference (€ million)	14.2	29.5	39.1
Relative difference	45.4%	57.6%	70.7%
* from risk parameters	9.3 pp	14.8 pp	62.9 pp
* from correlation	36.1 pp	42.8 pp	7.8 pp

Note: The internal model is the best fitted model (R-Vine-independence) in our estimation procedure. "pp" stands for percentage points.

The three standard models significantly overestimate the capital requirement, showing the variation size of 61.2% on average for the Korean case and 57.8% for the German case compared to the total risk estimated by the internal model.[148] The K-RBC provides the least

[147] We aggregate 12 consecutive monthly returns and losses to the annual size as follows:

Asset: $r^a = \prod_{i=1}^{12}(1 + r_i^m) - 1,$

Underwriting: $X^a = \sum_{i=1}^{12} X_i^m.$

The superscripts, a and m, indicate the annual value and the monthly value respectively.

[148] Previous studies criticize the SII standard formula comparing with internal modelling approaches, but have different aspects on the criticism. For instance, Gatzert and Martin (2012) argue that the predefined scenarios by SII, which over- or underestimate the actual risk, do not sufficiently reflect the insurance company's risk situation by comparing the standard model with their partial internal model. Mittnik (2011) identifies that the

conservative measure with VaR 99% showing the smallest gap between the internal model and the standard model, whereas SST turns out to be the most conservative model with TVaR 99% showing the largest deviation among three standard models.[149] This might be due to the fact that the K-RBC does not consider the correlation across market risks, resulting in no diversification effect in this module. Furthermore, the categorization in the underwriting risk module does not appropriately reflect the diversified lines of business, thereby providing a less conservative measure than the other two models.

The observed deviations are relatively large in comparison with the literature. For example, Tang and Valdez (2009) document a 5.4% deviation (overestimation) considering the underwriting risk module only. Christiansen et al. (2012) do not quantify the deviation between the internal model and the standard formula, but they identify a significant overestimation of economic capital resulting from the correlation assumption of the life underwriting module in SII. Our results also surpass most practitioners' expectations of internal models.[150]

The commonly identified overestimation of the economic capital from the standard models comes from two factors. The first factor is that the distributional features of individual risk factors are not adequately taken into account. Thus, the calibrated parameters (risk coefficients in the K-RBC and SII) and the distributional assumption (SST) can amplify the capital requirement. To measure the size of overestimation due to the fixed and predefined risk parameters, we apply our aggregation model with assumed risk parameters predefined by the standard models, which produces economic capital of 645.1, 927.3 and 1,373.5 (in ₩ billion) for the Korean case and 34.2, 58.8 and 90.1 (in € million) for the German. The predefined risk parameters imposed by the standard models thus contributes on average 34.3 and 29.0 percentage points for the Korean and German cases respectively to the empirical deviations, which are 56% and 50% of the total overestimation sizes. The second factor is the correlation assumption between risk factors, which does not reflect the dependence structure inherent in firm-specific data. Ceteris paribus, the linear correlation assumptions imposed in the standard

misestimated correlation measures between assets do not appropriately reflect the diversification effects. Pfeifer and Strassburger (2008) discuss that the overall SCR can be misspecified in case of symmetric aggregate distribution (linear) and correlated underlying risk factors.

149 Some practice-oriented papers and reports have emphasized the differences in regulatory frameworks. For example, Holzmüller (2009) conducts a comparison analysis for the U.S. RBC standard, Solvency II and SST by evaluating eleven criteria. Siegel (2016) compares SST and Solvency II with a focus on insurance groups and using five criteria.

150 According to a survey with 160 insurance companies in 19 European countries, 37% of the respondents expect the internal model to decrease the capital requirement by 10-20%, 26% of them expect 20-30% and 14% of them expect more than 30%. See EY (2013).

models contribute to 26.9 percentage points on average for the Korean case and 28.8 percentage points on average for the German.

As robustness tests, we implement different asset allocation strategies to analyze the extent to which differences between the standard model and the internal model are driven by the risk on the asset side (see Tables D2 and G5 in Appendix D and G). We find that the riskier the asset portfolio is, the more capital is required, resulting in a bigger gap between the internal model and the standard models. In Appendix D, we analyze the diversification benefits of the internal models compared to benchmark models; our best fit model provides a higher diversification benefit compared to other possible internal models, possibly offering a better optimization of the capital structure (see Figure D1).

5.2 Sensitivity analyses with respect to company size

One critical concern in internal models is how far the use of internal models might distort competition, if small companies lack the resources and expertise to implement such models (see Eling, Schmeiser and Schmit, 2007). Our empirical setup allows us to analyze this important policy question, because we have data for the complete Korean market and can analyze companies of different sizes. We therefore estimate the economic capital for four additional non-life insurers (in different size quartiles of total assets) for the Korean market in Table 9.[151] All selected companies operate the five lines of business considered in our study so that the outcomes are not driven by the complexity of the businesses; considering less complex companies will reduce the benefit of implementing an internal risk model. We first examine whether our internal model (R-Vine at the base level and independence at the top level) also fits well for these companies as robustness check and find that it holds for all these cases (see Appendix F). We then use the internal model to calculate the capital size for each company and compare it with the required capital under the corresponding standard model (the K-RBC).

The result shows an overestimation of 39.0% on average, which declines with company size. Large insurers as well as medium-size companies (such as the 1st and 2nd qrt company in Table 9) can obtain a significant financial advantage by using the internal model, by saving ₩287 billion, ₩133 billion and ₩45 billion in capital. In contrast, implementing the internal risk model might not be worthwhile for smaller companies (such as the 3rd qrt or smallest company in Table 9), since the capital savings might not justify the development of an internal model. We can

[151] We also obtain the insurance loss data for additional four companies from the same database as the one we obtain the main dataset. In this implementation, we apply the same benchmark portfolio used in the main model to four companies.

illustrate the potential cost saving by multiplying the capital reduction for the largest and smallest companies (₩287 billion and ₩3.3 billion) with a standard cost of capital rate of 6% as used, for instance, in Solvency II (EC, 2014, p. 39), leading to a cost reduction of ₩17.2 billion and ₩0.198 billion, respectively per annum.

For the largest company, investment in the construction of an internal model could be worthwhile, but not necessarily for the smallest company. For example, if developing an internal model needs five employees, it might already cost ₩0.162 billion over a year without consideration of further costs, based on statistics for 2016 average Korean salary (= $32,400; Statista, 2018) with an exchange rate of ₩1,000/$. Additional costs or resources to the development of an internal model might come from e.g., maintenance and validation of the internal model, the collection of the necessary, up-to-date data as inputs or the internal assessment prior to the approval process (Milliman, 2008). For a small company it is thus less clear whether developing an internal model is worthwhile. Note that the advantage of implementing an internal model decreases both on an absolute and on a relative basis, in relation to the K-RBC and in relation to asset size.

Table 9. Variation of Company Size

	(Largest)	(1st qrt)	(2nd qrt)	(3rd qrt)	(Smallest)
Internal Model (in ₩ billion)	559.7	292.7	103.0	25.4	10.9
*RBC ratio	620%	418%	258%	220%	812%
K-RBC (in ₩ billion)	846.3	425.2	148.2	34.7	14.2
*RBC ratio	410%	288%	179%	161%	623%
Absolute difference (in ₩ billion)	286.6	132.5	45.2	9.3	3.3
Relative difference	51.2%	45.3%	43.9%	36.6%	30.3%
*Risk parameter	15.3 pp	16.7 pp	19.6 pp	15.7 pp	12.8 pp
*Correlation	35.9 pp	28.6 pp	24.3 pp	20.9 pp	17.5 pp
Invested assets (in ₩ billion)	18,577.8	9,025.2	3,420.3	691.2	178.6
Asset size (in ₩ billion)	21,659.4	10,562.9	3,987.0	821.3	285.4
Equity capital (in ₩ billion)	3,471.0	1,224.0	265.7	55.9	88.5
Abs. difference/asset size	1.32%	1.25%	1.13%	1.13%	1.16%

Note: Four companies are chosen out of the Korean non-life insurers, who operate the five lines of business considered in our study, by the quartiles of the total asset size. Company 1 is at the first quartile (25%) from the largest company, company 2 is at the second quartile (median), company 3 is at the third quartile and company 4 is the smallest. Invested assets (A0 in equation (12)), total asset size and equity capital are downsized by the share of five insurance lines compared to the total non-life business for each company based on net earned premium.

6 Conclusion

The aim of this paper is to construct an internal risk model using a recently developed dependence method and apply it to empirical datasets in order to compare the internal model with three regulatory standards models (the K-RBC, SII and SST). We introduce a two-step aggregation methodology with the pair copula construction model (vine copula model) to estimate the economic capital for several Korean non-life insurers and a mid-sized German non-life insurer. For all datasets, the R-Vine structure at the base level and the independence structure at the top level turn out to be the best fit based on the statistical testing procedure. The comparison between the proposed internal model and the standard models shows that the standard models overestimate the economic capital by on average 61.2% for the Korean case and 57.8% for the German, implying that insurers can significantly reduce their risk capital using the proposed internal risk model. Our internal model is also proven to be a good fit for additional companies of different sizes in the Korean market, which again supports our finding on the overestimation of the risk capital by the standard model (39.0% on average).

We focus on two regulatory factors resulting in the overestimation: the calibrated parameters for marginal risks and the linear correlation assumption. The regulated risk parameters in the standard models address on average 56% (34.3 percentage points) and 50% (29.0 percentage points) out of the total deviation for the Korean and the German cases respectively, whereas, ceteris paribus, the rest of the deviation is led by the correlation parameters in the linear setting for each case. We identify that the K-RBC system is less conservative than SII and SST, based on the fact that the K-RBC system requires the estimation at a lower quantile, does not consider the correlated risk between the sub-modules in the market risk leading to no diversification effect and does not appropriately elaborate the categorization of the non-life underwriting risk module. The proposed internal model provides a higher diversification benefit than other models do, possibly leading to a better optimization of the capital structure.

The literature has already found over- or under-estimation of the standard formula, especially for SII (see, e.g., Pfeifer and Strassburger, 2008; Mittnik, 2011; Christiansen et al., 2012; Gatzert and Martin, 2012); however, it focuses only on one part of the corporate risk structure (e.g., only on the market risk module, the life insurance module or the non-life insurance module). Therefore, our result contributes to the literature by offering an undertaking-specific risk model for a comprehensive risk structure with recently developed aggregation models and an insight into how significant a potential misestimation might be. The misestimation of the economic capital under the standard models can present insurers with several challenges. The standard formula requires insurers to hold more capital than necessary with an internal risk model, thus, lowering the potential return on equity. The overestimation of the economic capital can force insurers to restructure their business (asset allocation and liability management) to increase the amount of capital.

Based on the estimated models, we propose that regulators adjust their dependence assumptions to reduce the deviation between the estimates under the standard models and the estimates under our data-driven internal models. Despite its efficiency, our internal risk model might be too costly for small- and mid-sized companies. The regulator might support small- and mid-size companies who are interested in taking advantage of such models by reducing the implementation costs, e.g., by providing basic technical documents and tools. Thus, to increase regulatory efficiency, it might be useful to support companies in developing the know-how. This could help to reduce potential market distortions in the competition of solvency models, caused by significant differences in efficiency. A key question that remains for future research is to find the optimal balance between statistical sophistication and the ease of implementation. Our results, however, illustrate that the approach taken by regulators is too simplistic and might raise concerns on potential market distortions. Our proposal might also be applied to the currently discussed international capital standard (ICS), which is also limited to linear dependency between risks and pre-defined risk parameters, but allows an opportunity to develop an accredited internal risk model as well.

In this study, we consider empirical datasets for a non-life insurer. The literature to date has also been mostly limited to the risk aggregation for non-life insurance, perhaps because of the tractability of dependence modeling for non-life loss data using the risk model. However, investigating life and health insurance data with our internal risk model might be also a possible means of contributing to the literature. It might also be an interesting avenue for future research to integrate some emerging risks that are well known for non-linear dependence structures (e.g., cyber risk) into the model. Another avenue for future research could be to apply U.S. data to the internal model and then compare it with the U.S. RBC system. Moreover, some other risk modules in the standard models, such as operational risk and credit risk (counterparty default risk), might be included in the modeling, thereby constructing a more comprehensive and complete framework for risk aggregation.

Appendix A. Overview of literature and comparison with the present paper

Table A1. Summary of Literature on the Correlated Risk under Regulatory Frameworks

		PS08	**ET09**	**TV09**	**SC11**	**DEM12**	**Present paper**
Modeling perspective	Focus of Study	Aggregating non-normal risks	Integrating assets and liabilities	Aggregating underwriting risks	Aggregating underwriting risks	Aggregating underwriting risks	**Aggregating asset and underwriting risks**
	Comparison to Regulation	SII	-	Australian standard model (Prescribed method)	RBC ratios under SII	-	**K-RBC, SII, SST**
	Data type	Simulation with selected parameters	Simulation with selected parameters for asset and underwriting(bivariate for each)	Semi-annual loss ratios on Australian general insurance industry	Simulation with calibrated parameters	Monthly German underwriting data (non-life: multi-dimension)	**Monthly Korean and German asset and underwriting data** (multi-dimension for each)
	Level of Aggregation	Base-level (underwriting)	· Base-level (asset and underwriting) · Top-level	Base-level (underwriting)	Base-level (underwriting)	Base-level (underwriting)	· Base-level (asset and underwriting) · **Top-level**
	Use of Copula	Frechet-Hoeffding copulas	HAC and elliptical copulas	Elliptical copulas and Cauchy copula	HAC and elliptical copulas	Elliptical copulas and Bernstein copula	**PCC**, HAC, elliptical and **Bernstein (D-Vine)**
Outcome	Main points	· Identify the drawback of the square root formula (SII standard formula) in case of skewed underwriting risks (using beta distribution)	· Apply HAC to the dynamic financial analysis for non-life insurance with stylized parameters for asset and liability portfolio. · Investigate how the risk and return measures are affected by different dependence structures.	· Investigate how the economic capital for the underwriting portfolio is affected by different dependence structures. · Compare the estimated economic capitals with the Australian requirement and see how much they are deviated.	· The collective risk model for the underwriting risk is applied to SCR estimation using calibrated parameters. · HAC model is used to the underwriting risk aggregation.	· Bernstein copula fits well for the dependence modeling of underwriting risks, which is proven by goodness-of-fit analysis.	· Different types of dependence models including PCC and HAC are used to aggregate the two main risk modules (asset and underwriting). · Economic capitals from different copulas are compared with three regulatory standards.
	Limitation	There is no empirical study to check the robustness of the model and it is limited to underwriting risks.	There is no empirical study to check the robustness of the model and only a bivariate case is considered.	The data size is limited (only 19 observations for each risk factor) and it is limited to underwriting risks.	Calibrated parameters are not from SII standard and there is no empirical study to check the robustness of the model.	The fitting result might not be a global solution to all possible problems and the data size is relatively small.	Limited to the non-life insurance (no life insurance modeling)

Note: PS08: Pfeifer and Strassburger (2008); ET09: Eling and Toplek (2009); TV09: Tang and Valdez (2009); SC11: Savelli and Clemente (2011); DEM12: Diers, Eling and Marek (2012); DFA: dynamic financial analysis; SII: Solvency II; K-RBC: Korean RBC, PCC: Pair copula construction method (vine copula); HAC: Hierarchical Archimedean copula method. The bold indicates the contributing points of the present paper to the literature.

Appendix B. High-dimensional dependence models

We describe the mathematical definitions of vine models (PCC), particularly R-Vine, and hierarchical Archimedean copula (HAC), whose graphical differences are presented in Figure 1. Vine models can be categorized as D-Vine, C-Vine and R-Vine. The R-Vine links the variables by dependency without fixing a certain structure, so it can project the other two vine models (Cooke, Joe and Aas, 2011). We first factorize the joint density, $f(x_1, \ldots, x_d)$, of a vector of random variables, $X = (X_1, \ldots X_d)$, to form the copula distribution function as follows:

$$f(x_1, \ldots, x_d) = f_1(x_1) \cdot f_{2|1}(x_2|\, x_1) \cdots f_{d|1,\ldots,d-1}(x_d|\, x_1, \ldots, x_{d-1}). \tag{B.1}$$

Sklar's theorem (Sklar, 1959) and the chain rule with continuous marginal function, $F_i(x_i), i = 1, \ldots d$, lead us to obtain

$$f(x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d) \cdot c_{1,\ldots,d}[F_1(x_1), \ldots, F_d(x_d)], \tag{B.2}$$

where $c_{1,\ldots,d}[\cdot]$ is a d-dimensional copula density.

Equation (B.1) is equivalent to equation (B.2) by decomposing the conditional density in equation (B.1). Let us consider an example of a bivariate conditional density, $f_{2|1}(x_2|\, x_1)$, which can be defined as:

$$f_{2|1}(x_2|\, x_1) = \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} = \frac{c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_1(x_1) \cdot f_2(x_2)}{f_1(x_1)} = c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_2(x_2), \tag{B.3}$$

where $c_{1,2}[\cdot,\cdot]$ is a copula density function for the pair of X_1 and X_2, each of which has a density function, $f_i(x_i)$, $i = 1{,}2$, and a probability function, $F_i(x_i)$.

A three-dimensional case with two variables given to the condition in the density can be defined as:

$$f_{1|23}(x_1|\, x_2, x_3) = c_{1,2|3}\big[F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)\big] \cdot f_{1|3}(x_1|x_3). \tag{B.4}$$

One can define different decomposition of the conditional density from equation (B.4) according to the following dependence structure of variables:

$$f_{1|23}(x_1|\, x_2, x_3) = c_{1,3|2}\big[F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\big] \cdot f_{1|2}(x_1|x_2). \tag{B.5}$$

Equation (B.5) can be further factorized to the following:

$$f_{1|23}(x_1|\, x_2, x_3) = c_{1,3|2}\big[F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\big] \cdot c_{1,2}[F_1(x_1), F_2(x_2)] \cdot f_1(x_1). \tag{B.6}$$

We can generalize equation (B.1) by constituting the pairwise copula construction and a conditional marginal density using the development of the factorization above in the following (Aas et al., 2009):

$$f(x|\,\mathbf{\Theta}) = c_{x,\Theta_j|\mathbf{\Theta}_{-j}}\left[F\left(x|\mathbf{\Theta}_{-j}\right), F\left(\Theta_j|\mathbf{\Theta}_{-j}\right)\right] \cdot f\left(x|\mathbf{\Theta}_{-j}\right), \tag{B.7}$$

where $\mathbf{\Theta}$ is a d-dimensional vector, Θ_j is an arbitrarily selected component of the vector $\mathbf{\Theta}$ and $\mathbf{\Theta}_{-j}$ is a vector of $\mathbf{\Theta}$ without the j-th component.

Vine models are constructed by a number of trees, T_j, $j = 1, \ldots, d-1$, which start from unconditional marginal densities and develop with decomposition. Each tree, T_j, is comprised of $d-j$ nodes and $d-j-1$ edges and the entire decomposition of the vine density is defined by $n(n-1)/2$ edges and the marginal densities of d variables. The labels of the edges in the tree T_{j+1} are defined by the nodes in the tree T_j. The case that two edges in the tree T_j have a common node forms an edge in the tree T_{j+1}.

As mentioned above, the R-Vine flexibly connects the variables by dependency, offering a general form of a vine model. Definition B1 illustrates a regular vine model based on Bedford and Cooke (2001, 2002), Cooke, Joe and Aas (2011) and Dissmann et al. (2013). For the D-Vine and the C-Vine, Aas et al. (2009) provide more detail in specific algorithms.

Definition B1. (A regular vine) $\mathbf{X} = \{x_1, \ldots, x_d\}$ is a d-dimensional set. $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$ is a nested set of trees in a regular vine structure on d components if

(1) T_1 is the first tree with nodes $D_1 = \{1, \ldots, d\}$ and a set of edges, E_1.
(2) For $i = 2, \ldots, d-1$, T_i is a following tree with nodes $D_i = E_{i-1}$ and a set of edges, E_i.
(3) (Proximity condition) for $i = 2, \ldots, d-1$ and $\{a, b\} \in E_i$, $\#(a \cap b) = 1$ holds, where # indicates the cardinality of a set.

The HAC method uses Archimedean copulas, which incorporate dependence parameters estimated by a generating function. Prior to constructing a generalized HAC structure, the following equation defines an exchangeable structure with Archimedean copulas in a d-dimensional setting (Aas and Berg, 2009):

$$C(u_1, \ldots, u_d) = \psi^{-1}[\psi(u_1) + \cdots + \psi(u_d)], \tag{B.8}$$

where $u_i, i = 1, \ldots, d$, is a uniform margin for random variable i, C is a copula function and ψ is a generating function of an Archimedean copula with a monotonic decreasing function $\psi^{-1} \in [0,1]$.

To construct a HAC structure, let us denote hierarchical levels by h, each of which incorporates n_h distinct objects. Suppose that $\mathbf{X} = \{x_1, \ldots, x_d\}$ is a d-dimensional set with a set of uniform margins $\mathbf{U^d} = \{u_1, \dots, u_d\} \in [0,1]^d$ at the ground level $h = 0$. At the first level $h = 1$, the uniform margins are modeled by n_1 Archimedean copulas with the form, $C_{1,j}, j = 1, \dots, n_1$, defined as:

$$C_{1,j}(\mathbf{u}_{1,j}) = \psi_{1,j}^{-1}\left(\sum_{\mathbf{u}_{1,j}} \psi_{1,j}(\mathbf{u}_{1,j})\right), \tag{B.9}$$

where $\psi_{1,j}$ is the generating function of the copula $C_{1,j}, j = 1, \dots, n_1$, and $\mathbf{u}_{1,j}$ is a set of elements from uniform margins $\mathbf{U^d}$.

A set of copulas, $C_{1,j}$, at the first level is grouped into a set of copulas at the second level, $C_{2,j}, j = 1, \dots, n_2$, which can be represented as:

$$C_{2,j}(\mathbb{C}_{2,j}) = \psi_{2,j}^{-1}\left(\sum_{\mathbb{C}_{2,j}} \psi_{2,j}(\mathbb{C}_{2,j})\right), \tag{B.10}$$

where $\psi_{2,j}$ is the generating function of the copula $C_{2,j}, j = 1, \dots, n_2$, at the second level and $\mathbb{C}_{2,j}$ stands for a set of the copulas from the first level ($h = 1$) grouped into $C_{2,j}$.

The entire HAC structure is complete until the process reaches the final level h with the hierarchical Archimedean copula $C_{h,1}$ as a single object. As an example, the four-dimensional density function of HAC can be expressed as:

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{21}\big(C_{11}(u_1, u_2), C_{12}(u_1, u_2)\big) \\ &= \psi_{21}^{-1}\big[\psi_{21}\{\psi_{11}^{-1}\big(\psi_{11}(u_1) + \psi_{11}(u_2)\big)\} + \psi_{21}\{\psi_{12}^{-1}\big(\psi_{12}(u_3) + \psi_{12}(u_4)\big)\}\big], \end{aligned} \tag{B.11}$$

where $u_i, i = 1, \dots, 4$, is a uniform margin for random variable i, C is a copula function and ψ is a generating function of an Archimedean copula.

Aas and Berg (2009), Savu and Trede (2010) and Okhrin, Okhrin and Schmid (2013) provide more detail on the specification and the inference of the HAC structure.

Appendix C. Graphical and numerical results for modeling

Graphical diagnoses and numerical results of marginal modeling

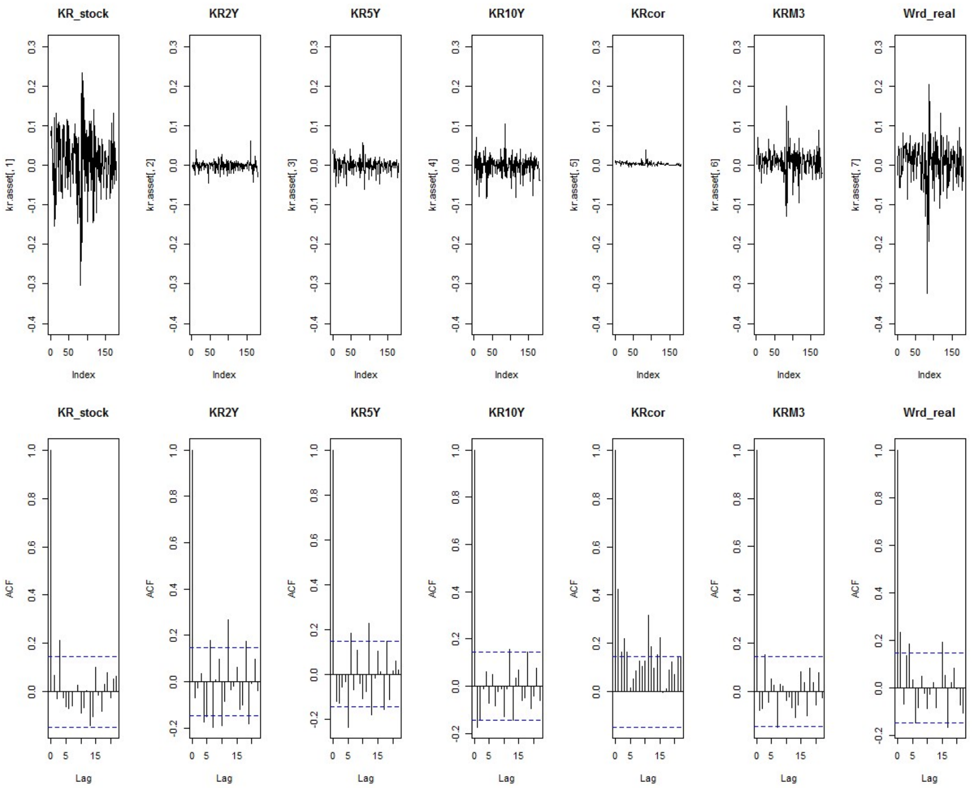


Figure C1. Time-series plots and autocorrelation function plots.

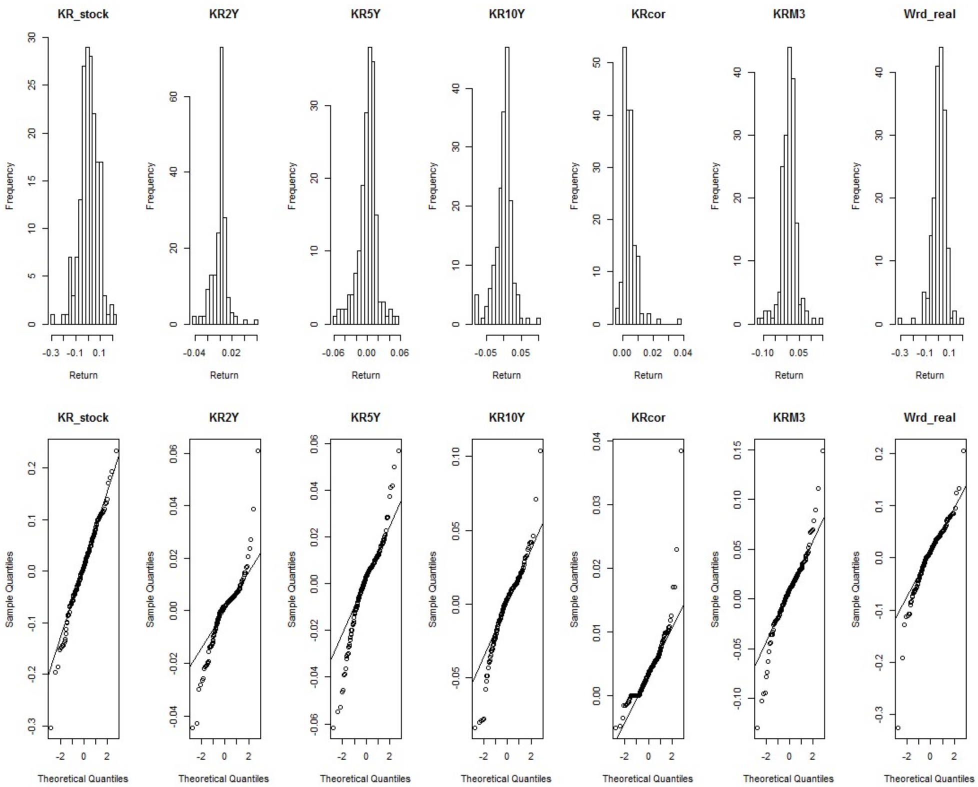


Figure C2. Histograms of asset returns and testing normality of innovations.

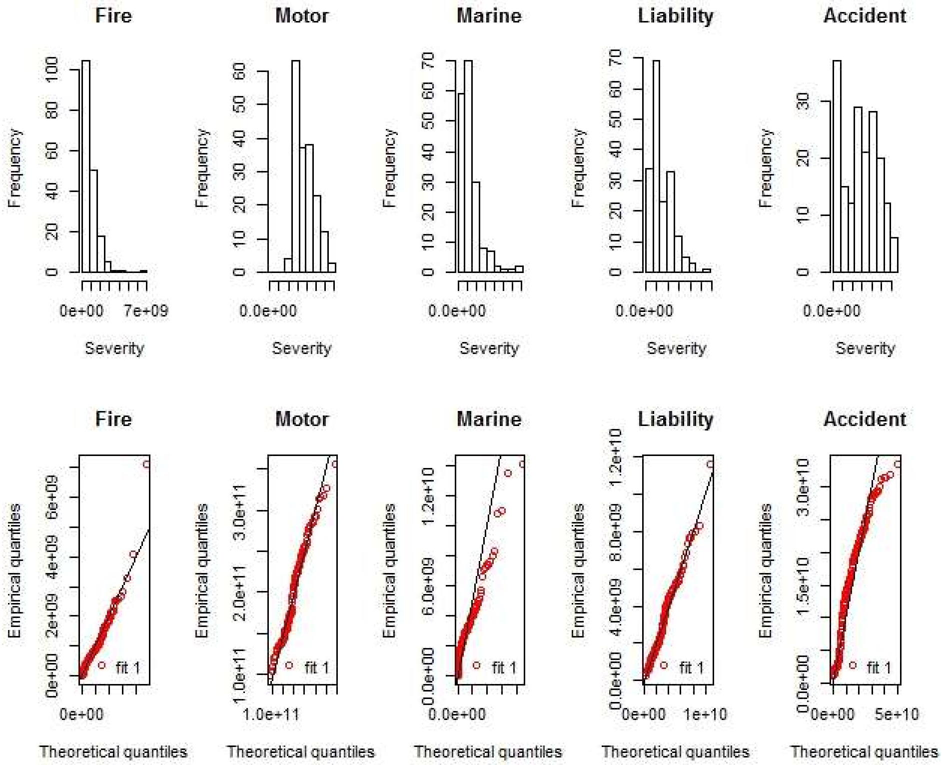


Figure C3. Histograms of insurance loss and QQ-plots with fitted distributions.

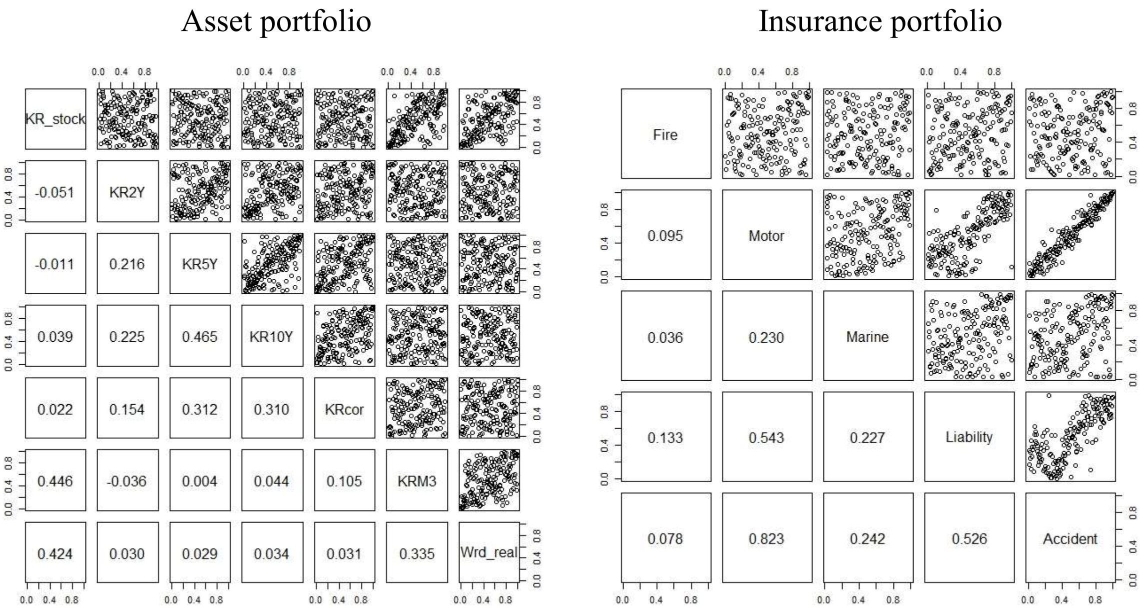


Figure C4. Pairwise plots and Kendall's correlations of marginal distributions. The pairwise scatter plots of transformed standardized residuals are displayed on the upper triangle and the corresponding bivariate Kendall's rank correlations are on the lower triangle.

Table C1. ARMA-GARCH Specifications

	ARMA-order	AIC for ARMA	Fitted distribution for innovation	AIC for GARCH(1,1)
KR_stock	(0,3)	-416.37	Skew normal	-439.28
KR2Y	(3,3)	-1,119.34	Student-t	-1,171.82
KR5Y	(3,2)	-993.47	Student-t	-1,000.89
KR10Y	(3,1)	-835.83	Skew student	-841.92
KRcor	(3,3)	-1,467.52	Student-t	-1,615.04
KRM3	(3,3)	-710.25	Skew normal	-785.15
Wrd_real	(2,3)	-540.72	Skew normal	-591.04

Table C2. Distributions Fitting Results for Insurance Losses

	Fire	Motor	Marine	Liability	Accident
Skew Normal	5,507.61 (0.426***)	18,411.32 (0.999***)	5,894.99 (0.426***)	5,884.48 (0.474***)	6,452.71 (0.458***)
Student-t	6,056.20 (0.560***)	7,550.89 (0.743***)	6,381.02 (0.539***)	5,756.39 (0.319***)	6,358.80 (0.424***)
Skew student-t	**5,277.97** **(0.071)**	7,144.08 (0.434***)	5,675.24 (0.108**)	5,685.14 (0.118**)	**6,254.18** **(0.072)**
Lognormal	5,595.35 (0.334***)	**6,898.17** **(0.073)**	5,958.91 (0.363***)	5,664.16 (0.072)	6,300.37 (0.199***)
Gamma	5,291.48 (0.213***)	6,902.25 (0.109**)	**5,637.89** **(0.075)**	**5,659.83** **(0.077)**	6,265.40 (0.161***)
Weibull	6,188.74 (0.823***)	8,497.73 (0.886***)	6,466.01 (0.803***)	6,819.57 (0.868***)	7,412.18 (0.884***)
Inverse Gaussian	8,725.86 (0.978***)	12,481.34 (0.999***)	9,200.98 (0.961***)	9,726.62 (0.999***)	10,699.46 (0.999***)
Cauchy	5,323.55 (0.125***)	7,000.54 (0.163***)	5,741.18 (0.138***)	5,761.80 (0.160***)	6,366.56 (0.133***)
Burr	5,988.23 (0.546***)	8,276.54 (0.620***)	6,273.73 (0.525***)	6,599.54 (0.574***)	7,185.69 (0.579***)
GPD	5,335.66 (0.201***)	7,418.68 (0.683***)	5,727.55 (0.191***)	5,768.63 (0.208***)	6,349.77 (0.221***)
POT 90% (Norm-GPD)	5,472.58 (0.437***)	7,564.20 (0.818***)	5,869.92 (0.426***)	5,896.56 (0.463***)	6,476.03 (0.453***)
POT 90% (Lognorm-GPD)	5,597.43 (0.303***)	8,341.07 (0.801***)	6,312.60 (0.697***)	6,661.56 (0.765***)	7,252.92 (0.757***)

Note: The numbers in the parentheses are Kolmogorov-Smirnov statistics and *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit for each insurance loss distribution.

Parameter estimation and graphical structure of dependence modeling

Table C3. Parameter Estimations of the Best Fitted Pair Copula Structure (First Tree)

Copula	Parameter	Lower-tail dependency	Upper-tail dependency
Asset Portfolio			
$\theta_{4,2}$; $C_{KR10Y,KR2Y}$ (Student-t)	0.36/3.78	0.195	0.195
$\theta_{3,4}$; $C_{KR5Y,KR10Y}$ (Student-t)	0.67/2.80	0.439	0.439
$\theta_{5,3}$; $C_{KRcor,KR5Y}$ (Student-t)	0.48/3.43	0.272	0.272
$\theta_{6,5}$; $C_{KRM3,KRcor}$ (Student-t)	0.17/3.72	0.129	0.129
$\theta_{1,6}$; $C_{KR\_stock,KRM3}$ (Student-t)	0.65/4.16	0.339	0.339
$\theta_{7,1}$; C_{Wrd_real,KR_stock} (Survival Gumbel)	1.82	0.537	-
Insurance Portfolio			
$\theta_{4,1}$; $C_{Liability,Fire}$ (Survival Gumbel)	1.16	0.181	-
$\theta_{2,4}$; $C_{Motor,Liability}$ (Gumbel)	1.93	-	0.568
$\theta_{5,2}$; $C_{Accident,Motor}$ (Frank)	20.34	-	-
$\theta_{5,3}$; $C_{Accident,Marine}$ (Survival Clayton)	0.86	-	0.341

Note: The table shows estimated parameters only in the first tree of the structure, which demonstrates the strongest dependency in the structure. Student-t copula has two parameters representing dependency and degree of freedom. The parametric R-Vine is sequentially determined by R-package `VineCopula`. The subscripts of the copulas indicate the following:
Asset (Korean): 1=KR_stock; 2=KR2Y; 3=KR5Y; 4=KR10Y; 5=KRcor; 6=KRM3; 7=Wrd_real
Insurance (Korean): 1=Fire; 2=Motor; 3=Marine; 4=Liability; 5=Accident

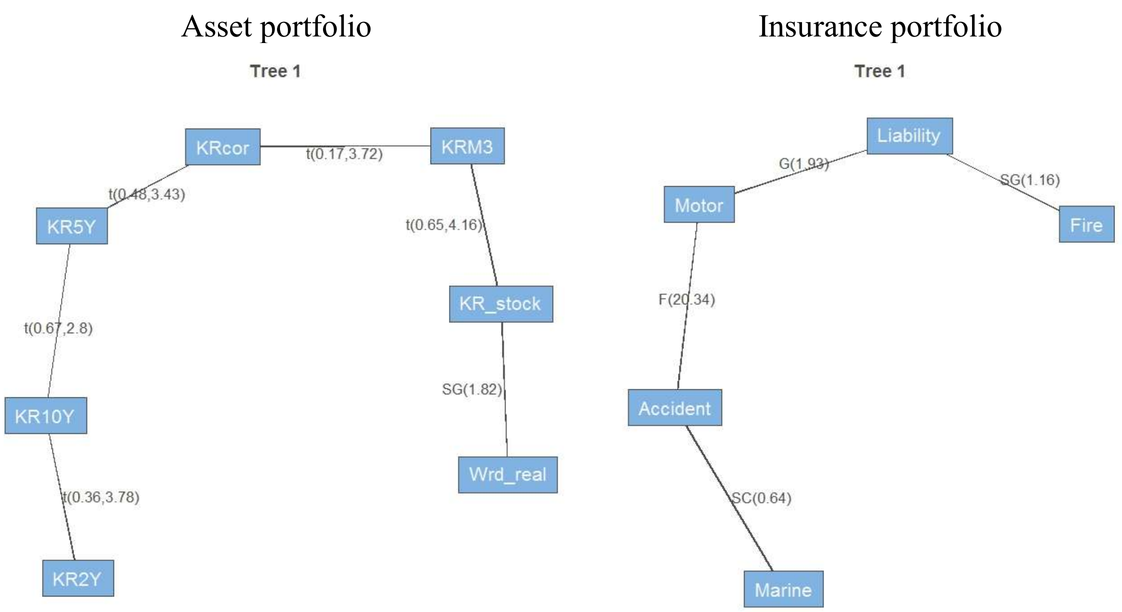


Figure C5. Graphical structure of dependency in the first tree. The estimated copula functions and their dependence parameters are illustrated at edges. The copula functions at edges are specified as follows: t: Student-t; SG: Survival Gumbel; G: Gumbel; SC: Survival Clayton; F: Frank.

Appendix D. Additional applications

In this application, we diversify the dependence structures at the base level to identify the impact of different base-level dependencies and also the structures at the top level with the data-driven correlation between asset and underwriting portfolios. Additionally, we investigate how the difference between internal models and standard models changes with different asset allocation strategies and how strong the diversification effect in our internal model could be.

Application to economic capital

Table D1 illustrates the results of risk measurement with diversified models at two levels. Regarding the models at the base level, we carry out two different approaches in order to apply HAC models and simple Archimedean models (AC); we consider HAC and AC with Gumbel for the asset portfolio and with Clayton for the insurance portfolio, which is regarded as the best-case scenario for a non-life insurer by HAC and AC models.[152] Conversely, HAC and AC with Clayton for the asset portfolio and with Gumbel for the insurance portfolio can be seen as the worst-case scenario by HAC and AC. Additionally, we employ the independence structure for both portfolios at the base level to see how significantly different economic capitals can be between independence assumption and intrinsic dependence structures at the base level aggregation.

The correlation applied at the top level is the empirical measure from the original data. Consistently, the independence structure at the top level generates the lowest economic capital across all base-level models, implying that the correlation assumption under regulations between market and insurance modules can overestimate the capital requirements. The models with the tail dependence measures (particularly, student-t and Gumbel) produce higher level of capitals at extreme quantiles. This also applies to the base-level results that R-Vine, student-t, HAC-worst and AC-worst cases with the correlated risk in the tail generally provide higher capital levels. In addition, the independence structure of *d*-dimensional case at the base level

[152] Since HAC with Gumbel copula can model the upper tail dependency for every estimated pair in the structure, it can derive the simultaneous positive returns from the asset portfolio. In contrast, since HAC with Clayton copula can model the lower tail dependency for every estimated pair, it can predict simultaneous lower claims from different lines of business in the insurance portfolio. However, these scenarios are not definitive best or worst scenarios across all considered models since the scenarios are modeled only by HAC, but they can provide different scenarios for estimating the economic capital. In the literature, several approaches are carried out to find the bounds on the risk measure. For example, Pfeifer and Strassburger (2008) use Frechet-Hoeffding bounds to aggregate stochastically dependent risks as extreme cases. Aas and Puccetti (2014) develop upper and lower bounds on the risk measure based on the rearrangement algorithm (RA).

can generate significantly lower level of the economic capital, since this structure does not incorporate the correlated risk that should be counted in the estimate.

Table D1. Comparison with Regulatory Standards

(in ₩ billion)			Top-level dependence (Bivariate)				
			Indep	Gauss	t	Gumbel	Clayton
Base-level dependence	VaR at 99%	R-Vine	**559.7**	582.1	587.9	605.6	577.4
		Gauss	521.1	554.7	571.5	569.1	548.5
		t	534.7	557.9	581.1	575.5	555.2
		HAC-best	464.3	488.7	516.7	504.5	482.3
		HAC-worst	507.8	528.9	563.8	535.1	524.8
		AC-best	398.3	419.5	438.5	430.2	413.5
		AC-worst	531.0	553.1	578.4	553.5	541.7
		Independence	129.3	148.5	165.6	165.4	146.9
		Bernstein	304.9	323.8	346.3	338.9	318.6
	VaR at 99.5%	R-Vine	**744.5**	757.3	768.6	763.4	755.0
		Gauss	693.3	711.0	697.5	682.0	680.6
		t	714.6	723.6	721.0	682.9	708.5
		HAC-best	619.0	629.8	625.7	614.5	625.4
		HAC-worst	672.4	698.8	692.7	669.6	693.2
		AC-best	533.9	547.4	546.8	529.2	543.4
		AC-worst	711.0	721.3	716.8	708.2	715.4
		Independence	243.5	265.7	252.9	248.9	253.2
		Bernstein	454.7	478.6	469.3	441.4	462.7
	TVaR at 99%	R-Vine	**842.5**	850.0	880.2	866.4	841.0
		Gauss	779.2	762.5	795.5	782.7	757.8
		t	795.0	780.6	811.8	790.5	777.2
		HAC-best	701.5	686.3	713.2	699.3	678.6
		HAC-worst	769.2	757.1	788.1	776.3	754.5
		AC-best	615.4	600.8	624.9	613.2	591.5
		AC-worst	801.5	805.0	829.8	819.1	794.8
		Independence	316.1	311.3	333.6	323.4	307.5
		Bernstein	540.7	527.3	551.8	534.5	520.6

	K-RBC (VaR 99%)	SII (VaR 99.5%)	SST (TVaR 99%)
Standard models	846.3	1,177.5	1,471.5

Note: HAC stands for hierarchical Archimedean copulas and AC indicates simple Archimedean copulas. The bold indicates the best fitted model for the empirical data.

The estimates of the Bernstein model are positioned between those of parametric models (R-Vine, Gaussian and student-t) and the independence model, which is in line with the outcome of the literature modeling non-life insurance loss with Bernstein copula (Diers et al., 2012).[153] Specifically, the Bernstein copula could generate a smaller size of the risk measures at extreme quantiles than those of parametric copulas. This result might be because the non-parametric estimation with Bernstein polynomials plays a role of an approximation to parametric copulas

[153] Few empirical studies have compared risk measurement between Bernstein copula and parametric copulas. Thus, our result in the application could be a benchmark to compare Bernstein model and other parametric models for future research.

and smoothing parameters for polynomial dimensions are dependent on the grid size (= m) (Sancetta and Satchell, 2004; Scheffer and Weiss, 2017).

Asset allocation strategies

The asset allocation considered in Table 8 reflects the realistic strategy that the main insurer adopted, giving 2.75% annual return and 7.51% annual standard deviation. In this case, slightly above 80% is invested in the fixed income securities, which can help an insurer meet the regulatory requirement (ruin probability of 0.5%) and be more likely to avoid insolvency (Eling et al., 2009). However, it is difficult for the insurer from the short-term perspective to increase the investment size so that the firm usually diversifies strategies by considering different allocations to the assets (Eling et al., 2009).[154] Based on this fact, we implement different asset allocations to derive the economic capital under our internal model and the standard model. Following Eling et al. (2009), we take into account two possible asset allocation strategies shown in panel A of Table D2[155], which have different levels of risk. Since the second case incorporates more weights on riskier assets (i.e., KR_stock and Wrd_real. See Table D2), it shows higher volatility (= 17.72%) on an annual basis, but higher return (= 6.17%) than that of the first case (2.82% volatility and 8.50% return). We estimate the economic capitals by VaR at 99% and the K-RBC level for each strategy to which the dataset is subject.

The results in panel B of Table D2 show that a riskier asset portfolio increases the economic capital and the capital requirement by the regulation. The economic capital of the riskier case under the best fit model is 38.7% higher than that of the less risky case, which implies that 1% increase in the portfolio volatility can lead to 4.21% increase in the economic capital. This is a reasonable outcome in that a riskier investment with a higher return and a higher volatility leads to a higher capital requirement. The economic capital with different asset strategies can be also overestimated by the standard model, showing that the K-RBC turns out to be 56.2% overestimated in the first strategy and 60.9% in the second strategy. It is observed from this result that the riskier the asset portfolio is, the larger the gap between the internal model and the standard model is.

---

[154] Gründl et al. (2016) state that firm size is a significant factor on risk-taking by insurers. Larger insurers are able to more diversify investment portfolios, which enables the insurers to take on more risks and control such risks.

[155] Eling et al. (2009) consider four examples of asset allocations. Among them, the first example accounts for the investment only on the money market and the third example addresses the equally weighted portfolio (see Table 2 in Eling et al., 2009). Since we need to take into consideration diversified portfolio under dependency with different riskiness we do not include the first and third examples in Eling et al. (2009).

Table D2. Diversified Investment Strategies

Panel A: Asset allocations

		Allocation 1	Allocation 2
Equity	KR_stock	10%	30%
Fixed income	KR2Y	20%	10%
	KR5Y	20%	10%
	KR10Y	20%	10%
	KRcor	20%	10%
Money Market	KRM3	0%	0%
Real estate	Wrd_real	10%	30%

Panel B: Application to the economic capital

			Top-level dependence					
(in ₩ trillion)			Indep	Gauss	t	Gumbel	Clayton	K-RBC
Base-level dependence	Allocation 1 (VaR 99%) · μ=2.82% · σ=8.50%	R-Vine	**603.4**	631.8	663.0	641.0	621.1	942.5
		Gauss	565.1	590.1	619.4	614.1	583.7	
		t	578.3	602.3	628.7	621.1	592.0	
		HAC-best	511.5	532.8	566.7	551.0	517.1	
		HAC-worst	552.0	570.9	608.1	585.8	566.5	
		AC-best	458.9	483.1	503.8	494.2	466.0	
		AC-worst	590.9	598.5	643.0	614.1	597.7	
		Independence	202.9	214.2	241.5	238.3	206.6	
		Bernstein	421.3	439.3	452.5	451.5	431.5	
			Indep	Gauss	t	Gumbel	Clayton	K-RBC
	Allocation 2 (VaR 99%) · μ=6.17% · σ=17.72%	R-Vine	**837.1**	871.1	924.1	899.8	848.8	1,346.7
		Gauss	787.7	795.2	841.9	832.1	793.3	
		t	790.2	817.6	851.6	844.5	797.5	
		HAC-best	659.5	682.5	717.3	705.4	672.0	
		HAC-worst	800.1	844.0	893.9	863.8	821.9	
		AC-best	654.7	678.2	709.2	702.5	669.9	
		AC-worst	852.5	876.9	928.7	892.4	854.0	
		Independence	392.8	405.4	433.0	431.0	399.3	
		Bernstein	650.8	669.9	694.0	688.4	668.9	

Note: μ is the mean of the portfolio and σ is the volatility of the portfolio on an annual basis. The bold in panel B indicates the best fitted model for the empirical data.

Diversification effect

The standard formula under regulations allows insurers to have the diversification benefit by reducing the economic capital in the aggregate distribution. The diversification effect in the standard formula is generated by the square root formula with the assumed correlation measures. However, it can be expected that the diversification effect under the linear dependence structure is smaller than the effect under the non-linear dependence structure (Eling and Jung, 2018). The diversification effect is estimated by:

$$div = \frac{SCR\left(\sum_{i=1}^{d} X_i\right) - \sum_{i=1}^{d} SCR_i}{\sum_{i=1}^{d} SCR_i}, \quad (D.1)$$

where $SCR\left(\sum_{i=1}^{d} X_i\right)$ is the solvency capital requirement of a portfolio with d-dimensional risks and $\sum_{i=1}^{d} SCR_i$ is the sum of individual SCRs for the portfolio.

Gatzert and Martin (2012) assess the diversification benefit of around 18% from the asset portfolio with stocks and bonds in their internal model and provide the benefit of the standard formula in SII (see Figure D1). To compare with their estimate as a benchmark, we also derive the diversification effect of the asset portfolio from different base-level settings. Here, we consider five competing base-level models: R-Vine (PCC), Gaussian, student-t, HAC with Gumbel, HAC with Clayton and Bernstein copula. The diversification benefits from our empirical case and Gatzert and Martin (2012) are comparable in that the same diversification formula is applied and only stocks and bonds are considered in the portfolio. The difference between two implementations is that the asset portfolio in our empirical case consists of less number of assets than the portfolio in Gatzert and Martin (2012) (5 vs. 14). Overall, the pair copula model drives the highest diversification benefit of around 42.6%, followed by student-t (37.9%), Gaussian (37.2%) and Bernstein (37.0%). This effect accounts for a significant reduction in the economic capital (from the simple summation of individual risks estimated in the same model) by the data-driven dependence structure, compared to that of the standard formula (7.7%) in Figure D1.

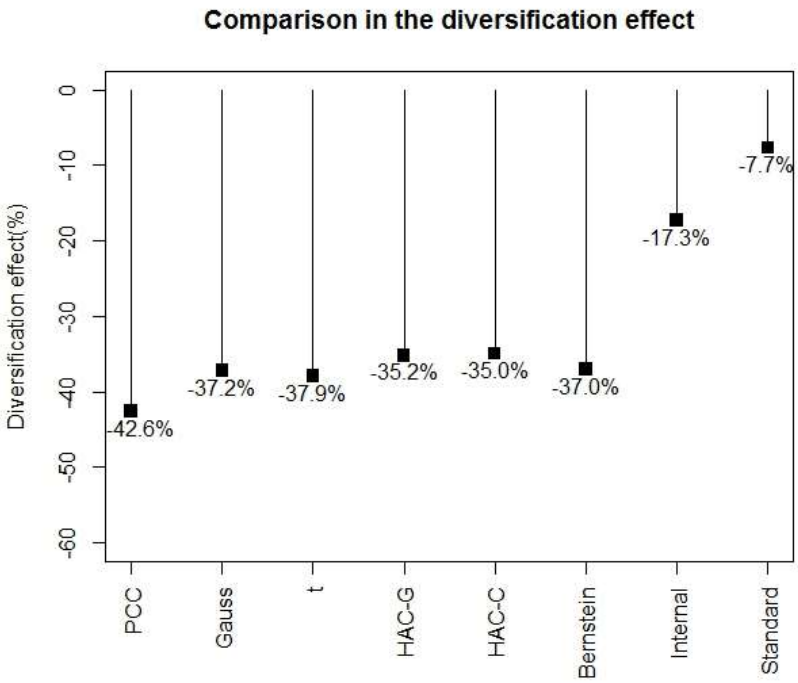


Figure D1. Diversification benefit of the asset portfolio. The diversification benefit is obtained by measuring how much the SCR of a portfolio is reduced from the sum of individual SCRs. “HAC-G” stands for HAC model with Gumbel copula and “HAC-C” for HAC model with Clayton copula. “Internal” indicates the diversification benefit from the internal model in Gatzert and Martin (2012) and “Standard” from the standard model in the same paper.

Appendix E. Comparing standard models and the internal model

In this section, we additionally note the comparison between the internal model and the standard models in Section 5. First, we apply the risk category of each regulatory framework to the dataset as explained in Section 2.1. For example, we categorize our underwriting data to two lines of business (automobile and general insurance) under the K-RBC and to five lines of business (Fire, Motor, Marine, Liability and Miscellaneous) under SII. This categorization is also applied to the German data in Appendix G. The following describes several particular points of comparison.

Korean RBC

As mentioned in Section 2.1, no correlation between market risk factors is taken into consideration in the standard model. Here, we assume that the equity and bond assets in the Korean benchmark portfolio belong to the category of short-term securities (see footnote 124). Thus, the market risk module in the K-RBC framework is based on the simple summation of risk amounts from different factors (FSS, 2017, p. 131):

$$Risk_{mkt} = \sum_{i} Risk_i, \tag{E.1}$$

where $Risk_i$ is the exposure of i-th risk factor multiplied by the corresponding risk coefficient.

The standard risk coefficient for each factor from both asset and underwriting portfolios is given to an insurer, however, the K-RBC allows for adjusting the coefficient only to the underwriting portfolio. The adjusted risk coefficient (= γ_i') for the underwriting risk is calculated with the difference in the combined ratio between the undertaking and the industry average (FSS, 2017, p. 54):

$$\gamma_i' = \min\{\mathrm{Max}[\gamma_i + (CR_i - CR_{avg}) \times 0.5, \gamma_i \times 0.7], 1\}, \tag{E.2}$$

where γ_i is the standard risk coefficient for i-th line of business, CR_i is the combined ratio for i-th line and CR_{avg} is the industry average of the combined ratio for i-th line.

The risk size for each factor is conceptually described by VaR 99% as seen in Figure E1, which is reflected in the risk coefficient. To replace the risk coefficient given in the standard model by our undertaking-specific parameters, we apply the multiplication of 99% quantile under the normality assumption and the standard deviation (= 2.32σ) for the asset portfolio in the internal model, which is equivalent to the size of the deviation in Figure E1. Our undertaking-specific risk coefficients can better reflect the degree of riskiness, for example, from the returns in the realistic asset portfolio (see Table E1).

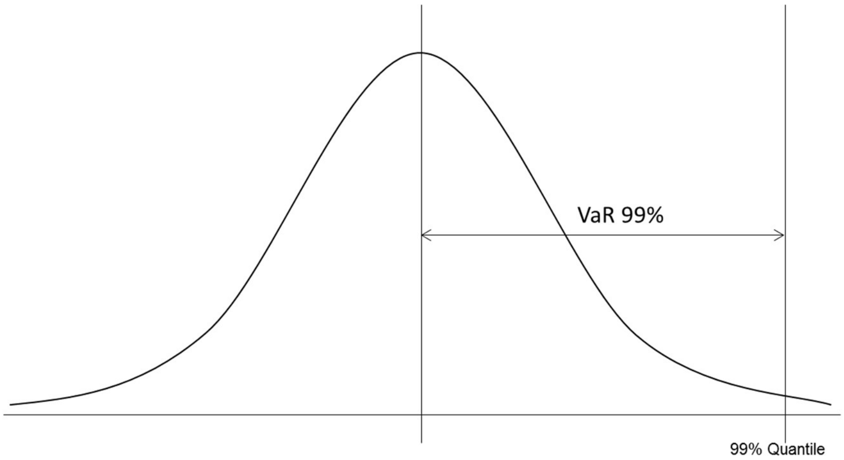


Figure E1. Conceptual plot for individual risk exposure in the K-RBC (FSS, 2017, p. 29).

Table E1. Comparison of Risk Coefficients for the Asset Portfolio

	KRstock	KR2Y	KR5Y	KR10Y	KRcor	KR3MCD	Wrd_real
Internal model	0.179	0.028	0.040	0.059	0.124	0.119	0.131
K-RBC	0.120	0.012	0.035	0.060	0.030	0.012	0.060

Note: For the risk coefficients on the government bonds, we assume that the term to maturity for each bond is following: 1-2 years for KR2Y, 4-5 years for KR5Y and 7-10 years for KR10Y.

Solvency II

SII differentiates the correlation assumptions of the market risk module into two possible scenarios: "up" and "down" shocks of interest rate. The SCR calculation considering interest rate shocks is following (EC, 2010, p. 108):

$$SCR_{mkt} = \max\left[\sum_{i,j} CorrMktUp_{i,j} \cdot Mkt_{up,i} \cdot Mkt_{up,j}\,; \sum_{i,j} CorrMktDown_{i,j} \cdot Mkt_{down,i} \cdot Mkt_{down,j}\right], \quad \text{(E.3)}$$

where $CorrMktUp_{i,j}$ is the correlation between market risk factor i and j under the interest rate up stress, $Mkt_{up,i}$ is the required capital for the i-th market risk sub-module under the interest rate up stress, $CorrMktDown_{i,j}$ is the correlation between market risk factor i and j under the interest rate down stress and $Mkt_{down,i}$ is the required capital for the i-th market risk sub-module under the interest rate down stress.

The SII estimate in Table 8 is the one under the interest rate up stress. In terms of the equity risk module, SII differentiates the shock scenario in two categories: "Global" and "Other" ("Type 1" and "Type 2" in EC, 2014). "Global" category contains the equities belonging to the members of the European Economic Area (EEA) or the Organization for Economic

Cooperation and Development (OECD), whereas "Other" category incorporates other equities not belonging to the list of "Global" equities (EC, 2014, p. 108). The equities in our Korean and German cases are considered in the "Global" category.

The required capital for each risk module is calculated by applying the calibrated parameters by EC (2014). With regard to the required capital for the non-life risk module, the calculation is based on equation (3). For each sub-risk module, we calculate the required capital:

$$SCR_{i,NL} = 3 \times \sigma_i \times V_i, \tag{E.4}$$

where σ_i is the standard deviation for i-th submodule based on the calibrated standard deviation for premium and reserve risks in SII and V_i is the volume measure for i-th submodule. $3 \times \sigma_i$ indicates an approximate measure of the combined standard deviation for the module, which is assumed upon the lognormal losses (EC, 2010, p. 199).

Swiss Solvency Test

SST is designed to evaluate the amount of risks on a market-consistent basis (FINMA, 2006). The SST measure in Table 8 is upon the standard model, which assumes the linear dependence of P&L on risk factors and multivariate normal for all risk factors. That is, SST assumes a multivariate normal distribution with mean 0 and the volatility based on the sensitivity analysis for the asset portfolio (FINMA, 2006). It calibrates the volatilities and the correlation matrix with monthly data. Thus, we also use variance-covariance approach for the asset portfolio with the delta measure defined in SST as follows:

$$\Delta V(P) \sim N\left(0, \sqrt{\delta' \textstyle\sum \delta}\right), \tag{E.5}$$

where $\Delta V(P)$ is the change in the market portfolio value, δ is a vector of the sensitivities of the portfolio with respect to the risk factors and $\sum$ is the covariance matrix of the risk factors (7×7 for the Korean case and 10×10 for the German case in Appendix G).

Regarding the underwriting portfolio, SST requires different estimations for normal and large losses, which are above either the threshold of 1 million or 5 million CHF (see Section 2.1). For normal sizes, the first two moments are used to estimate, whereas lognormal assumption is employed to estimate large sizes. In this application, we assume that our loss data are the losses in the current accident year and the exchange rate between CHF and Korean Won is 1,000 Won/CHF to translate the threshold. Then, we apply the distributional assumptions to the different sizes of losses and implement convolution to aggregate the market risk and the underwriting risk.

Appendix F. Robustness of the best fit model

We implement our modeling procedure on the base-level dependence structure of the underwriting portfolio and the top-level dependence structure from both sides for four companies considered in Table 9. Panel A of Table F1 illustrates that the statistical testing results for four companies at the base level turn out to be similar to the result with the main dataset in Section 4. R-Vine model is proven to be superior to other models, followed by elliptical models (Gaussian and student-t). Hierarchical Archimedean models fit better than simple Archimedean family, but cannot compete with R-Vine and elliptical family. At the top level, we again find that the independence assumption needs to be taken into account as evidenced by statistical testing in panel B; thus, we can conclude that our internal risk model is robust.

Table F1. Implementation of the Best Fit Aggregation Model for Four Companies

Panel A: Base-level estimation (underwriting)

		Company 1		Company 2		Company 3		Company 4	
Copula		LogLik	AIC	LogLik	AIC	LogLik	AIC	LogLik	AIC
Elliptical	Gauss	154.4	-288.9	118.1	-216.1	52.5	-85.0	118.6	-217.3
	t	161.7	-301.4	120.6	-219.2	53.6	-85.2	133.5	-245.0
Archimedean	Gumbel	26.5	-50.9	80.9	-159.7	-0.0	2.0	60.6	-119.2
	Clayton	26.7	-51.4	39.3	-76.5	1.8	-1.7	86.2	-170.3
HAC	Gumbel	142.4	-276.9	128.0	-248.0	14.0	-20.0	87.6	-167.1
	Clayton	120.2	-232.4	11.5	-15.0	13.4	-18.8	86.5	-165.0
PCC(R-Vine)		**202.4**	**-396.7**	**157.7**	**-301.3**	**74.3**	**-130.5**	**162.0**	**-306.0**
Bernstein		193.5	-122.0	220.5	-138.8	103.8	-40.3	242.1	-134.4

Panel B: Top level estimation (with R-Vine at the base level)

	Company 1		Company 2		Company 3		Company 4	
	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Indep	**0.866**	**0.386**	**0.027**	**0.978**	**0.596**	**0.551**	**1.725***	**0.085**
Gauss	0.488***	0.008	0.245*	0.092	0.198	0.176	0.586***	0.006
t	0.511**	0.011	0.297*	0.066	0.204	0.179	0.580***	0.005
Gumbel	0.465**	0.015	0.276*	0.071	0.184	0.213	0.562***	0.005
Clayton	0.550**	0.011	0.273*	0.082	0.228	0.139	0.632***	0.001

Note: Elliptical and Archimedean copulas are estimated via R package `copula`, HAC models are estimated via `copula` and `HAC`, R-Vine model is implemented via `VineCopula` and Bernstein D-Vine model is estimated via `penDvine`. *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit method for each portfolio. Four companies in panel B are chosen out of the Korean non-life insurers, who operate the five lines of business considered in our study, by the quartile of the total asset size. Company 1 is at the first quartile (25%) from the largest company, company 2 is at the second quartile (median), company 3 is at the third quartile and company 4 is the smallest.

Appendix G. Fitting result for German data

The considered realistic asset portfolio consists of 10 asset indices (Table G1) as a benchmark for a German non-life insurer introduced and employed in Eling et al. (2009).[156] The asset data are monthly returns from January 1998 to December 2006, each of which contains 108 observations. We also use historical claims data of the same insurer on a monthly basis for the same period.[157] The insurance portfolio is formed with six lines of business: industry fire, other fire, household storm, homeowner fire, homeowner storm and water damage insurances.

Table G1. List of Benchmark Indices and Descriptive Statistics (Eling et al., 2009)

Panel A: List of benchmark indices

Asset class	Index	Description	Asset allocation
Equity	Wrd_stock	MSCI Worldwide stock indices without EMU	5.0%
	EMU_stock	MSCI Stock indices in EMU without Germany	5.0%
	DE_stock	MSCI Stock index in Germany	5.0%
Fixed Income	US2Y	US 2 year Sovereign Bond Index	15.0%
	DE2Y	Germany 2 year Sovereign Bond Index	15.0%
	EMU2Y	Euro Zone 2 year Sovereign Bond Index	15.0%
	IBOXX corp	IBOXX Euro AAA 3 year Corporate bond index	15.0%
Money Market	EURM3	European Central Bank M3 money supply index	5.0%
Real Estate	Wrd_real	MSCI Worldwide real estate index	10.0%
	Euro_real	MSCI Europe real estate index	10.0%

Panel B: Descriptive statistics

Assets	mean	sd	skewness	kurtosis	Max	median	min	JB-test
Wrd_stock	0.0053	0.0414	-0.6789	0.6728	0.0891	0.0092	-0.1400	11.007***
EMU_stock	0.0072	0.0542	-0.7148	1.1359	0.1212	0.0194	-0.1699	16.085***
DE_stock	0.0065	0.0688	-0.8105	2.8450	0.2126	0.0104	-0.2791	51.446***
US2Y	0.0035	0.0053	0.1033	0.2326	0.0173	0.0032	-0.0112	0.584
DE2Y	0.0027	0.0035	0.1327	-0.7752	0.0098	0.0025	-0.0052	2.747
EMU2Y	0.0035	0.0092	-0.2086	-0.6162	0.0248	0.0044	-0.0172	2.275
IBOXX corp	0.0033	0.0087	-0.1727	-0.3203	0.0242	0.0039	-0.0219	0.880
EURM3	0.0054	0.0030	0.3758	1.0624	0.0163	0.0051	-0.0030	8.449**
Wrd_real	0.0090	0.0522	0.1560	1.8597	0.2002	0.0116	-0.1243	17.579***
Euro_real	0.0112	0.0431	-0.3219	0.5933	0.1359	0.0115	-0.1022	3.883

Underwriting	mean	sd	skewness	kurtosis	Max	median	min
Ind_fire	803,228.2	1,240,581.2	2.521	6.028	6,186,363.1	295,276.2	949.3
HO_fire	2,228,594.6	991,556.9	-0.238	0.042	4,663,559.9	2,251,845.3	25,119.4
Other_fire	1,313,010.5	917,484.8	1.054	0.543	4,146,761.6	984,767.3	3,384.7
HH_storm	44,862.9	74,288.1	4.431	26.823	594,514.1	18,223.3	0.0
HO_storm	1,833,345.2	3,990,165.4	6.633	53.926	37,075,463.2	462,420.4	5,411.4
Water	4,007,250.9	1,445,809.7	-1.097	2.166	7,930,423.2	4,277,080.0	101,092.1

Note: In panel A, MSCI stands for Morgan Stanley Capital International and EMU stands for European Monetary Union. *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. JB-test stands for Jarque-Bera test for normality assumption on the residual. Underwriting statistics are in *€ (Euro).*

[156] The portfolio consisting of bonds, stocks, real estate and money market securities accounts for 99.5% of all investments by insurance companies (Eling et al., 2009). The equity and real estate indices are the performance indices counting the dividend reinvestment performance provided by MSCI.

[157] Due to the limitation of available claims data, we test the model with German portfolios for a shorter period than with the Korean data. However, the implementation with the German data is meaningful because it can describe the adequacy of the estimated model in the different period (effectiveness for pre-financial crisis in 2008) and different market (European market vs. Asian market) as well as the firm size (large insurer vs. mid-sized insurer). Note that this underwriting dataset has been also used in Diers et al. (2012), who also investigates the adequacy of dependence models for the underwriting risk aggregation.

Marginal modeling

ARMA-GARCH fit is implemented for the asset portfolio and 12 distribution candidates are applied to the insurance portfolio, showing that skewed and long-tailed distributions are also mainly fitted for the German assets (see Table G2).

Table G2. Fitting Distributions for Asset and Insurance Portfolios

Panel A: Asset portfolio

	ARMA-order	AIC for ARMA	Fitted distribution	AIC for GARCH
Wrd_stock	(0,0)	-379.11	Skew normal	-396.16
EMU_stock	(2,2)	-322.21	Skew student	-351.18
DE_stock	(0,0)	-269.68	Skew student	-282.58
US2Y	(1,0)	-819.58	Normal	-822.95
DE2Y	(2,3)	-916.45	Normal	-934.22
EMU2Y	(1,1)	-699.01	Normal	-735.19
IBOXX corp	(3,0)	-720.41	Normal	-729.12
EURM3	(3,3)	-954.08	Skew normal	-973.93
Wrd_real	(3,1)	-346.53	Skew student	-355.91
Euro_real	(0,1)	-367.46	Normal	-380.42

Panel B: Insurance portfolio

	Ind_fire	HO_fire	Other_fire	HH_storm	HO_storm	Water
Skew Normal	3,359.63 (0.426***)	3,452.52 (0.549***)	3,369.76 (0.428***)	2,732.08 (0.274***)	3,604.26 (0.426***)	3,572.18 (0.600***)
Student-t	3,145.12 (0.683***)	3,516.66 (0.637***)	3,390.33 (0.693***)	2,795.55 (0.487***)	3,295.83 (0.655***)	3,635.47 (0.633***)
Skew student-t	**3,128.23** **(0.056)**	3,337.36 (0.309***)	3,246.36 (0.109)	2,514.79 (0.075)	3,292.71 (0.093)	3,457.85 (0.441***)
Lognormal	3,137.60 (0.103)	3,391.84 (0.267***)	3,285.51 (0.148**)	2,529.24 (0.077)	**3,289.55** **(0.095)**	3,520.44 (0.377***)
Gamma	3,140.02 (0.133**)	3,337.01 (0.210***)	3,247.31 (0.087)	2,514.24 (0.079)	3,310.62 (0.164***)	3,455.35 (0.345***)
Weibull	3,134.27 (0.100)	3,311.14 (0.141**)	**3,244.18** **(0.068)**	2,509.65 (0.073)	3,300.03 (0.149**)	3,409.59 (0.282***)
Inverse Gaussian	3,256.48 (0.411***)	3,462.41 (0.428***)	3,418.60 (0.421***)	2,958.41 (0.776***)	3,337.78 (0.213***)	3,582.77 (0.506***)
Cauchy	3,203.37 (0.220***)	3,321.02 (0.078)	3,289.30 (0.139**)	2,606.33 (0.252***)	3,381.99 (0.264***)	**3,309.34** **(0.062)**
Burr	3,499.17 (0.509***)	3,907.38 (0.540***)	3,759.83 (0.540***)	2,813.60 (0.501***)	3,669.61 (0.529***)	4,048.19 (0.536***)
GPD	3,128.36 (0.070)	3,315.54 (0.240***)	3,247.90 (0.114)	**2,509.01** **(0.061)**	3,289.94 (0.114)	3,430.96 (0.369***)
POT 90% (Norm-GPD)	3,267.40 (0.215***)	**3,296.22** **(0.070)**	3,266.07 (0.160***)	2,639.78 (0.201***)	3,451.95 (0.220***)	3,371.74 (0.246***)
POT 90% (Lognorm-GPD)	3,513.02 (0.657***)	3,403.94 (0.389***)	3,797.46 (0.741***)	2,552.05 (0.207***)	3,709.15 (0.694***)	4,085.97 (0.742***)

Note: The numbers in the parentheses are Kolmogorov-Smirnov statistics and *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold indicates the best fit for each insurance claims distribution.

Dependence modeling

With the fitted marginal factors, we carry out the same set of the dependence models as done with the Korean data and panel A of Table G3 presents testing results for such dependence models.

Table G3. Dependence Modeling Results

Panel A: Statistical testing for the dependence modeling at the base level

Family	Copula	Asset Portfolio Log-lik	Asset Portfolio AIC	Asset Portfolio GoF	Insurance Portfolio Log-lik	Insurance Portfolio AIC	Insurance Portfolio GoF
Elliptical	Gaussian	236.12	-382.24	0.016	175.74	-321.47	0.119***
	Student-t	256.51	-421.02	0.019*	175.75	-319.49	0.100***
Archimedean	Gumbel	8.81	-15.62	0.015	76.87	-151.73	0.230***
	Clayton	23.86	-45.72	0.011	122.78	-243.56	0.055
HAC	Gumbel	182.86	-347.72	0.175	142.77	-275.54	0.642
	Clayton	204.12	-390.24	0.175	162.36	-314.73	0.759
PCC	**R-Vine**	**289.42**	**-522.83**	**108.00**	**203.00**	**-374.00**	**56.43**
Bernstein (D-Vine)		278.95	-41.06	-	181.88	-73.72	-

Panel B: Dependence parameters in the first tree at the base level (R-Vine)

Copula	Parameter	Lower-tail dependency	Upper-tail dependency
Asset Portfolio			
$\theta_{3,8}$; $C_{DEstock,EURM3}$ (Gumbel)	1.09	-	0.116
$\theta_{2,4}$; $C_{EMUstock,US2Y}$ (90° Rotated Joe)	-1.73	-	-
$\theta_{3,2}$; $C_{DEstock,EMUstock}$ (Survival Gumbel)	2.39	0.663	-
$\theta_{1,3}$; $C_{Wrdstock,DEstock}$ (Student-t)	0.81/2.53	0.583	0.583
$\theta_{6,7}$; $C_{EMU2Y,IBOXXcor}$ (Survival Gumbel)	3.18	0.756	-
$\theta_{5,6}$; $C_{DE2Y,EMU2Y}$ (Survival Clayton)	0.30	-	0.098
$\theta_{9,1}$; $C_{Wrdreal,Wrdstock}$ (Student-t)	0.48/2.60	0.328	0.328
$\theta_{9,5}$; $C_{Wrdreal,DE2Y}$ (Frank)	1.56	-	-
$\theta_{10,9}$; $C_{EURreal,Wrdreal}$ (Survival Joe)	1.91	0.563	-
Insurance Portfolio			
$\theta_{2,6}$; $C_{HOfire,HOwater}$ (Survival Joe)	1.83	0.540	-
$\theta_{2,3}$; $C_{HOfire,Otherfire}$ (Survival Joe)	1.93	0.569	-
$\theta_{4,1}$; $C_{HHstorm,INDfire}$ (Survival Gumbel)	1.44	0.383	-
$\theta_{5,2}$; $C_{HOstorm,HOfire}$ (Survival Gumbel)	1.62	0.467	-
$\theta_{5,4}$; $C_{HOstorm,HHstorm}$ (Gaussian)	0.93	-	-

Panel C: Statistical testing at the top level

		R-Vine (Base level) Statistics	R-Vine (Base level) P-value	Linear (Base level) Statistics	Linear (Base level) P-value
Independence test		**0.8083**	**0.419**	**0.2283**	**0.819**
Elliptical	Gaussian copula	0.1293	0.415	0.1795	0.217
	Student-t copula	0.1639	0.260	0.1979	0.167
Archimedean	Gumbel copula	0.1295	0.376	0.1751	0.238
	Clayton copula	0.1585	0.282	0.2140	0.157

Note: The table shows the statistical testing results for dependence models (panel A), estimated parameters only in the first tree of the structure representing the strongest dependency in the structure (panel B) and goodness-of-fit test results for the top-level dependency (panel C). The numbers of the goodness-of-fit results in Panel A indicate the test statistics except for those of HAC models showing p-values (see footnote 142). *,**,*** indicate that the p-value is less than the significance levels, 10%, 5% and 1% respectively. The bold in panel A indicates the best fit method for each portfolio. The parametric R-Vine is sequentially determined by R-package `VineCopula`. The subscripts of copulas in panel B indicate the following:
Asset: 1=Wrdstock; 2=EMUstock; 3=DEstock; 4=US2Y; 5=DE2Y; 6=EMU2Y; 7=IBOXXcor; 8=EURM3; 9=Wrdreal; 10=EUreal
Insurance: 1=INDfire; 2=HOfire; 3=Otherfire; 4=HHstorm; 5=HOstorm; 6=HOwater

It is observed that R-Vine model best describes the dependence structures of both portfolios and elliptical copulas demonstrate a better fit than HAC and Bernstein models as discovered with the Korean data. For the top-level dependency in panel C, it can be concluded, based on the p-values, that the independence structure reflects the top-level structure of the German data better than other dependence structures under the correlation assumption by the standard models. The parameter estimation result by R-Vine shows that tail risks are observed in most pairs of both portfolios, particularly lower tail risks for the underwriting portfolio.

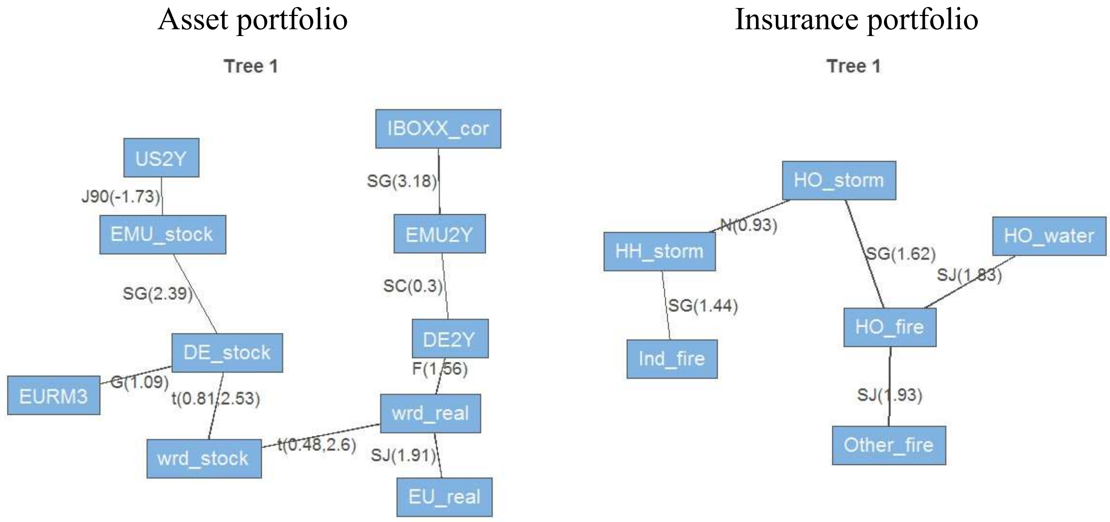


Figure G1. Graphical structure of dependency in the first tree. The estimated copula functions and their dependence parameters are illustrated at edges. The copula functions at edges are specified as follows: J: Joe; SG: Survival Gumbel; G: Gumbel; t: Student-t; SJ: Survival Joe; SC: Survival Clayton; F: Frank; N: Normal (Gaussian).

Full application to risk measurement

As a robustness of our finding from Korean case, we generate the economic capitals for the German data from different dependence structures. The result of the comparison between the estimates can be found in Table G4. We can clearly observe that the top-level dependence structure of the German data ends up with the same result as in the Korean case, showing that the lowest level of capital is generated from the independence structure, whereas student-t and Gumbel capturing the correlated risk in the right tail produce higher levels. Similar patterns are identified at the base level in that the independence structure in the high-dimension significantly lowers the level of the economic capital and the Bernstein model estimates it between the parametric models and the independence model.

Table G4. Comparison with Regulatory Standards

(in € million)			Top-level dependence (Bivariate) Indep	Gauss	t	Gumbel	Clayton
Base level dependence	VaR at 99%	R-Vine	**31.27**	35.56	38.78	36.99	33.53
		Gauss	31.04	34.60	36.26	36.17	33.48
		t	32.93	36.23	38.87	37.86	35.58
		HAC-best	27.36	32.34	31.98	31.19	29.78
		HAC-worst	44.21	48.11	51.63	51.03	47.81
		AC-best	16.32	20.93	21.01	20.41	18.47
		AC-worst	33.06	36.17	38.30	37.96	35.62
		Independence	5.93	7.63	9.47	8.07	7.33
		Bernstein	22.64	26.65	27.39	27.32	25.07
	VaR at 99.5%	R-Vine	**51.24**	58.77	59.42	59.32	53.94
		Gauss	50.26	56.63	58.36	56.84	52.39
		t	52.76	59.03	60.51	59.71	55.24
		HAC-best	47.15	51.56	52.70	52.37	48.29
		HAC-worst	68.06	74.23	75.63	75.17	68.22
		AC-best	35.01	39.24	40.61	40.33	36.90
		AC-worst	54.21	59.48	62.32	61.18	55.40
		Independence	23.61	28.12	29.17	29.03	24.71
		Bernstein	42.02	46.56	47.28	46.82	42.12
	TVaR at 99%	R-Vine	**55.27**	59.81	62.19	60.99	56.11
		Gauss	53.44	57.28	59.95	58.89	55.25
		t	56.05	59.98	62.84	62.08	57.54
		HAC-best	48.88	52.65	54.37	53.13	50.27
		HAC-worst	71.49	75.07	79.38	78.72	72.42
		AC-best	36.23	39.95	41.30	40.87	38.05
		AC-worst	56.54	60.32	64.11	63.80	58.44
		Independence	24.61	29.09	30.03	29.53	26.57
		Bernstein	45.03	47.95	50.00	48.91	45.68

	K-RBC	SII	SST
Standard models	45.51	80.68	94.39

Note: HAC stands for hierarchical Archimedean copulas and AC indicates simple Archimedean copulas. The bold in the upper panel indicates the best fitted model for the empirical data.

Application to asset allocation

As done with the Korean case in Table D2, we also investigate the effect of different asset allocation strategies with the German case. The corresponding regulatory measure (SII) and confidence level (VaR 99.5%) are used for comparison. Two examples of the allocation are considered in panel A of Table G5, showing that the second case is riskier than the first case. The economic capital of the riskier case under the best fit model is 63.6% higher than that of the less risky case, implying that 1% increase in the portfolio volatility can lead to 6.72% increase in the economic capital. Therefore, the German case again confirms that a riskier asset portfolio can require a higher capital size for an insurer, possibly leading to an increase in the restructuring cost. It also supports the finding that the higher the risk is, the higher the difference between the internal and standard model is, showing 64.6% overestimation in the first case and 77.4% in the second case.

Table G5. Diversified Investment Strategies

Panel A: Asset allocations

		Allocation 1	Allocation 2
Equity	Wrd_stock	10%	20%
	EMU_stock	10%	20%
	DE_stock	10%	20%
Fixed income	US2Y	10%	0%
	DE2Y	15%	0%
	EMU2Y	15%	0%
	IBOXX corp	10%	0%
Money Market	EURM3	0%	0%
Real estate	Wrd_real	10%	20%
	Euro_real	10%	20%

Panel B: Application to the economic capital

			Top-level dependence					
(in € million)			Indep	Gauss	t	Gumbel	Clayton	SII
Base-level dependence	Allocation 1 (VaR 99.5%) · μ=6.89% · σ=12.40%	R-Vine	**65.63**	73.63	74.95	74.01	68.58	108.0
		Gauss	65.07	70.68	72.31	70.74	65.67	
		t	66.28	74.28	74.97	74.67	69.30	
		HAC-best	60.42	64.73	65.87	65.56	60.64	
		HAC-worst	82.75	90.48	91.79	90.97	82.98	
		AC-best	40.34	44.50	45.87	45.44	41.80	
		AC-worst	60.28	66.48	68.94	68.17	61.72	
		Independence	26.99	31.48	32.85	32.42	27.61	
		Bernstein	52.53	56.81	58.53	57.48	52.76	
			Indep	Gauss	t	Gumbel	Clayton	SII
	Allocation 2 (VaR 99.5%) · μ=9.84% · σ=21.86%	R-Vine	**107.37**	117.08	118.47	117.43	110.68	190.5
		Gauss	104.26	111.66	113.08	112.52	104.36	
		t	111.16	120.90	122.49	121.40	112.54	
		HAC-best	97.22	102.66	104.68	103.85	98.13	
		HAC-worst	128.43	138.03	140.40	138.17	129.13	
		AC-best	62.25	66.80	68.67	68.00	62.41	
		AC-worst	84.75	92.31	95.32	93.51	85.40	
		Independence	44.12	49.09	51.42	50.54	44.25	
		Bernstein	85.51	91.58	94.23	92.13	86.94	

Note: μ is the mean of the portfolio and σ is the volatility of the portfolio on an annual basis. The bold in panel B indicates the best fitted model for the empirical data.

References

Aas, K., and Berg, D. (2009). Models for construction of multivariate dependence - a comparison study. *European Journal of Finance, 15*(7-8), 639-659.

Aas, K., and Puccetti, G. (2014). Bounds on total economic capital: the DNB case study. *Extremes, 17*(4), 693-715.

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics,* 44(1), 182-198.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.

Avanzi, B., Taylor, G., and Wong, B. (2016). Correlations between insurance lines of business: An illusion or a real phenomenon? Some methodological considerations. *ASTIN Bulletin*, *46*(2), 225-263.

Bedford, T., and Cooke, R. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence, 32*(1-4), 245-268.

Bedford, T., and Cooke, R. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics, 30*(4), 1031-1068.

Bermudez, L., Ferri, A., and Guillen, M. (2013). A correlation sensitivity analysis of non-life underwriting risk in solvency capital requirement estimation. *ASTIN Bulletin, 43*(1), 21-37.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.

Bollerslev, T., and Wooldridge, J. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews, 11*(2), 143-172.

Bølviken, E., and Guillen, M. (2017). Risk aggregation in Solvency II through recursive log-normals. *Insurance: Mathematics and Economics, 73*, 20-26.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). New Jersey: Prentice-Hall.

Brechmann, E. C., and Czado, C. (2013). Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50. *Statistics and Risk Modeling, 30*(4), 307-342.

Brechmann, E. C., and Schepsmeier, U. (2013). Modeling dependence with C- and D-vine copulas: The R-package CDVine. *Journal of Statistical Software, 52*(3), 1-27.

Brechmann, E. C., Czado, C., and Paterlini, S. (2014). Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *Journal of Banking and Finance, 40*, 271-285.

Christiansen, M., Denuit, M., and Lazar, D. (2012). The Solvency II square-root formula for systematic biometric risk. *Insurance: Mathematics and Economics, 50*(2), 257-265.

Clarke, K., (2007). A simple distribution-free test for nonnested model selection. *Political Analysis, 15*(3), 347-363.

Cooke, R., Joe, H., and Aas, K. (2011). Vines arise. In *Dependence modeling: Vine copula handbook* (ed. D. Kurowicka and H. Joe), pp. 37-71. Singapore: World Scientific.

Côté, M. P., and Genest, C. (2015). A copula-based risk aggregation model. *Canadian Journal of Statistics*, *43*(1), 60-81.

Cummins, J. D., (1991). Statistical and financial models of insurance pricing and the insurance firm. *Journal of Risk and Insurance, 58*(2), 261-302.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula theory and its applications: Proceedings of the workshop in Warsaw* (eds. P. Jaworski, F. Durante, W. Härdle and T. Rychlik), pp. 93-109. Springer.

Czado, C., Jeske, S., and Hofmann, M. (2013). Selection strategies for regular vine copulae. *Journal de la Société Française de Statistique, 154*(1), 174-191.

Czado, C., Schepsmeier, U., and Min, A. (2012). Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling, 12*(3), 229-255.

Daul, S., De Giorgi, E., Lindskog, F., and McNeil, A. (2003). The grouped t-copula with an application to credit risk. *Risk, 16*, 73-76.

Devineau, L., and Loisel, S. (2009). Risk aggregation in Solvency II: How to converge the approaches of the internal models and those of the standard formula?. *Bulletin Français d'Actuariat*, *9*(18), 107-145.

Diers, D., Eling, M., and Marek, S. (2012). Dependence modeling in non-life insurance using the Bernstein copula. *Insurance: Mathematics and Economics, 50*(3), 430-436.

Dissmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis, 59*, 52-69.

Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics, 51*(2), 239-248.

Eling, M. (2014). Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics, 59*, 45-56.

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

Eling, M., and Toplek, D. (2009). Modeling and management of nonlinear dependencies - Copulas in dynamic financial analysis. *Journal of Risk and Insurance, 76*(3), 651-681.

Eling, M., Gatzert, N., and Schmeiser, H. (2009). Minimum standards for investment performance: A new perspective on non-life insurer solvency. *Insurance: Mathematics and Ecnomics, 45*(1), 113-122.

Eling, M., Schmeiser, H., and Schmit, J. (2007). The Solvency II process: Overview and critical analysis. *Risk Management and Insurance Review, 10*(1), 69-85.

Embrechts, P. (2002). Insurance analytics. *British Actuarial Journal, 8*(4), 639-641.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance.* Berlin Heidelberg: Springer-Verlag.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas and applications to risk management. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich*.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (ed. M. A. H. Dempster), pp. 176-223. Cambridge University Press.

Engle, R., (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica, 50*(4), 987-1007.

Ernst & Young (EY). (2013). *European Solvency II survey.* London: Ernst & Young.

European Commission (EC). (2010). *QIS5 Technical Specifications.* Brussels.

European Commission (EC). (2014). *Commission Delegated Regulation (EU) 2015/35: Supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).* Official Journal of the European Union.

Filipovic, D. (2009). Multi-level risk aggregation. *ASTIN Bulletin, 39*(2), 565-575.

Financial Services Commission (FSC). (2014). *Regulations for Insurance Act (in Korean).* Seoul: Financial Services Commission.

Financial Supervisory Service (FSS). (2017). *Guide to Korea's Risk Based Capital (in Korean).* Seoul: Financial Supervisory Service (FSS).

FINMA. (2006). *Technical Document on the Swiss Solvency Test*. Bern: Swiss Financial Market Supervisory Authority (FINMA).

Frachot, A., Georges, P., and Roncalli, T. (2001). Loss distribution approach for operational risk. *Working paper*. Retrieved from https://ssrn.com/abstract=1032523.

Frees, E., (2015). Analytics of insurance markets. *Annual Review of Financial Economics, 7*, 253-277.

Frees, E., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks, 4*(1), 4.

Fu, L., and Moncher, R. (2004). Severity distributions for GLMs: Gamma or lognormal? Evidence from Monte Carlo simulations. *Casualty Actuarial Society Discussion Paper Program*, pp. 149-230.

Gatzert, N., and Martin, M. (2012). Quantifying credit and market risk under Solvency II: Standard approach versus internal model. *Insurance: Mathematics and Economics, 51*(3), 649-666.

Genest, C., and Remillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test, 13*(2), 335-369.

Genest, C., and Remillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semi-parametric models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 44*(6), 1096-1127.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika, 82*(3), 543-552.

Gilbert, C. (2016). Asset liability management. In *the IAA Risk Book* (Chapter 13). International Actuarial Association.

Gründl, H., Dong, M., and Gal, J. (2016). The evolution of insurer portfolio investment strategies for long-term investing. *OECD Journal: Financial Market Trends*, 1-55.

Haff, I. H. (2013). Parameter estimation for pair-copula constructions. *Bernoulli, 19*(2), 462-491.

Hofert, M., and Mächler, M. (2014). A graphical goodness-of-fit test for dependence models in higher dimensions. *Journal of Computational and Graphical Statistics, 23*(3), 700-716.

Holzmüller, I. (2009). The United States RBC standards, Solvency II and the Swiss solvency test: a comparative assessment. *The Geneva Papers on Risk and Insurance – Issues and Practice*, *34*(1), 56-77.

Huang, W., and Prokhorov, A. (2014). A goodness-of-fit test for copulas. *Econometric Reviews, 33*(7), 751-771.

International Association of Insurance Supervisors (IAIS). (2018). Risk-based global insurance capital standard version 2.0: Public consultation document. Basel.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m − 1)/2 bivariate dependence parameters. In *Distributions with fixed marginals and related topics* (ed. L. Rüschendorf, B. Schweizer and M. Taylor)*, 28*, 120-141.

Joe, H. (1997). *Multivariate models and dependence concepts.* New York: Chapman & Hall.

Joe, H., and Xu, J. J. (1996). *The estimation method of inference functions for margins for multivariate models.* Department of Statistics. University of British Columbia.

Jondeau, E., and Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance, 25*(5), 827-853.

Kauermann, G., and Schellhase, C. (2014). Flexible pair-copula estimation in D-vines using bivariate penalized splines. *Statistics and Computing, 24*(6), 1081-1100.

Kauermann, G., Schellhase, C., and Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical B-splines. *Scandinavian Journal of Statistics, 40*(4), 685-705.

Kim, S.-J. (2015). A study of efficiency about nonlife insurance asset management to low interest. *Journal of Service Research and Studies (in Korean), 5*(2), 35-49.

Kurowicka, D., and Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling.* West Sussex: John Wiley & Sons.

Low, R. K. Y., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it?. *Journal of Banking and Finance, 37*(8), 3085-3099.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools* (Revised Edition). New Jersey: Princeton University Press.

Milliman. (2008). *Benefits and challenges of using an internal model for Solvency II.* London: Milliman.

Mittnik, S. (2011). *Solvency II calibrations: where curiosity meets spuriosity*. Center for Quantitative Risk Analysis (CEQURA), University of Munich.

Moscadelli, M. (2004). The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee. *Working paper*. Retrieved from https://ssrn.com/abstract=557214.

Nelsen, R. B. (2006). *An Introduction to Copulas* (2nd ed.). New York: Springer.

Okhrin, O., and Ristig, A. (2014). Hierarchical Archimedean copulae: The HAC package. *Journal of Statistical Software, 58*(4), 1-20.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics, 173*(2), 189-204.

Pfeifer, D., and Strassburger, D. (2008). Solvency II: Stability problems with the SCR aggregation formula. *Scandinavian Actuarial Journal, 2008*(1), 61-77.

Pfeifer, D., Strassburger, D., and Philipps, J. (2009). Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas. *Working Paper*. University of Oldenburg.

Rosenberg, J., and Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics, 79*(3), 569-614.

Sancetta, A., and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory, 20*(3), 535-562.

Savelli, N., and Clemente, G. P. (2011). Hierarchical structures in the aggregation of premium risk for insurance underwriting. *Scandinavian Actuarial Journal, 2011*(3), 193-213.

Savu, C., and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance, 10*(3), 295-304.

Scheffer, M., and Weiss, G. (2017). Smooth nonparametric Bernstein vine copulas. *Quantitative Finance, 17*(1), 139-156.

Schepsmeier, U. (2015). Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review. *Journal of Multivaraite Analysis, 138*, 34-52.

Schepsmeier, U. (2019). A goodness-of-fit test for regular vine copula models. *Econometric Reviews*, *38*(1), 25-46.

Schepsmeier, U., and Czado, C. (2016). Dependence modelling with regular vine copula models: a case-study for car crash simulation data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 65*(3), 415-429.

Shevchenko, P. (2011). *Modelling operational risk using Bayesian inference.* Heidelberg: Springer.

Siegel, C. (2016). Solvency assessment for insurance groups in the United States and Europe—A comparison of regulatory frameworks. *In The Geneva Papers* (ed. C. Courbage), pp. 139-168. London: Palgrave Macmillan.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Inst. Statist. Univ Paris 8*, 229-231.

Statista. (2018). *Average annual wage in South Korea adjusted for Purchasing Power Parity from 2000 to 2016.* Retrieved from https://www.statista.com/statistics/557759/south-korea-average-annual-wage/.

Swiss Federal Office of Private Insurance (FOPI). (2004). *White paper of the Swiss solvency test.* Bern: Swiss Financial Market Supervisory Authority (FINMA).

Swiss Re. (2018). *Sigma 03/2018: World insurance in 2017.* Zurich: Swiss Re.

Tang, A., and Valdez, E. (2009). Economic capital and the aggregation of risks using copulas. *Working paper*. Retrieved from https://ssrn.com/abstract=1347675.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307-333.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50*(1), 1-25.

Yang, J., Chen, Z., Wang, F., and Wang, R. (2015). Composite Bernstein copulas. *ASTIN Bulletin, 45*(2), 445-475.

Summary table on the doctoral dissertation

	Essay I	Essay II (Single-authored)	Essay III	Essay IV
Title	Copula approaches for modeling cross-sectional dependence of data breach losses	Probable maximum cyber loss: Empirical estimation and reinsurance design with public-private partnership	Decision-making on cyber risk management: Interaction between market insurance and risk control measures under prospect theory	Risk aggregation in non-life insurance: Standard models vs. internal models
Author(s)	Martin Eling Kwangmin Jung	Kwangmin Jung	Martin Eling Kwangmin Jung	Martin Eling Kwangmin Jung
Main topic	Cyber risk	Cyber risk	Cyber risk	Risk aggregation
Data	Data breach records from Privacy Rights Clearinghouse (Jan 2005 - Dec 2016)	Data breach records from Privacy Rights Clearinghouse (Jan 2005 - Dec 2018)	· Theoretical modeling · Data breach records from Privacy Rights Clearinghouse (Jan 2005 - Dec 2018)	· Korean asset and underwriting data (Jan 2002 - Dec 2016) · German asset and underwriting data (Jan 1998 - Dec 2006)
Methodology	Collective risk model, Vine copulas, Elliptical copulas, Archimedean copulas	Block maxima with extreme value theory Time series analysis for cyber loss maxima Extreme value copulas	Prospect theory	ARMA-GARCH, Individual risk model, Vine copulas, Hierarchical (nested) Archimedean copulas, Elliptical copulas
Main points	· Information on data breach events has been classified in the different categorization settings and analyzed in different dependent structures. · It shows how differently high-dimensional dependence constructions influence in cyber-insurance premiums and risk measures.	· Investigate the statistical features of data breach maximum losses from a variety of time blocks (weekly, bi-weekly and monthly) by analyzing temporal dependency and fitting GEV distribution. · Provide the estimates of probable maximum cyber loss to be used as the cover limit of a reinsurance scheme.	· Establish a conceptual model to explain a behavioral bias against cyber risk with two concepts: reference-dependent utility model under prospect theory and the presence of interdependent risk. · Set self-protection effort on cyber security as the reference point reflecting the status quo in the current business environment.	· Different types of dependence models including PCC and HAC are used to aggregate two risk modules, asset and underwriting. · Economic capitals from different copulas are compared with different regulatory standards (Korean RBC, Solvency II and Swiss Solvency Test).
Contribution	A comprehensive analysis on dependence structure of data breach risks is suggested, which reveals non-linear dependency between risk factors proved by significant asymmetric tail dependency.	· The analysis focuses purely on extremes (maxima) of cyber losses for a certain time period and can show the statistical features of extremes for the forecasting. · The essay offers an empirical benchmark on a potential reinsurance portfolio against cyber risk.	It helps explain a behavioral bias observed in the market that an agent with the reference point of self-protection as an essential effort against cyber risk tends to not implement additional measures for cyber risk management.	· Two-step aggregation process is constructed and implemented by a cutting-edge copula model. · Two sides of balance sheet (asset and liability) are considered and two market datasets (Korean and German) are used.
Status	Published in *Insurance: Mathematics and Economics* (September 2018)	Presentations: · ARIA (2019) in San Francisco · S.S.Huebner Doctoral Colloquium · APRIA (2019) in Seoul · DVfVW (2019) in Berlin · Ph.D. in Finance seminar · I.VW. Research seminar	Presentations: · EGRIE (2019) in Rome · ARIA (2019) in San Francisco · DVfVW (2019) in Berlin · APRIA (2018) in Singapore · ICIR workshop (2018) in Frankfurt · PEF/PiF poster seminar · I.VW. Research seminar	· Conditionally accepted upon minor revisions in *ASTIN Bulletin* and resubmitted after thorough revisions.

Curriculum Vitae

Personal Information

Name:	Kwangmin Jung
Date of Birth:	09.08.1986
Place of Birth:	Seoul, Republic of Korea
Nationality:	South Korean

Education

09/2016 – 02/2020	**University of St. Gallen**, St. Gallen, Switzerland Ph.D. in Finance
09/2014 – 06/2016	**EBS Business School**, Oestrich-Winkel, Germany MSc in Automotive Management (Dual-degree program)
09/2013 – 08/2014	**Cass Business School** (City University of London), London, U.K. MSc in Quantitative Finance (Distinction)
03/2006 – 02/2010	**Yonsei University**, Seoul, Republic of Korea BA in Applied Statistics (Honors) / Minor: Economics

Work Experience

09/2016 – 01/2020	**Institute of Insurance Economics** (I.VW.) Project manager and research assistant
03/2010 – 06/2012	**Republic of Korea Army** (ROKA) Chief of Intelligence (60th Artillery Battalion) – 1st Lieutenant