

Zängerle, Daniel; Schiereck, Dirk

Article — Published Version

Modelling and predicting enterprise-level cyber risks in the context of sparse data availability

The Geneva Papers on Risk and Insurance - Issues and Practice

Provided in Cooperation with:

Springer Nature

Suggested Citation: Zängerle, Daniel; Schiereck, Dirk (2022) : Modelling and predicting enterprise-level cyber risks in the context of sparse data availability, The Geneva Papers on Risk and Insurance - Issues and Practice, ISSN 1468-0440, Palgrave Macmillan, London, Vol. 48, Iss. 2, pp. 434-462, <https://doi.org/10.1057/s41288-022-00282-6>

This Version is available at:

<https://hdl.handle.net/10419/309580>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Modelling and predicting enterprise-level cyber risks in the context of sparse data availability

Daniel Zängerle¹ · Dirk Schiereck¹

Received: 1 April 2022 / Accepted: 1 December 2022 / Published online: 10 December 2022
© The Author(s) 2022

Abstract

Despite growing attention to cyber risks in research and practice, quantitative cyber risk assessments remain limited, mainly due to a lack of reliable data. This analysis leverages sparse historical data to quantify the financial impact of cyber incidents at the enterprise level. For this purpose, an operational risk database—which has not been previously used in cyber research—was examined to model and predict the likelihood, severity and time dependence of a company’s cyber risk exposure. The proposed model can predict a negative time correlation, indicating that individual cyber exposure is increasing if no cyber loss has been reported in previous years, and vice versa. The results suggest that the probability of a cyber incident correlates with the subindustry, with the insurance sector being particularly exposed. The predicted financial losses from a cyber incident are less extreme than cited in recent investigations. The study confirms that cyber risks are heavy-tailed, jeopardising business operations and profitability.

Keywords Cyber risk modelling · Cyber risk management · Cyber insurance · Vine copula · Sparse time series

Introduction

Cyber risks are one of the greatest threats of the twenty-first century (WEF 2021). Originally arising from the use of information technology (IT), cyber risks have since increased in both number and financial impact, especially due to rapidly progressing digitisation, worldwide interconnection and the introduction of new digital

✉ Daniel Zängerle
daniel.zaengerle@stud.tu-darmstadt.de

Dirk Schiereck
dirk.schiereck@tu-darmstadt.de

¹ Department of Corporate Finance, Technical University of Darmstadt, Hochschulstr. 1, 64289 Darmstadt, Germany



products and services (Njegomir and Marović 2012; Rakes et al. 2012; Aldasoro et al. 2020). The cost of cyber incidents is estimated at more than USD 1 trillion (McAfee 2020) globally. Cyber incidents not only jeopardise private customers but also pose new challenges for companies and organisations (Njegomir and Marović 2012; Choudhry 2014; Bendovschi 2015; Wrede et al. 2018; Aldasoro et al. 2020). Despite the high awareness of cyber risk among corporate decision makers (Smidt and Botzen 2018) and insurance companies (Pooser et al. 2018), enterprise risk management (ERM) still neglects the associated risks, with some industries and firms even adopting a passive stance (Ashby et al. 2018; Pooser et al. 2018).

Effective cyber risk management should be comprehensively incorporated into ERM rather than analysed in an isolated manner, such as exclusively in IT departments (Marotta and McShane 2018; Shetty et al. 2018; Poyraz et al. 2020). Furthermore, there is evidence that cyber risk management processes are generally qualitative and are missing quantitative findings (Pálsson et al. 2020). The usual method to quantify cyber risk is through an analysis of historical cyber incidents from verifiable sources and the performance of empirical, statistical and actuarial examinations to determine the financial impact and likelihood of a cyber incident in a specific organisation (Smidt and Botzen 2018; Pálsson et al. 2020). However, the lack of data restrains the quality of such assessments and constitutes the main research gap in the cyber risk literature (Eling and Schnell 2016; Marotta et al. 2017; Boyer 2020).

To address this, we quantitatively assess the financial impact of cyber risks at the enterprise level using sparse historical data. Our analysis is based on the Öffentliche Schadenfälle OpRisk (ÖffSchOR) database—an operational risk (OpRisk) database on publicly disclosed loss events in the European financial sector—which has not been adapted to cyber risk research. We apply advanced modelling techniques suggested by Shi and Yang (2018), Eling and Wirfs (2019) and Fang et al. (2021) to predict the likelihood and loss exposure of a potential cyber incident. We specifically use statistical dependence, modelled by a D-vine copula structure, to cope with the sparsity of events in a multivariate time series setting. In doing so, we provide new empirical evidence and quantitative results on actual cyber risk losses at the company level. Our findings suggest that cyber risks are less severe than recent studies claim and that subindustries must be separately modelled. Additionally, our results support the insight that cyber risks are heavy tailed, with an extreme cyber incident as a worst-case scenario that would seriously harm or default a company (Eling and Wirfs 2019; Wheatley et al. 2021).

The results of this study provide one of the first quantitative insights on the nature of cyber risks and introduce a new dataset to cyber research. The outlined methodology allows researchers and practitioners, in particular cyber insurers, to assess cyber risks despite the lack of larger datasets and to combine with existing pricing tools in order to evaluate risk-based premiums (Nurse et al. 2020; Cremer et al. 2022). Our study, thus, contributes to the limited research available on the empirical quantification of cyber risks and to a better understanding within the field.

The remainder of this paper is structured as follows. The next section provides a summary of the most relevant literature. Then, we introduce the dataset and methodology. The fourth section presents the results of our analysis. The final section



concludes with a discussion of the findings and limitations of the study as well as future research possibilities.

Literature review

Compared to the prevailing research on operational risk modelling (see e.g. Cox 2012; MacKenzie 2014), cyber risk analyses are still very limited (Eling 2020).¹ This lack of research is often linked to the limited availability of cyber loss data (Maillart and Sornette 2010; Biener et al. 2015), which is typically not disclosed by organisations in an effort to avoid reputational damage (Giudici and Raffinetti 2020). Despite several public and private initiatives to form databases (see the next section), companies have little incentive to share loss information in a public or consortium repository (Palsson et al. 2020). Initiatives such as the introduction of new reporting requirements for cyber incidents and data breaches—in the U.S. by the National Conference of State Legislatures (NCSL 2016) and in Europe by the European Union (EU 2016)—might improve modelling techniques (Eling and Wirfs 2019) but are still incapable of delivering new insights. Further, the introduction of individual cyber risk definitions leads to a maze of terms rather than a comprehensive and unified terminology and understanding of cyber risks (Zängerle and Schiereck 2022).

The lack of cyber risk data has also been addressed in recent publications. In particular, Cremer et al. (2022) conduct a comprehensive and systematic review of cyber data availability, identifying only 79 datasets from a preliminary 5,219 peer-reviewed cyber studies. Furthermore, most of these databases focus on technical cybersecurity aspects, such as intrusion detection and machine learning, with only a fraction of available datasets on cyber risks. The authors find that the lack of available data on cyber risks is a serious problem for stakeholders that undermines collective efforts to better manage these risks. This interpretation is supported by Romanosky et al. (2019), who show that (cyber) insurers in the U.S. have no historic or credible data to assess the loss expectation of cyber insurance coverages.

Due to the scarcity of cyber loss information, data breaches, mainly in the U.S., have received the most attention in empirical research (see e.g. Maillart and Sornette 2010; Edwards et al. 2016; Wheatley et al. 2016; Eling and Loperfido 2017; Eling 2018; Xu et al. 2018; Wheatley et al. 2021). Attempts have also been made to assess the monetary impact of data breaches (see e.g. Layton and Watters 2014; Romanosky 2016; Ruan 2017; Poyraz et al. 2020). However, as addressed by Woods and Böhme (2021), these studies have produced contradictory results, depending on the dataset and methodology applied. Furthermore, Eling and Wirfs (2019) found that data breaches account for just 25% of all cyber events, and the estimated distribution of breached records does not align with that of the actual financial cost of cyber

¹ For a comprehensive review and the status quo of cyber risk research, we refer to Eling and Schnell (2016), Marotta et al. (2017), Eling (2020) and Woods and Böhme (2021).



incidents. To this day, only a few studies have assessed the financial impact of cyber incidents in a comprehensive way.

Biener et al. (2015) analyse cyber losses from the SAS operational loss database and emphasise the distinct characteristics of cyber risks, including the lack of data, information asymmetries and highly interrelated losses. However, the authors focus on insurability rather than the modelling and prediction of cyber losses. Romanosky (2016) provides the first quantitative insights from actual loss information based on the Advisen dataset but concentrates on descriptive statistics.² Later, Palsson et al. (2020) use the same database to model the financial cost of different cyber event types by applying a random forest algorithm. Although the data are not sufficiently detailed to construct a predictive model with high accuracy, the researchers identify relevant factors affecting the expenses of such incidents. Similar to our examination, Eling and Wirfs (2019) analyse the actual costs of cyber incidents from the SAS loss database with statistical and actuarial methods. By applying the peaks-over-threshold (POT) method from extreme value theory (EVT), they find that cyber risks are distinct from other risk categories and argue that researchers must distinguish between ‘cyber risks of daily life’ and ‘extreme cyber risks’. In addition, they present a simulation study for practical application. We apply techniques similar to those of Eling and Wirfs (2019), who focus on monthly aggregated observations from all entities available, treated as one sample from a single distribution. We, however, utilise enterprise-level sparse time series data from a database that has not yet been used in the context of cyber risk modelling.

A second research stream focusing on the modelling of dependence structures has recently emerged (Eling 2020). In particular, the application of copula theory is widely accepted due to the ability to use any marginal distribution, which is essential for diverse cyber risk classes, and to address non-linear dependencies (see e.g. Böhme and Kataria 2006; Herath and Herath 2011; Mukhopadhyay et al. 2013). Further studies have extended these approaches to multivariate settings using vine copulas (see e.g. Joe 1997; Bedford and Cooke 2002; Kurowicka and Cooke 2006; Aas et al. 2009), which generate a multivariate copula based on iterative and bivariate pairwise copula constructions (PCC). The D-vine, a distinct vine copula, is particularly structured and simple to interpret in the time series context (Zhao et al. 2020). For example, Peng et al. (2016) use honeypot data to model multivariate and extreme cyber risks with marked point processes and vine copulas, later progressing with a vine copula GARCH model (Peng et al. 2018). Shi and Yang (2018) analyse the temporal dependence in longitudinal data by a D-vine copula. Xu et al. (2018) model the interarrival times of data breaches by ARMA-GARCH and joint density with copula. Eling and Jung (2018) apply the Privacy Rights Clearinghouse (PRC) dataset and model the cross-sectional dependence of data breaches. They find that vine structures exhibit a better fit than simple elliptical or Archimedean copulas. Fang et al. (2021) also study the same dataset, but in a multivariate time series setting with sparse observations at the enterprise level. Therefore, they propose

² The author uses a logistic regression model to analyse the actual costs of data breaches only. A new approach for assessing the monetary impact of mega data breaches is suggested by Poyraz et al. (2020).



a D-vine copula to model the serial trend. We adopt this framework to model the financial impact of actual cyber incidents rather than data breaches alone.

The current emergence of network models also offers a new, more appealing path for cyber risk modelling (see e.g. Fahrenwaldt et al. 2018; Jevtić and Lanchier 2020; Wu et al. 2021). However, these advanced predictive models are currently limited to simulation studies, as applying such methods to real-world data requires a vast amount of unfiltered data points in order to provide accurate predictions (Tavabi et al. 2020). These techniques are consequently not applicable to our setting due to the lack of sufficient data.

Data and methodology

In addition to the fact that information on cyber risks is typically not publicly available, the systematic collection of known cyber incidents poses further challenges (Eling and Wirfs 2016b). As Romanosky (2016) illustrates, only a fraction of actual cyber incidents is recorded in associated loss databases. A limited number of cyber databases (see Table 1) do exist, mainly established by private and public companies and consortia. Nevertheless, it is challenging to gain access to them, and there is no standard practice in the recording and collection of cyber incidents.

OpRisk databases from the U.S. have been primarily used to model cyber risks in the existing literature, including Advisen (see e.g. Romanosky 2016; Kesan and Zhang 2019; McShane and Nguyen 2020; Palsson et al. 2020) and the SAS OpRisk database (see e.g. Biener et al. 2015; Eling and Wirfs 2016a, 2019). Furthermore, other organisations and consortia collaborate and share data on operational and cyber risks to build systematic databases. Specific databases focusing on data breaches in the U.S. (e.g. Privacy Rights Clearinghouse) and private initiatives (e.g. Hackmagedon) have emerged. However, only some of the above-mentioned initiatives provide information on the economic loss of reported cyber incidents. For our analyses, we use the German Öffentliche Schadenfälle OpRisk database due to the following reasons. First, the database is rather small, which emphasises the introduced motivation of sparse cyber risk modelling. Second, ÖffSchOR focuses on OpRisk losses from the financial sector in Europe, providing some of the first insights both from this important industry in the European Union and from Europe overall. In particular, the size of the recorded losses and relative number of cyber incidents are comparable to previous studies (see e.g. Eling and Wirfs 2019). Third, ÖffSchOR provided free access to the database to conduct this research project and to promote quantitative cyber research. Fourth, and to the best of our knowledge, this is the first scientific analysis based on ÖffSchOR in the context of cyber risk research.³

³ A few other studies have previously used ÖffSchOR, mainly in OpRisk research (see e.g. Sturm 2013; Kaspereit et al. 2017; Eckert et al. 2020).



Table 1 Overview of relevant cyber risk databases

| Database | | Scope | Focus | Sample | Selected references from cyber risk literature |
|---------------|-------------------------------------------------------------|---------------|---------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. Commercial | Advisen | OpRisk | U.S | > 90,000 | Romanosky (2016), Kesan and Zhang (2019), McShane and Nguyen (2020), Palsson et al. (2020) |
| | IBM FIRST Risk Case Studies | OpRisk | Global | > 17,000 | /. |
| | Öffentliche Schadenfälle OpRisk (ÖffSchOR) | OpRisk | DACH | > 3,300 | /. |
| | SAS OpRisk | OpRisk | U.S | > 32,000 | Biener et al. (2015), Eling and Wirfs (2016a), Eling and Wirfs (2019) |
| B. Consortium | Associazione Italiana per la Sicurezza Informatica (Clusit) | Cyber | Global | > 6,800 | Giudici and Raffinetti (2020) |
| | Datenkonsortium OpRisk (DakOR) | OpRisk | Germany | > 38,000 | /. |
| | Italian Database of Operational Losses (DIPO) | OpRisk | Italy | /. | /. |
| | Operational Riskdata eXchange Association (ORX) | OpRisk | Global | > 800,000 | Bouveret (2018) |
| C. Public | Hackmaggdon | Cyber | Global | > 10,000 | /. |
| | Privacy Rights Clearinghouse (PRC) | Data breaches | U.S | > 8,500 | Edwards et al. (2016), Wheatley et al. (2016), Eling and Loperfido (2017), Eling and Jung (2018), Xu et al. (2018), Eling and Wirfs (2019), Jung (2019), Fang et al. (2021), Kamiya et al. (2021), Wheatley et al. (2021) |
| | Vocabulary for Event Recording and Incident Sharing (VERIS) | Data breaches | U.S | > 8,000 | /. |
| | | | | | |

References from OpRisk research are not included in this overview



ÖffSchOR database

ÖffSchOR is an information database on publicly disclosed loss events of operational risks in the financial sector. The database is operated by VÖB-Service GmbH, a subsidiary of the Federal Association of Public Banks (Bundesverband Öffentlicher Banken Deutschlands, VÖB) in Germany. In general, losses of a gross amount of EUR 100,000 or more are recorded in the database, including reputational risks and risk scenarios. The industry focus is on financial services and insurance companies in Europe. In addition, interesting loss events can be examined from other economic sectors or regions. ÖffSchOR uses print and online media services to collect data.

All loss events are categorised according to the Capital Requirements Regulation (CRR) specifications (EU 2013). Loss incidents are assigned to different subcategories, such as conduct risk, legal risk, information and communication technology (ICT) risk or sustainability risk. To date, however, there is no unique identifier for cyber risk in the ÖffSchOR database. Therefore, subcategories distinguishing cyber and non-cyber events are necessary.

Methodology

Motivated by the framework of Fang et al. (2021), the methodology of this study is organised into six key components: (1) data preparation, analysis and transformation; (2) marginal model; (3) modelling frequency; (4) modelling severity; (5) modelling temporal dependence and (6) predicting the next time period.

Data preparation, explorative data analysis and data transformation

As of 30 September 2021, the ÖffSchOR database consists of 3,261 operational loss events between 2002 and 2021. Given that the database does not categorise cyber events, it is first necessary to allocate the sample to cyber and non-cyber incidents. Cyber risk is defined as “any risk emerging from the use of ICT that compromises the confidentiality, availability, or integrity of data or services [...]. Cyber risk is either caused by natural disasters or is man-made where the latter may emerge from human failure, cyber criminality (e.g. extortion, fraud), cyber war or cyber terrorism” (Eling et al. 2016). Based on this definition, which has been suggested as the most comprehensive in the cyber risk literature (Strupczewski 2021), Tables 2 and 3 present the search strategy employed to identify 341 cyber events in the ÖffSchOR database. The strategy combines both systematic and manual search steps to maximise and validate the categorisation of cyber events. In order to gain preliminary insights from the data, a descriptive analysis is conducted.

The dataset is then transformed into a time series, where y_{it} is the amount of all cyber losses of company i in year t , n is the number of companies in the data and T is the time horizon. Non-cyber incidents and companies without a single cyber



Table 2 Search and identification of cyber events in the ÖffSchOR database and remaining data points (in bold)

| Step | Task | Data points |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| 0 | Extraction of ÖffSchOR database as of 30 September 2021 | 3,261 |
| 1 | Systematic and manual search of cyber events according to the definition of Eling et al. (2016) by identifying cyber keywords in the event description, according to Table 3 | (– 2898) 363 |
| 2 | Manual review of <i>all</i> tagged cyber incidents in terms of validity and consistency including recategorisation, if necessary | (– 38) 325 |
| 3 | Manual review of <i>randomly selected</i> non-cyber incidents in terms of validity and consistency including recategorisation, if necessary | (+ 21) 341 |

Values in italics used to highlight the difference or add-on of data points



Table 3 Unique keywords (incl. frequency) to identify cyber events in the ÖffSchOR database

| Threats | Vulnerability | Assets |
|------------------------|--------------------|---------------------|
| Hacker (120) | Data breach (39) | Customer data (23) |
| Skimming (98) | Data security (23) | Smartphone (5) |
| Data theft (87) | Security gap (15) | Bitcoin (4) |
| Cyber (60) | pushTAN/mTAN (10) | Security system (3) |
| Phishing (51) | E-mail (7) | Mobile (2) |
| Stealing data (10) | Password (3) | |
| DDoS (7) | PIN (1) | |
| IT attack (2) | | |
| Social engineering (1) | | |
| Spam (1) | | |

incident are not considered in the analysis. If several companies are affected by an event, the reported loss figure is equally distributed. Further, if a company experiences several cyber incidents in one year, the total amount is accumulated. This transformation results in a multivariate time series with many instances in which $y_{it} = 0$, indicating that many companies, i , did not suffer any cyber loss in year t .

Marginal model

In the following, let the random variable Y_{it} represent the cyber loss of company i in year t . Thus, the distribution function, F_{it} , of Y_{it} , and the corresponding density function, f_{it} , can be described as follows (Shi and Yang 2018; Fang et al. 2021):

$$\begin{aligned} F_{it}(y) &= p_{it} + (1 - p_{it})M_{it}(y), \\ f_{it}(y) &= p_{it}I(y = 0) + (1 - p_{it})m_{it}(y), \end{aligned} \quad (1)$$

where $I(\cdot)$ is the indicator function, p_{it} is the probability that company i experiences *no* cyber incident in year t , M_{it} is the distribution function and m_{it} is the corresponding density function of Y_{it} under the condition that $Y_{it} > 0$ (i.e. a cyber loss has occurred).

Modelling frequency

A logistic regression is performed to determine the probability of occurrence $1 - p_{it}$. As the dataset consists of limited additional information, the temporal trend t and the categories industry (*ind*) and region (*reg*) are included, such that the overall model M equals:



$$\begin{aligned}
\log \left(\frac{p_{it}}{1 - p_{it}} \right) = & \beta_0 + \beta_1 t + \beta_2 t^2 \\
& + (\beta_3, \beta_4, \beta_5)^T X_{ind} + (\beta_6, \beta_7, \beta_8)^T X_{reg} \\
& + (\beta_9, \beta_{10}, \beta_{11})^T X_{ind} \cdot t + (\beta_{12}, \beta_{13}, \beta_{14})^T X_{reg} \cdot t \\
& + (\beta_{15}, \beta_{16}, \beta_{17})^T X_{ind} \cdot t^2 + (\beta_{18}, \beta_{19}, \beta_{20})^T X_{reg} \cdot t^2,
\end{aligned} \tag{2}$$

where $X_{ind} = \{\text{Bank, Municipal bank, Insurance, Other}\}$ and $X_{reg} = \{\text{DACH, Europe, Americas, Other}\}$. The category *other* serves as the base reference. The Results section analyses further model variants. The aim is to obtain a robust and valid estimate of p_{it} while minimising the coefficient β .

Modelling severity

Given that cyber risks are skewed and heavy tailed (Maillart and Sornette 2010; Eling and Wirfs 2016b, 2019; Wheatley et al. 2016; Fang et al. 2021), a mixed-model approach based on EVT is chosen to model the distribution of M_{it} :

$$M_{it}(y | \Theta) = \begin{cases} (1 - \phi_{\mu_i}) H_i(y | \Theta), & y \leq \mu_i, \\ (1 - \phi_{\mu_i}) + \phi_{\mu_i} G_i(y | \Theta), & y > \mu_i, \end{cases} \tag{3}$$

where Θ is the parameter vector and $H_i(y | \Theta)$ and $G_i(y | \Theta)$ are the distribution functions below and above the threshold μ_i , respectively.

This approach offers a high degree of flexibility in the choice of distribution functions. Due to the sparse data and for robust estimation of Θ , a probability distribution with few shape parameters is preferred, especially for the distribution below the threshold. Based on the explorative data analysis and for a simple illustration of the methodology, we set $H_i(y | \Theta) \sim N(\mu_{H_i}, \sigma_{H_i})$ to the normal distribution and $G_i(y | \Theta) \sim GPD(\mu_{G_i}, \sigma_{G_i}, \xi)$ to the generalised Pareto distribution (GPD) while considering the log-transformed losses (Eling and Wirfs 2016b; Eling and Loperfido 2017; Fang et al. 2021). The parameter vector Θ is then estimated numerically using the maximum likelihood estimation (MLE). To further analyse the robustness of the model, the estimation of Θ is performed and evaluated with different timeframes $t = \{t_0, \dots, T^*\}$ with $T^* \in \{T - 5, \dots, T\}$.

Modelling temporal dependence

To model the temporal dependence (i.e. serial trend, between Y_{i1}, \dots, Y_{iT}), we use a copula structure such that

$$F_i(\mathbf{y}) = C(F_{i1}(y_1), \dots, F_{iT}(y_T)), \tag{4}$$

where $\mathbf{y} = (y_1, \dots, y_T)$, F_{i1}, \dots, F_{iT} are the marginal distributions from Eq. (1) and C is the copula. A large variety of copulas could be used. As our focus is on modelling



the serial trend in a time series setting, the D-vine copula provides a good fit—offering flexibility and efficiency and incorporating the temporal structure of time series data (Shi and Yang 2018; Fang et al. 2021).

Considering the random variable $Y_i = (Y_{i1}, \dots, Y_{iT})$ of company i , the joint density of its cyber losses can be expressed as follows:

$$f_i(\mathbf{y}) = \prod_{i=1}^T f_{it}(y_t) \prod_{t=2}^T \prod_{s=1}^{t-1} \tilde{f}_{i,s,t|(s+1):(t-1)}(y_s, y_t | y_{(s+1):(t-1)}), \quad (5)$$

with $\mathbf{y} = (y_1, \dots, y_T)$, $f_{it}(y_t)$ from Eq. (1) and where $\tilde{f}_{i,s,t|(s+1):(t-1)}(y_s, y_t | y_{(s+1):(t-1)})$ is the ratio of bivariate distribution to the product of the marginals (see Equations (A.1) – (A.3) in the Appendix and Smith (2015), Shi and Yang (2018), and Fang et al. (2021) for technical details). To model the multivariate time series, the log-likelihood function (LL) equals

$$LL(\mathbf{y}_1, \dots, \mathbf{y}_n) = \sum_{i=1}^n \sum_{t=1}^T \log f_{it}(y_{it}) + \sum_{i=1}^n \sum_{t=2}^T \sum_{s=1}^{t-1} \log (\tilde{f}_{i,s,t|(s+1):(t-1)}(y_{is}, y_{it} | y_{i,(s+1):(t-1)})), \quad (6)$$

where $\mathbf{y}_k = (y_{k1}, \dots, y_{kT})$. To estimate the model's parameters, the two-stage inference functions for margins (IFM) approach is used, which are practical for predictive applications and computationally efficient (Joe 2005). Applying the sequential approach (Shi and Yang 2018), we estimate and fix the dependence structure for each tree by selecting the bivariate copula with the lowest Akaike information criterion (AIC). Starting from the first tree, we estimate the next tree using the estimates of the previous tree(s). For practical reasons, we also fix the dependence structure in each tree, which leads to the same copula within one tree but can differ between trees.

Due to the flexible bivariate, pairwise copula construction, a variety of copulas can be used. With regard to the limited data, the following one-parametric copulas are considered for the pairwise copula construction: the independence copula, the Gaussian copula, the Frank copula, the Joe copula, the Clayton copula and the Gumbel copula. Special attention is dedicated to the Frank copula, which can represent both positive and negative dependence structures and has already been successfully applied in comparable scientific investigations (Fang et al. 2021; Kularatne et al. 2021):

$$c(u_1, u_2) = -\eta^{-1} \log \left(1 + \frac{(e^{-\eta u_1} - 1)(e^{-\eta u_2} - 1)}{e^{-\eta} - 1} \right), \quad (7)$$

with $\eta \neq 0$ being the copula parameter. Further information on the other copulas can be found in Nelsen (2006).

Predicting the next time period

The final aim is to predict the probability of an occurrence and the economic impact of a potential cyber event on company i for the future time period $T + 1$. In the first step, the probability of occurrence $1 - p_{i,T+1}$ of a cyber event for company i is



estimated. In the second step, the loss amount $Y_{i,T+1}$ is determined under the condition that a cyber event has occurred. Given historical loss data $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})$ for company i , the conditional density of $Y_{i,t+1} | \mathbf{y}_i$ can be expressed as follows:

$$f_{i,t+1|1:t}(\mathbf{y}) = f_{i,t+1}(\mathbf{y}) \sum_{s=2}^t \tilde{f}_{i,s,t+1|(s+1):t}(\mathbf{y}_s, \mathbf{y} | \mathbf{y}_{(s+1):t}), \quad (8)$$

with $i \in \{1, \dots, n\}$. This method allows prediction of the cyber loss distribution of entity i one step ahead of time. As there is no closed formula for $Y_{i,T+1}$, a Monte Carlo simulation based on the rejection sampling method is used for the prediction (Robert and Casella 2004; Fang et al. 2021). The validity and goodness of fit of the predicted model are measured by the ranked probability score (RPS), which is a commonly used accuracy measure (Epstein 1969; Gneiting and Raftery 2007). Due to the nature of the (mixed-model) approach, standard measures such as mean absolute error (MAE) and mean standard error (MSE) are not applicable.

Results

Data preparation, explorative data analysis and data transformation

Since the ÖffSchOR database does not consist of any identifier for cyber events, all 3,261 data points are first categorised into cyber and non-cyber events. Based on the cyber definition of Eling et al. (2016) and the iterative search and identification presented in Tables 2 and 3, 22 unique keywords for cyber risks are identified and further categorised into threats (10 keywords), vulnerabilities (7) and risk objects (5) in line with Böhme et al. (2019). The most frequently identified keywords from the *threats* category are hacker (120), skimming (98), data theft (87), cyber (60) and phishing (51). The most frequently identified keywords from the *vulnerabilities* category are data breach (39) and data security (23), and the most frequently identified keyword from the *assets* category is customer data (23). In total, 341 cyber (10%) and 2,920 non-cyber events are determined in the database and manually checked for correct categorisation.

Table 4 summarises the descriptive analysis of the ÖffSchOR database, split into four different panels. According to Panel A, cyber risks differ from non-cyber (operational) risks. Cyber risks have a lower average loss severity and skewness. For example, the average loss of a cyber event is EUR 17.4 million, while a non-cyber event costs on average EUR 210.9 million. In terms of skewness, the 95% quantile of cyber events is approximately EUR 82 million and the maximum is more than 10 times that amount at EUR 877 million. For non-cyber events, the 95% quantile is almost 10 times as large as the same figure for cyber events, at almost EUR 800 million. The maximum for non-cyber events of EUR 24.6 billion corresponds to a 30-fold multiplier between the maximum and the 95% quantile.

Panel B focuses solely on cyber risks and demonstrates a detailed split regarding the event origin. Almost three quarters (74%) of all cyber incidents have an external origin, which also correspond to the highest loss amounts (95% quantile:



Table 4 Descriptive analysis of ÖffSchOR database regarding the loss severity (in EUR million)

| | <i>N</i> | % | Ø | Min | Quantiles | | | | Max |
|------------------------------------------------------------------------|----------|----|-------|-----|-----------|-----|------|-------|----------|
| | | | | | 25% | 50% | 75% | 95% | |
| <i>Panel A: Cyber vs. non-cyber</i> | | | | | | | | | |
| Cyber | 341 | 10 | 17.4 | 0.0 | 0.2 | 0.5 | 2.0 | 82.0 | 877.0 |
| Non-cyber | 2920 | 90 | 210.9 | 0.0 | 0.3 | 2.0 | 38.6 | 799.8 | 24,643.0 |
| <i>Panel B: Cyber risks only—risk origin (according to CRR)</i> | | | | | | | | | |
| External | 251 | 74 | 18.3 | 0.0 | 0.2 | 0.4 | 1.9 | 86.3 | 877.0 |
| Internal processes | 11 | 3 | 5.3 | 0.1 | 0.6 | 1.0 | 2.2 | 24.5 | 39.0 |
| Human failure | 43 | 13 | 19.1 | 0.2 | 0.5 | 1.5 | 13.2 | 77.8 | 149.0 |
| System failure | 36 | 11 | 6.7 | 0.7 | 0.9 | 1.3 | 7.5 | 23.5 | 30.3 |
| <i>Panel C: Cyber risks only—regional distribution</i> | | | | | | | | | |
| DACH | 209 | 61 | 2.7 | 0.0 | 0.1 | 0.3 | 0.7 | 6.9 | 149.0 |
| Europe | 52 | 15 | 8.6 | 0.1 | 0.8 | 2.0 | 5.4 | 41.4 | 75.0 |
| Americas | 55 | 16 | 60.3 | 0.1 | 0.8 | 4.0 | 61.5 | 306.5 | 624.2 |
| Other | 25 | 7 | 75.0 | 0.1 | 0.6 | 5.2 | 34.0 | 326.2 | 877.0 |
| <i>Panel D: Cyber risks only—single vs. multiple entities affected</i> | | | | | | | | | |
| Single | 236 | 69 | 11.2 | 0.0 | 0.1 | 0.4 | 2.0 | 39.8 | 624.2 |
| Multiple | 105 | 31 | 30.0 | 0.1 | 0.3 | 0.5 | 2.0 | 123.4 | 877.0 |

EUR 82 million, maximum: EUR 877 million). The human factor (13%) is a considerable source of risk—with similarly high figures. Cyber events caused by internal processes (3%) and systems (11%) exhibit a lower average loss, but a comparable median of EUR 1–1.3 million.

Regarding the regional distribution of cyber incidents (Panel C), more than two thirds of all cases are reported in Germany, Austria and Switzerland (DACH) (61%) and Europe (15%). Although a minority of events originates from the Americas (16%) and the rest of the world (7%), these events demonstrate a much higher loss value, both on average and in the quantiles. For example, the average loss amount in the Americas and the rest of the world is between EUR 60–75 million, and the 95% quantile is around EUR 300–330 million. In comparison, the average loss amount in DACH is EUR 2.7 million and the 95% quantile is EUR 6.9 million. One could, therefore, conclude that cyber losses in DACH and Europe are mild in relative terms. However, the focus of the ÖffSchOR database is on loss data from Germany and Europe, which is why non-European losses are recorded from a relatively high absolute loss figure.

Finally, Panel D indicates that more than one company is affected in one third of all cyber incidents. In these cases, the average loss amount as well as the 95% quantile are approximately three times higher compared to a singular (e.g. one company is affected) incident.

In summary, cyber and non-cyber risks must be distinguished and separately modelled. In particular, external cyber events have historically corresponded to the highest loss amounts and almost every third cyber event has affected multiple



entities. Nevertheless, non-cyber risks statistically exhibit higher mean values and skewness compared to cyber risks.

To conclude, the cyber data subset needs to be transformed into a time series. Further limitations of the dataset arise due to the incompleteness of the data and restrictions of the time horizon. Specific loss amount information is available for only 207 of the 341 cyber incidents. As the dataset is already very small, all 341 cyber incidents are used to model the frequency according to Eq. (2), and the remaining 207 data points are used thereafter—in particular for modelling the severity. If several companies are affected by a cyber incident, the loss amount is distributed equally among all companies. The time horizon corresponds to one year, which is why losses during the year are aggregated at the annual level. The relevant timeframe is set to $t_0 = 2005$ and $T = 2018$, meaning that we predict the likelihood and severity of an enterprise-level cyber incident for year $T + 1 = 2019$.⁴ In total, the time series consists of $n = 275$ companies for the modelling of frequency and $n = 184$ companies thereafter, over a time span of 14 years.

Modelling frequency

To estimate the probability of a cyber event $1 - p_{it}$, a logistic regression is applied according to the base model M from Eq. (2). In addition, three model variants are considered: $M1$ (without quadratic interaction terms, i.e. $\beta_{15}, \dots, \beta_{20} = 0$), $M2$ (without regional dummy variable X_{reg} , i.e. $\beta_6, \dots, \beta_8, \beta_{12}, \dots, \beta_{14}, \beta_{18}, \dots, \beta_{20} = 0$) and $M3$ (without quadratic interaction terms and without regional dummy variable, thus, $\beta_3, \dots, \beta_5, \beta_{12}, \dots, \beta_{20} = 0$). Table 5 summarises the regression results.

First, all parameters of models $M1$ and $M3$ are significant at $p \leq 0.05$, which do not hold for M and $M2$. In particular, the regional dummy variables X_{reg} are only significant with the temporal interaction term t . Furthermore, we find that $\beta_3, \dots, \beta_5 > 0$ and $\beta_9, \dots, \beta_{11} > 0$, implying an initially positive and subsequently negative non-linear trend for the dummy variable X_{ind} . The reverse trend (first negative, then positive) is observed for the regional dummy variable due to $\beta_6, \dots, \beta_8 < 0$ and $\beta_{12}, \dots, \beta_{14} > 0$.

Comparing the four models, M exhibits the lowest AIC and highest log-likelihood value (LL) as well as the highest pseudo R^2 (McKelvey and Zavoina 1975) and area under curve (AUC). $M2$ demonstrates only slightly less favourable values for the considered ratios, followed by $M1$ and $M3$. The Hosmer–Lemeshow test (HLT) can be rejected for all models. Thus, in principle, model M could be the most appropriate. However, we choose model $M2$ for the following reasons. First, the ANOVA test confirms that there is no significant difference between models $M1$ and $M3$ with respect to $M2$ ($p < 0.01$) such that model $M2$ is preferred over the variants $M1$ and $M3$. Second, M and $M2$ reveal similar goodness-of-fit statistics, with $M2$ having fewer regression parameters. Under the condition of variable reduction, $M2$ is, thus, selected as the preferred model. The adequacy and accuracy of model $M2$

⁴ For $t < 2005$ and $t > 2019$, a limited number of complete data points are observed.



Table 5 Results of the logistic regression

| | M | | M1 | | M2 | | M3 | |
|----------------|--------|----------|--------|----------|--------|----------|--------|----------|
| | Est | SD | Est | SD | Est | SD | Est | SD |
| β_0 | 3.724 | 1.475* | 3.639 | 0.706*** | 0.429 | 0.427 | 1.623 | 0.329*** |
| β_1 | −0.430 | 0.479 | −0.659 | 0.146*** | 0.386 | 0.208 | −0.340 | 0.107** |
| β_2 | 0.038 | 0.037 | 0.076 | 0.009*** | −0.006 | 0.020 | 0.073 | 0.009*** |
| β_3 | 3.372 | 0.666*** | 2.471 | 0.436*** | 4.047 | 0.603*** | 2.841 | 0.409*** |
| β_4 | 4.501 | 1.135*** | 2.663 | 0.564*** | 5.898 | 1.022*** | 2.937 | 0.496*** |
| β_5 | 14.025 | 6.635* | 7.488 | 1.436*** | 14.518 | 6.497* | 7.632 | 1.416*** |
| β_6 | −1.473 | 1.495 | −1.604 | 0.630* | − | − | − | − |
| β_7 | −3.834 | 1.533* | −1.742 | 0.704* | − | − | − | − |
| β_8 | −4.124 | 1.468** | −2.330 | 0.658*** | − | − | − | − |
| β_9 | −1.095 | 0.277*** | −0.528 | 0.091*** | −1.309 | 0.254*** | −0.579 | 0.087*** |
| β_{10} | −1.452 | 0.410*** | −0.577 | 0.105*** | −1.979 | 0.367*** | −0.607 | 0.096*** |
| β_{11} | −3.139 | 1.605 | −1.169 | 0.181*** | −3.352 | 1.571* | −1.185 | 0.178*** |
| β_{12} | 0.129 | 0.459 | 0.239 | 0.086** | − | − | − | − |
| β_{13} | 1.058 | 0.480* | 0.269 | 0.098** | − | − | − | − |
| β_{14} | 1.086 | 0.461* | 0.338 | 0.095*** | − | − | − | − |
| β_{15} | 0.064 | 0.025* | − | − | 0.079 | 0.023*** | − | − |
| β_{16} | 0.087 | 0.034* | − | − | 0.130 | 0.031*** | − | − |
| β_{17} | 0.147 | 0.096 | − | − | 0.166 | 0.094 | − | − |
| β_{18} | 0.012 | 0.034 | − | − | − | − | − | − |
| β_{19} | −0.062 | 0.035 | − | − | − | − | − | − |
| β_{20} | −0.063 | 0.034 | − | − | − | − | − | − |
| AIC | 1903.3 | | 1926.1 | | 1917.3 | | 1930.5 | |
| LL | −930.6 | | −948.1 | | −946.6 | | −956.3 | |
| R ² | 0.38 | | 0.30 | | 0.34 | | 0.28 | |



Table 5 (continued)

| | M | | M1 | | M2 | | M3 | |
|-----|------|----|------|----|------|----|------|----|
| | Est | SD | Est | SD | Est | SD | Est | SD |
| AUC | 0.72 | | 0.71 | | 0.70 | | 0.69 | |
| HLT | 13.0 | | 26.1 | | 17.5 | | 18.7 | |

Est: Estimated parameter, SD: Standard deviation, AIC: Akaike information criterion, LL: Log-likelihood, R^2 : Pseudo-measure according to McKelvey and Zavoina (1975), AUC: Area under curve of the ROC curve, HLT: Hosmer–Lemeshow test, ***: p value ≤ 0.001 , **: p value $\in (0.001;0.01]$, *: p value $\in (0.01;0.05]$



Table 6 Mean absolute error (MAE) and mean squared error (MSE) of the four regression models

| | MAE | | | | | MSE | | | | |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Σ | B | MB | I | O | Σ | B | MB | I | O |
| M | 2.1% | 2.9% | 4.0% | 6.5% | 4.0% | 0.1% | 0.1% | 0.3% | 1.0% | 0.3% |
| M1 | 1.8% | 2.8% | 4.4% | 7.0% | 5.6% | 0.1% | 0.1% | 0.3% | 1.2% | 0.6% |
| M2 | 2.0% | 2.8% | 4.0% | 6.4% | 4.1% | 0.1% | 0.1% | 0.3% | 1.0% | 0.3% |
| M3 | 1.8% | 2.8% | 4.5% | 7.0% | 5.6% | 0.1% | 0.1% | 0.3% | 1.2% | 0.6% |

Σ : Total, B: Bank, MB: Municipal bank, I: Insurance, O: Other

Bold highlights the selected model- Model M2

are further assessed in terms of MAE and MSE, as reflected in Table 6. Both the MAE and the MSE are low, with 2.0% and 0.1% respectively. Furthermore, *M2* has the lowest MAE and MSE regarding the subcategories *municipal bank* (MB) and *insurance* (I). Hence, the adequacy and accuracy of model *M2* can be sufficiently confirmed. Based on *M2*, the probability $1 - p_{i,T+1}$ of a cyber event for company *i* will be predicted.

Modelling severity

We subsequently model the severity of cyber losses according to the proposed mixed model from Eq. (3). As described previously, only 207 cyber losses with a loss amount $y_{it} > 0$ are used in the following analysis. Due to the very small dataset and the fact that six parameters of the vector $\Theta = \{\mu, \sigma_\mu, \xi, \phi_\mu, \mu_G, \sigma_G\}$ need to be estimated, separated modelling by subindustry—analogous to the probability of occurrence—cannot be conducted in order to ensure convergence and robust estimation.

Figure 1 depicts the plotted log-transformed cyber losses and the fitted mixed model, which exhibits a good overall fit to the data. In greater detail, Table 7 presents the estimated values and standard deviations of the parameter vector Θ for $T = 2018$ and the estimation results with a truncated period from $t_0 = 2005$ to $T = 2013, \dots, 2017$. The truncated analysis is performed to affirm the overall robustness of the model due to data scarcity.

For $T = 2018$, we observe a log threshold $\mu = 7.12$, which nominally is equal to EUR 13.2 million ($10^{7.12}$). Regarding the normal distribution below the threshold, the log-expected value is $\mu_G = 5.62$ (nominally EUR 417,000) with a standard deviation of $\log(\sigma_G) = 0.67$. The scaling parameter σ_μ of the generalised Pareto distribution is equal to 0.06 and the shape parameter $\xi = 1.56$. With $\phi_\mu = 0.21$, approximately every fifth cyber loss is above the threshold value μ . Furthermore, Table 7 presents that there are no significant changes in the estimated parameters while truncating the time period, indicating a very robust estimation of the parameter vector Θ and a robust mixed-model approach.



Fig. 1 Histogram of the log-transformed cyber losses and plot of the estimated mixed model (red line)

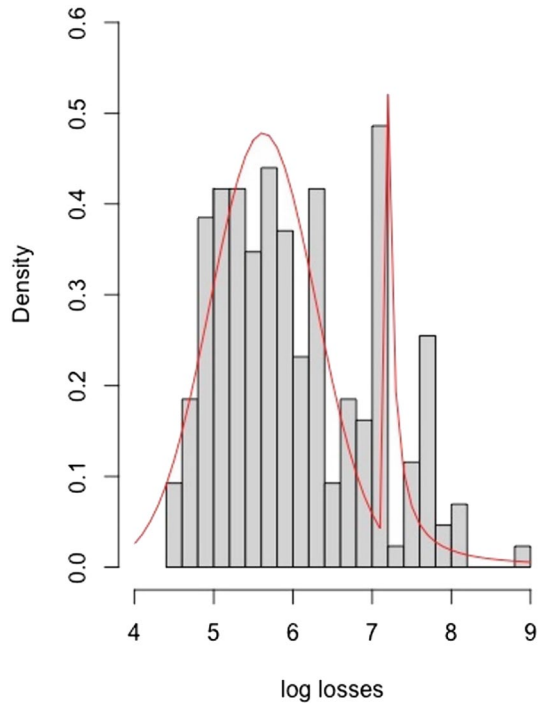


Table 7 Estimated parameter values and standard deviations (SD) of Θ while using different time intervals $t_0 = 2005$ and $T = 2013, \dots, 2018$

| | μ | $SD(\mu)$ | σ_μ | $SD(\sigma_\mu)$ | ξ | $SD(\xi)$ | ϕ_μ | $SD(\phi_\mu)$ | μ_G | $SD(\mu_G)$ | σ_G | $SD(\sigma_G)$ |
|-------------|-------------|-------------|--------------|------------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|----------------|
| 2013 | 7.12 | 0.00 | 0.02 | 0.01 | 1.29 | 0.32 | 0.21 | 0.04 | 5.56 | 0.06 | 0.63 | 0.05 |
| 2014 | 7.12 | 0.00 | 0.03 | 0.01 | 1.56 | 0.46 | 0.17 | 0.03 | 5.50 | 0.05 | 0.60 | 0.04 |
| 2015 | 7.12 | 0.00 | 0.05 | 0.02 | 1.59 | 0.44 | 0.19 | 0.03 | 5.58 | 0.06 | 0.67 | 0.05 |
| 2016 | 7.12 | 0.00 | 0.08 | 0.04 | 1.45 | 0.56 | 0.21 | 0.03 | 5.58 | 0.05 | 0.66 | 0.04 |
| 2017 | 7.12 | 0.00 | 0.08 | 0.04 | 1.47 | 0.54 | 0.20 | 0.03 | 5.61 | 0.05 | 0.66 | 0.04 |
| 2018 | 7.12 | 0.00 | 0.06 | 0.03 | 1.56 | 0.46 | 0.21 | 0.03 | 5.62 | 0.05 | 0.67 | 0.04 |

[bold] highlights the selected year - 2018 is selected

Modelling temporal dependence

Following the methodology, we next model the serial trend based on the D-vine copula from Eq. (5). Regarding the pair-copula construction, six bivariate copulas are considered $\Omega = \{\text{Independent, Gaussian, Clayton, Frank, Gumbel, Joe}\}$. Given the time frame $T - t_0 = 13$, a maximum of 13 trees could be estimated. However, due to the sparse information on serial trends in the database, we decide to limit the estimation to five trees, meaning that the temporal dependence of the last six years is



taken into account in the copula model. For each tree Tr_1, \dots, Tr_5 , the bivariate linking copula with the lowest AIC is chosen.

As reflected in the results in Table 8 (Panel A), the Frank copula demonstrates the lowest AIC for all trees, which is why the Frank copula is selected to represent the pairwise serial trend. It is important to note that for the Gumbel and Joe copula $\hat{\eta} \approx 1$ and for the Clayton copula $\hat{\eta} \approx 0$ regarding the trees Tr_1, \dots, Tr_5 , suggesting very little to no temporal dependence. However, the log-likelihood and AIC are significantly less favourable in comparison to the Frank and Gaussian copula.

Panel B provides the estimated parameter value, $\hat{\eta}$, of the selected bivariate linking copula (Frank), its standard deviation and the Kendall rank correlation coefficient τ for Tr_1, \dots, Tr_5 . The parameter $\hat{\eta}$ of the Frank copula is negative for all trees, indicating a negative temporal dependence. This result suggests that if a company has not yet experienced a cyber loss, it is relatively likely that a loss will occur in the next time period. However, if there has been a previous cyber incident, it is relatively unlikely that another cyber incident will occur within the next five years. This negative dependence may be a result of the fact that (external) attackers are not interested in breaching the same company twice. In the aftermath of an attack, companies tend to close security gaps and invest in their cyber risk management (Kamiya et al. 2021).

Further assessment of the goodness of fit reveals that the RPS of the mixed D-vine with Frank is the lowest at 0.219, followed by the mixed D-vine with Gauss (0.251) and independence copula (0.311). Therefore, the mixed D-vine with Frank provides the best fit and is chosen to represent the serial trend.

Predicting the next time period

Finally, we predict the frequency and severity of an enterprise-level cyber event with respect to the next time period $T + 1 = 2019$. Table 9 summarises the key statistical values derived from the distribution $Y_{i,T+1} | y_i$ for randomly selected companies in the four industry categories. Values for the maximum and tail value at risk (TVaR) are not presented because the shape parameter $\xi > 1$ (i.e. we deal with infinite mean models with extreme uncertainties in very high quantiles; see e.g. Chavez-Demoulin et al. 2016; Eling and Wirfs 2019).

Regarding the probability of occurrence $1 - p_{i,T+1}$, the chance of a cyber incident in the next year is predicted to be 0.6% for a selected municipal bank, 2.0% for a bank, 8.7% for an insurer and 1.4% for any other financial services company. Under the condition that a cyber incident does occur in the next year, the minimum loss value is estimated to be around EUR 1,000–3,000, while the median loss ranges between EUR 455,000 (*insurance*) and EUR 585,000 (*other*). With respect to value at risk (VaR) measures, the VaR(90%) is equal to EUR 14.6 million for the municipal bank and EUR 15.4 million for the selected bank, while the VaR(95%) is approximately EUR 18.5–22.3 million. At a higher confidence level, the VaR(99%) ranges from EUR 69.5 million (*municipal bank*) to EUR 543 million (*insurance*). Even more extreme values are observed for the VaR(99.5%), indicating a worst-case incident that can cause the collapse of a company.



Table 8 Results of the linking copula selection for the five-dimensional D-vine structure

| | Tr_1 | | | Tr_2 | | | Tr_3 | | | Tr_4 | | | Tr_5 | | |
|-----------------------------------------|---------------|--------------|--|---------------|--------------|--|--------------|--------------|--|-------------|--------------|--|-------------|--------------|--|
| | LL | AIC | | LL | AIC | | LL | AIC | | LL | AIC | | LL | AIC | |
| <i>A. Copula selection</i> | | | | | | | | | | | | | | | |
| Independence | -27.34 | 56.68 | | -16.84 | 35.67 | | -7.97 | 17.93 | | -0.16 | 2.32 | | 2.45 | 2.90 | |
| Gaussian | -18.73 | 39.45 | | -17.92 | 37.85 | | -6.79 | 15.57 | | 1.71 | -1.43 | | 3.23 | -4.46 | |
| Clayton | -27.34 | 56.67 | | -16.83 | 35.66 | | -7.96 | 17.93 | | -0.16 | 2.32 | | 2.45 | -2.90 | |
| Frank | -18.30 | 38.60 | | -15.96 | 34.93 | | -5.59 | 13.17 | | 2.67 | -3.34 | | 4.28 | -6.56 | |
| Gumbel | -27.34 | 56.68 | | -16.84 | 35.67 | | -7.94 | 17.87 | | -0.13 | 2.25 | | 2.97 | -3.93 | |
| Joe | -27.34 | 56.68 | | -16.84 | 35.67 | | -7.88 | 17.76 | | -0.11 | 2.22 | | 3.06 | -4.12 | |
| <i>B. Statistics of selected copula</i> | | | | | | | | | | | | | | | |
| $\hat{\eta}$ | -2.2565 | | | -0.9591 | | | -1.4439 | | | -1.6284 | | | -1.1577 | | |
| SD | 0.5588 | | | 0.5103 | | | 0.5890 | | | 0.5890 | | | 0.7529 | | |
| Kendall τ | -0.2383 | | | -0.1054 | | | -0.1566 | | | -0.1759 | | | -0.1268 | | |

Bold highlights the selected parameter - Frank copula is selected



Table 9 Predicted probability of occurrence $1 - p_{i,T+1}$ and statistical values (in EUR thousand) of the distribution $Y_{i,T+1} | Y_i > 0$ for randomly selected companies

| | Bank | Municipal bank | Insurance | Other |
|-----------------|---------|----------------|-----------|---------|
| $1 - p_{i,T+1}$ | 2.0% | 0.6% | 8.7% | 1.4% |
| Minimum | 1 | 1 | 1 | 3 |
| 25% quantile | 182 | 123 | 147 | 205 |
| 50% quantile | 533 | 465 | 455 | 585 |
| 75% quantile | 2,218 | 2,033 | 1,707 | 2,679 |
| VaR(90%) | 15,373 | 14,611 | 14,244 | 14,971 |
| VaR(95%) | 22,309 | 18,446 | 18,957 | 20,326 |
| VaR(99%) | 351,255 | 69,496 | 542,616 | 292,680 |
| VaR(99.5%) | 490,671 | 107,435 | 1,902,275 | 720,306 |

Discussion and conclusion

This study provides new insights on the empirical nature and prediction of cyber risks at the enterprise level under data scarcity. We introduced the ÖffSchOR database to cyber risk research and applied advanced modelling techniques adapted from the work of Shi and Yang (2018), Eling and Wirfs (2019) and Fang et al. (2021) to predict the frequency, severity and serial trend of enterprise cyber risks. Our findings first suggest that cyber risks are indeed different from operational risks. In particular, we found that cyber risks are lower on average, less skewed and less extreme compared to non-cyber risks in the dataset (Biener et al. 2015; Woods and Böhme 2021). Second, the industry subcategories exhibited different probabilities of occurrence, a finding which has not yet been addressed in previous studies. Third, in modelling the impact of the log-transformed cyber incidents, the POT model with a normal distribution below the threshold demonstrated a satisfying fit, which is in line with previous empirical results and supports the differentiation of daily and extreme cyber risks (Eling and Loperfido 2017; Eling and Wirfs 2019). Due to the limited data, a separate loss modelling for each subcategory was not possible. By leveraging the serial dependence, the D-vine copula was able to predict the impact of a potential cyber incident in the next time period with a negative correlation over time (Fang et al. 2021). The prediction results provide some of the first quantitative insights on the financial impact of a cyber incident at the enterprise level based on historic data.

Our results underline that high-level descriptive statistics from commercial datasets might be misleading for enterprise risk managers due to information asymmetry and interdependence of loss events (Eling and Wirfs 2016a; Marotta et al. 2017; Zeller and Scherer 2021). In particular, our model predicted a median enterprise-level loss amount of EUR 455,000–585,000, only a fraction of the millions of dollars often cited in surveys (e.g. USD 3.86 million; IBM Security 2020). In a U.K. survey, the maximum loss is around GBP 310,000 (~EUR 370,000; Heitzenrater and Simpson 2016), while Romanosky (2016) estimates the average data breach loss to be even lower, at USD 200,000 (~EUR 180,000), bearing in mind that data breaches only account for 25% of cyber events and that the transfer from data



breaches to actual costs is misrepresentative (Eling and Wirfs 2019). Moreover, the estimated extreme losses of VaR(99%) and VaR(99.5%) can be compared to mega breaches such as those of Home Depot (USD 340 million), Anthem (USD 407 million) and Yahoo (USD 502 million) in the U.S. (Poyraz et al. 2020) or to General Data Protection Regulation (GDPR) fines on Whatsapp (EUR 225 million) and Amazon (EUR 745 million) in Europe (CNPD 2021; EDPB 2021). The estimated VaR(95%) can be interpreted as the lower limit of a GDPR penalty, at a minimum of EUR 20 million (Poyraz et al. 2020). This information supports the impression that our estimates are reasonable in size.

Furthermore, our findings suggest that cyber risks are less heavy tailed than often anticipated. For example, Eling and Wirfs (2019) simulate VaR measures for a small bank with 5,000 employees, which is comparable to our *municipal bank* category. Our estimated figures are significantly lower, such as EUR 18.5 million vs. EUR 48 million (USD 55 million) for the VaR (95%) and EUR 69.5 million vs. EUR 422 million (USD 480 million) for the VaR (99%). One conclusion from these findings is that cyber risks are just not that harmful (Woods and Böhme 2021). Another reason suggested by practitioners is that attackers have focused on easier targets while the financial services industry is comparably well protected due to regulated risk management and anti-money laundering systems. However, cyber risks are still heavy tailed and extreme. With every fifth cyber incident above the threshold of EUR 13.2 million, there is still a (small) chance of a devastating cyber event seriously harming an individual company (Eling et al. 2016; Wheatley et al. 2021).

Comparing the four subcategories, our findings imply that bigger banks suffer from a higher potential loss than smaller (municipal) banks, indicating that the loss amount might be correlated to the company size (i.e. revenue or number of employees; Poyraz et al. 2020). Furthermore, the selected insurance company exhibited a four-times greater chance of a cyber incident, with the highest estimated risk measures for VaR (99%) and VaR (99.5%). Similar heavy tails were observed for the category *other* consisting of payment providers, stock exchanges and other financial services providers, which seems plausible due to their high interconnectivity to other companies.

In practice, most risk and expert assessments are solely qualitative due to the limited data available on cyber incidents. For example, the Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) provided one of the first frameworks to identify and manage information security risks by analysing a company's asset, threat and vulnerability information (Alberts et al. 1999). Further information security risk assessment (ISRA) methods have been developed, with the Core Unified Risk Framework (CURF) being the most comprehensive and all-inclusive approach (Wangen et al. 2018). A specific cyber risk classification framework named Quantitative Bow-tie (QBowTie) has been suggested by Sheehan et al. (2021) combining proactive and reactive barriers to reduce a company's risk exposure and quantify the risk. However, all these (qualitative) methods are generally based on the assessment of probability of occurrence and of the associated consequence of an event, i.e. requiring a quantification of the (cyber) risk.

Compared to that, our analysis provides a helpful tool in the ongoing quantification of cyber risks. Nevertheless, the method also comes with limitations. First,



researchers have argued that the rapidly changing cyber risk environment may render historic data useless (CRO Forum 2014; Eling and Schnell 2016). Given ongoing digitisation and in times of a global pandemic, the usefulness of historic data can be questioned. A further limitation is the assumption of independence between entities. Particularly for extreme cyber risks and mega breaches, there is a high correlation between companies (Biener et al. 2015). However, our framework could be extended to model both the serial and cross-company dependence, as conceptually shown by Acar et al. (2019) and Zhao et al. (2020) for dense data. Further limitations arise due to the use of the ÖffSchOR database. In particular, ÖffSchOR relies on print and online media to detect operational risk events which could bias the recorded loss events and in turn the modelling results. The latter could also be influenced by the historical (log-normal) distribution of the cyber loss severity. Furthermore, the total number of identified cyber events is rather small compared to other studies within the research (e.g. 1,579 cyber incidents are analysed by Eling and Wirfs 2019), challenging the robustness of our results. Finally, due to the limited dataset, we did not distinguish between different cyber risks or loss categories even though different types of cyber risks follow different distributions (Eling and Loperfido 2017; Eling and Jung 2018) and cyber risks do not only cause economic losses, but also intangible losses, including reputational damage (Xie et al. 2020).

Despite these limitations and the dynamic nature of cyber risks (Boyer 2020), this study contributes to the literature on cyber risk measurement and can help practitioners such as risk managers, insurers and policymakers by providing a quantitative and data-driven cyber risk assessment. Insurance stakeholders particularly face a major challenge in assessing and understanding cyber risk due to the lack of historical data (Cremer et al. 2022). We believe that the provided methodology could be combined and integrated with existing pricing tools and factors from cyber insurers to better evaluate cyber risk and the required risk-based premiums at the enterprise level (Nurse et al. 2020).

There are plenty of future research opportunities to further develop quantitative approaches. With better and more data, more accurate models can be designed, for example by including both cyber incident data and corporate financial data as proposed by Palsson et al. (2020) or by using network models as seen in Fahrenwaldt et al. (2018), Jevtić and Lanchier (2020) and Wu et al. (2021). The integration of different approaches from diverse disciplines poses extensive future opportunities in the field of cyber risk measurement (Falco et al. 2019).

Appendix

Additional formulas

See Fang et al. (2021), Shi and Yang (2018), and Smith (2015) for further technical details.



$$\begin{aligned} & \tilde{f}_{i,s,t|(s+1):(t-1)}(y_s, y_t | y_{(s+1):(t-1)}) \\ &= \begin{cases} \frac{C_{s,t|(s+1):(t-1)}(F_{is|(s+1):(t-1)}(0|y_{(s+1):(t-1)}), F_{it|(s+1):(t-1)}(0|y_{(s+1):(t-1)}))}{F_{is|(s+1):(t-1)}(0|y_{(s+1):(t-1)})F_{it|(s+1):(t-1)}(0|y_{(s+1):(t-1)})}, & y_s = 0, y_t = 0, \\ \frac{C_{1,s,t|(s+1):(t-1)}(F_{is|(s+1):(t-1)}(y_s|y_{(s+1):(t-1)}), F_{it|(s+1):(t-1)}(0|y_{(s+1):(t-1)}))}{F_{it|(s+1):(t-1)}(0|y_{(s+1):(t-1)})}, & y_s > 0, y_t = 0, \\ \frac{C_{2,s,t|(s+1):(t-1)}(F_{is|(s+1):(t-1)}(0|y_{(s+1):(t-1)}), F_{it|(s+1):(t-1)}(y_t|y_{(s+1):(t-1)}))}{F_{is|(s+1):(t-1)}(0|y_{(s+1):(t-1)})}, & y_s = 0, y_t > 0, \\ C_{s,t|(s+1):(t-1)}(F_{is|(s+1):(t-1)}(y_s|y_{(s+1):(t-1)}), F_{it|(s+1):(t-1)}(y_t|y_{(s+1):(t-1)})), & y_s > 0, y_t > 0 \end{cases} \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} & F_{is|(s+1):(t-1)}(y_s | y_{(s+1):(t-1)}) \\ &= \begin{cases} \frac{C_{s,t-1|(s+1):(t-2)}(F_{is|(s+1):(t-2)}(y_s|y_{(s+1):(t-2)}), F_{it-1|(s+1):(t-2)}(0|y_{(s+1):(t-2)}))}{F_{it-1|(s+1):(t-2)}(0|y_{(s+1):(t-2)})}, & y_{t-1} = 0, \\ C_{2,s,t-1|(s+1):(t-2)}(F_{is|(s+1):(t-2)}(y_s|y_{(s+1):(t-2)}), F_{it-1|(s+1):(t-2)}(y_{t-1}|y_{(s+1):(t-2)})), & y_{t-1} > 0. \end{cases} \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} & F_{it|(s+1):(t-1)}(y_t | y_{(s+1):(t-1)}) \\ &= \begin{cases} \frac{C_{t,s+1|(s+2):(t-1)}(F_{it|(s+2):(t-1)}(y_t|y_{(s+2):(t-1)}), F_{is+1|(s+2):(t-1)}(0|y_{(s+2):(t-1)}))}{F_{is+1|(s+2):(t-1)}(0|y_{(s+2):(t-1)})}, & y_{s+1} = 0, \\ C_{2,t,s+1|(s+2):(t-1)}(F_{it|(s+2):(t-1)}(y_t|y_{(s+2):(t-1)}), F_{is+1|(s+2):(t-1)}(y_{s+1}|y_{(s+2):(t-1)})), & y_{s+1} > 0. \end{cases} \end{aligned} \quad (\text{A.3})$$

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44 (2): 182–198. <https://doi.org/10.1016/j.insmatheco.2007.02.001>.
- Acar, Elif F., Claudia Czado, and Martin Lysy. 2019. Flexible dynamic vine copula models for multivariate time series data. *Econometrics and Statistics* 12: 181–197. <https://doi.org/10.1016/j.ecosta.2019.03.002>.



- Alberts, Christopher J., Sandra G. Behrens, Richard D. Pethia, and William R. Wilson. 1999. Operationally critical threat, asset, and vulnerability evaluation (OCTAVE) Framework, Version 1.0. Fort Belvoir, VA.
- Aldasoro, Iñaki, Leonardo Gambacorta, Paolo Giudici, and Thomas Leach. 2020. The drivers of cyber risk. *BIS Working Papers No 865*. <https://www.bis.org/publ/work865.pdf>. Accessed May 20, 2021
- Ashby, Simon, Trevor Buck, Stephanie Nöth-Zahn, and Thomas Peisl. 2018. Emerging IT risks: insights from German banking. *The Geneva Papers on Risk and Insurance — Issues and Practice* 43 (2): 180–207. <https://doi.org/10.1057/s41288-018-0081-8>.
- Bedford, Tim, and Roger M. Cooke. 2002. Vines: a new graphical model for dependent random variables. *The Annals of Statistics* 30 (4): 1031–1068.
- Bendovschi, Andreea. 2015. Cyber-attacks—trends, patterns and security countermeasures. *Procedia Economics and Finance* 28: 24–31. [https://doi.org/10.1016/S2212-5671\(15\)01077-1](https://doi.org/10.1016/S2212-5671(15)01077-1).
- Biener, Christian, Martin Eling, and Jan Wirfs. 2015. Insurability of cyber risk: an empirical analysis. *The Geneva Papers on Risk and Insurance — Issues and Practice* 40 (1): 131–158. <https://doi.org/10.1057/gpp.2014.19>.
- Böhme, Rainer, and Gaurav Kataria. 2006. Models and measures for correlation in cyber-insurance. workshop on the economics of information security (WEIS). <https://core.ac.uk/download/pdf/162458449.pdf>. Accessed February 11, 2021
- Böhme, Rainer, Stefan Laube, and Markus Riek. 2019. A fundamental approach to cyber risk analysis. *Casualty Actuarial Society* 12 (2): 161–185.
- Bouveret, Antoine. 2018. Cyber risk for the financial sector: a framework for quantitative assessment. *IMF Working Papers No. 143*. <https://doi.org/10.5089/9781484360750.001>.
- Boyer, M.M. 2020. Cyber insurance demand, supply, contracts and cases. *The Geneva Papers on Risk and Insurance — Issues and Practice* 45 (4): 559–563. <https://doi.org/10.1057/s41288-020-00188-1>.
- Chavez-Demoulin, Valérie, Paul Embrechts, and Marius Hofert. 2016. An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance* 83 (3): 735–776. <https://doi.org/10.1111/jori.12059>.
- Choudhry, Umar. 2014. *Der Cyber-Versicherungsmarkt in Deutschland: Eine Einführung*. Aufl. 2014. essentials. Wiesbaden: Springer Gabler.
- Commission Nationale Pour La Protection Des Données (CNPD). 2021. Decision Regarding Amazon Europe Core S.À R.L. <https://cnpd.public.lu/en/actualites/international/2021/08/decision-amazon-2.html>. Accessed February 17, 2022
- Cox, Jr., and Louis Anthony. 2012. Evaluating and improving risk formulas for allocating limited budgets to expensive risk-reduction opportunities. *Risk Analysis* 32 (7): 1244–1252. <https://doi.org/10.1111/j.1539-6924.2011.01735.x>.
- Cremer, Frank, Barry Sheehan, Michael Fortmann, Arash N. Kia, Martin Mullins, Finbarr Murphy, and Stefan Materne. 2022. Cyber risk and cybersecurity: a systematic review of data availability. *The Geneva Papers on Risk and Insurance — Issues and Practice* 47 (3): 698–736. <https://doi.org/10.1057/s41288-022-00266-6>.
- CRO Forum. 2014. Cyber resilience—the cyber risk challenge and the role of insurance. <https://www.thecroforum.org/wp-content/uploads/2015/01/Cyber-Risk-Paper-version-24-1.pdf>. Accessed April 01, 2021
- de Smidt, Guido, and Wouter Botzen. 2018. Perceptions of corporate cyber risks and insurance decision-making. *The Geneva Papers on Risk and Insurance — Issues and Practice* 43 (2): 239–274. <https://doi.org/10.1057/s41288-018-0082-7>.
- Eckert, Christian, Nadine Gatzert, and Dinah Heidinger. 2020. Empirically assessing and modeling spillover effects from operational risk events in the insurance industry. *Insurance Mathematics and Economics* 93: 72–83. <https://doi.org/10.1016/j.insmathco.2020.04.003>.
- Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. 2016. Hype and heavy tails: a closer look at data breaches. *Journal of Cybersecurity* 2 (1): 3–14. <https://doi.org/10.1093/cybsec/tyw003>.
- Eling, Martin. 2018. Cyber risk and cyber risk insurance: Status Quo and future research. *The Geneva Papers on Risk and Insurance — Issues and Practice* 43 (2): 175–179. <https://doi.org/10.1057/s41288-018-0083-6>.
- Eling, Martin. 2020. Cyber risk research in business and actuarial science. *European Actuarial Journal* 10 (2): 303–333. <https://doi.org/10.1007/s13385-020-00250-1>.
- Eling, Martin, and Kwangmin Jung. 2018. Copula approaches for modeling cross-sectional dependence of data breach losses. *Insurance: Mathematics and Economics* 82: 167–180. <https://doi.org/10.1016/j.insmathco.2018.07.003>.



- Eling, Martin, and Nicola Loperfido. 2017. Data breaches: goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics* 75: 126–136. <https://doi.org/10.1016/j.insmatheco.2017.05.008>.
- Eling, Martin, and Werner Schnell. 2016. What do we know about cyber risk and cyber risk insurance? *The Journal of Risk Finance* 17 (5): 474–491. <https://doi.org/10.1108/JRF-09-2016-0122>.
- Eling, Martin, Werner Schnell, and Fabian Sommerrock. 2016. Ten key questions on cyber risk and cyber risk insurance. The Geneva Association. https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/cyber-risk-10_key_questions.pdf. Accessed April 06, 2021
- Eling, Martin, and Jan H. Wirfs. 2016a. *Cyber Risk: Too Big to Insure? Risk Transfer Options for a Mercurial Risk Class*. I.VW HSG SchriftenreiheUR, no. 59: Verlag Institut für Versicherungswirtschaft der Universität St. Gallen, St. <http://hdl.handle.net/10419/226644>. Accessed April 06, 2021
- Eling, Martin, and Jan H. Wirfs. 2016b. Modelling and management of cyber risk. Working Paper. <http://www.actuaries.org/oslo2015/papers/iaals-wirfs&eling.pdf>. Accessed April 05, 2021
- Eling, Martin, and Jan Wirfs. 2019. What are the actual costs of cyber risk events? *European Journal of Operational Research* 272 (3): 1109–1119. <https://doi.org/10.1016/j.ejor.2018.07.021>.
- Epstein, Edward S. 1969. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology* 8 (6): 985–987.
- European Data Protection Board (EDPB). 2021. Binding decision 1/2021 on the dispute arisen on the draft decision of the Irish supervisory authority regarding Whatsapp Ireland under Article 65(1)(A) GDPR: EDPB. https://edpb.europa.eu/system/files/2021-09/edpb_bindingdecision_202101_ie_sa_whatsapp_redacted_en.pdf. Accessed February 17, 2022
- European Union (EU). 2013. Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and amending Regulation (EU) No 648/2012 (Text with EEA relevance). <http://data.europa.eu/eli/reg/2013/575/2022-07-08>. Accessed September 20, 2022
- European Union (EU). 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on 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) (Text with EEA Relevance): EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679>. Accessed February 14, 2022
- Fahrenwaldt, Matthias A., Stefan Weber, and Kerstin Weske. 2018. Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin* 48 (3): 1175–1218. <https://doi.org/10.1017/asb.2018.23>.
- Falco, Gregory, Martin Eling, Danielle Jablanski, Matthias Weber, Virginia Miller, Lawrence A. Gordon, Shaun S. Wang, et al. 2019. Cyber risk research impeded by disciplinary barriers. *Science* 366 (6469): 1066–1069. <https://doi.org/10.1126/science.aaz4795>.
- Fang, Zijian, Xu. Maochao, Xu. Shouhuai, and Hu. Taizhong. 2021. A framework for predicting data breach risk: leveraging dependence to cope with sparsity. *IEEE Transactions on Information Forensics and Security* 16: 2186–2201. <https://doi.org/10.1109/TIFS.2021.3051804>.
- Giudici, Paolo, and Emanuela Raffinetti. 2020. Cyber risk ordering with rank-based statistical models. *ASta Advances in Statistical Analysis*. <https://doi.org/10.1007/s10182-020-00387-0>.
- Gneiting, Tilmann, and Adrian E. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102 (477): 359–378. <https://doi.org/10.1198/016214506000001437>.
- Heitzenrater, Chad D., and Andrew C. Simpson. 2016. Policy, statistics and questions: reflections on UK cyber security disclosures. *Journal of Cybersecurity* 2 (1): 43–56. <https://doi.org/10.1093/cybsec/tyw008>.
- Herath, Hemantha S. B., and Tejaswini C. Herath. 2011. Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies: Analyses and Actuarial Computations* 2 (1).
- IBM Security. 2020. Cost of a Data Breach Report 2020. <https://www.ibm.com/security/data-breach>. Accessed May 25, 2021
- Jevtić, Petar, and Nicolas Lanchier. 2020. Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based LAN topology. *Insurance Mathematics and Economics* 91: 209–223. <https://doi.org/10.1016/j.insmatheco.2020.02.005>.
- Joe, Harry. 1997. *Multivariate models and multivariate dependence concepts*. New York: Chapman and Hall/CRC.
- Joe, Harry. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* 94 (2): 401–419. <https://doi.org/10.1016/j.jmva.2004.06.003>.



- Jung, Kwangmin. 2019. Probable maximum cyber loss: empirical estimation and reinsurance design with private-public partnership. 2019 German Insurance Science Association (DVfVW) annual meeting. Berlin.
- Kamiya, Shinichi, Jun-Koo. Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz. 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics* 139 (3): 719–749. <https://doi.org/10.1016/j.jfineco.2019.05.019>.
- Kasperer, Thomas, Kerstin Lopatta, Suren Pakhchanyan, and Jörg. Prokop. 2017. Systemic operational risk: spillover effects of large operational losses in the European banking industry. *The Journal of Risk Finance* 18 (3): 252–267. <https://doi.org/10.1108/JRF-11-2016-0141>.
- Kesan, Jay P., and Linfeng Zhang. 2019. Analysis of cyber incident categories based on losses. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3489436>.
- Kularatne, Thilini D., Jackie Li, and David Pitt. 2021. On the use of archimedean copulas for insurance modelling. *Annals of Actuarial Science* 15 (1): 57–81. <https://doi.org/10.1017/S1748499520000147>.
- Kurowicz, Dorota, and Roger Cooke. 2006. *Uncertainty analysis with high dimensional dependence modelling*. Wiley series in probability and statistics. Chichester: Wiley.
- Layton, Robert, and Paul A. Watters. 2014. A methodology for estimating the tangible cost of data breaches. *Journal of Information Security and Applications* 19 (6): 321–330. <https://doi.org/10.1016/j.jisa.2014.10.012>.
- MacKenzie, Cameron A. 2014. Summarizing risk using risk measures and risk indices. *Risk Analysis* 34 (12): 2143–2162. <https://doi.org/10.1111/risa.12220>.
- Maillart, T., and D. Sornette. 2010. Heavy-tailed distribution of cyber-risks. *The European Physical Journal B* 75 (3): 357–364. <https://doi.org/10.1140/epjb/e2010-00120-8>.
- Marotta, Angelica, Fabio Martinelli, Stefano Nanni, Albina Orlando, and Artsiom Yautsiukhin. 2017. Cyber-insurance survey. *Computer Science Review* 24: 35–61. <https://doi.org/10.1016/j.cosrev.2017.01.001>.
- Marotta, Angelica, and Michael McShane. 2018. Integrating a proactive technique into a holistic cyber risk management approach: a holistic cyber risk management approach. *Risk Management and Insurance Review* 21: 435–452. <https://doi.org/10.1111/rmir.12109>.
- McAfee. 2020. The hidden costs of cybercrime. <https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hidden-costs-of-cybercrime.pdf>. Accessed April 20, 2021.
- McKelvey, Richard D., and William Zavoina. 1975. A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology* 4 (1): 103–120. <https://doi.org/10.1080/0022250X.1975.9989847>.
- McShane, Michael, and Trung Nguyen. 2020. Time-varying effects of cyberattacks on firm value. *The Geneva Papers on Risk and Insurance — Issues and Practice* 45 (4): 580–615. <https://doi.org/10.1057/s41288-020-00170-x>.
- Mukhopadhyay, Arunabha, Samir Chatterjee, Debashis Saha, Ambuj Mahanti, and Samir K. Sadhukhan. 2013. Cyber-risk decision models: to insure IT or not? *Decision Support Systems* 56: 11–26. <https://doi.org/10.1016/j.dss.2013.04.004>.
- National Conference of State Legislatures (NCSL). 2016. Security breach notification laws. <https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx>. Accessed February 15, 2022
- Nelsen, Roger B. 2006. *An introduction to copulas*. Springer Series in Statistics, 2nd ed. New York: Springer.
- Njegomir, Vladimir, and Boris Marović. 2012. Contemporary trends in the global insurance industry. *Procedia - Social and Behavioral Sciences* 44: 134–142. <https://doi.org/10.1016/j.sbspro.2012.05.013>.
- Nurse, Jason, Louise Axon, Arnau Erola, Ioannis Agrafiotis, Michael Goldsmith, and Sadie Creese. 2020. The data that drives cyber insurance: a study into the underwriting and claims processes. In 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). 15–19 June 2020
- Palsson, Kjartan, Steinn Gudmundsson, and Sachin Shetty. 2020. Analysis of the impact of cyber events for cyber insurance. *The Geneva Papers on Risk and Insurance — Issues and Practice* 45 (4): 564–579. <https://doi.org/10.1057/s41288-020-00171-w>.
- Peng, Chen, Xu. Maochao, Xu. Shouhuai, and Hu. Taizhong. 2016. Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics* 44 (14): 2534–2563. <https://doi.org/10.1080/02664763.2016.1257590>.



- Peng, Chen, Xu. Maochao, Xu. Shouhuai, and Hu. Taizhong. 2018. Modeling multivariate cybersecurity risks. *Journal of Applied Statistics* 45 (15): 2718–2740. <https://doi.org/10.1080/02664763.2018.1436701>.
- Pooser, David M., Mark J. Browne, and Oleksandra Arkhangelska. 2018. Growth in the perception of cyber risk: evidence from U.S. P&C Insurers. *The Geneva Papers on Risk and Insurance — Issues and Practice* 43 (2): 208–223. <https://doi.org/10.1057/s41288-017-0077-9>.
- Poyraz, Omer I., Mustafa Canan, C.A. Michael McShane, and Pinto, and T. S. Cotter. 2020. Cyber assets at risk: monetary impact of U.S. personally identifiable information mega data breaches. *The Geneva Papers on Risk and Insurance — Issues and Practice* 45 (4): 616–638. <https://doi.org/10.1057/s41288-020-00185-4>.
- Rakes, Terry R., Jason K. Deane, and Loren Paul Rees. 2012. IT security planning under uncertainty for high-impact events. *Omega* 40 (1): 79–88. <https://doi.org/10.1016/j.omega.2011.03.008>.
- Robert, Christian P., and George Casella. 2004. *Monte Carlo statistical methods*. New York: Springer, New York.
- Romanosky, Sasha. 2016. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity* 2 (2): 121–135. <https://doi.org/10.1093/cybsec/tyw001>.
- Romanosky, Sasha, Lillian Ablon, Andreas Kuehn, and Therese Jones. 2019. Content analysis of cyber insurance policies: how do carriers price cyber risk? *Journal of Cybersecurity* 5 (1): 1–19. <https://doi.org/10.1093/cybsec/tyz002>.
- Ruan, Keyun. 2017. Introducing cybernomics: a unifying economic framework for measuring cyber risk. *Computers & Security* 65: 77–89. <https://doi.org/10.1016/j.cose.2016.10.009>.
- Sheehan, Barry, Finbarr Murphy, Arash N. Kia, and Ronan Kiely. 2021. A Quantitative Bow-Tie cyber risk classification and assessment framework. *Journal of Risk Research* 24 (12): 1619–1638. <https://doi.org/10.1080/13669877.2021.1900337>.
- Shetty, Sachin, Michael McShane, Linfeng Zhang, Jay P. Kesan, Charles A. Kamhoua, Kevin Kwiat, and Laurent L. Njilla. 2018. Reducing informational disadvantages to improve cyber risk management. *The Geneva Papers on Risk and Insurance — Issues and Practice* 43 (2): 224–238. <https://doi.org/10.1057/s41288-018-0078-3>.
- Shi, Peng, and Lu. Yang. 2018. Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association* 113 (521): 122–133. <https://doi.org/10.1080/01621459.2017.1330692>.
- Smith, Michael S. 2015. Copula modelling of dependence in multivariate time series. *International Journal of Forecasting* 31 (3): 815–833. <https://doi.org/10.1016/j.ijforecast.2014.04.003>.
- Strupczewski, Grzegorz. 2021. Defining cyber risk. *Safety Science* 135: 105143. <https://doi.org/10.1016/j.ssci.2020.105143>.
- Sturm, Philipp. 2013. Operational and reputational risk in the european banking industry: the market reaction to operational risk events. *Journal of Economic Behavior & Organization* 85: 191–206. <https://doi.org/10.1016/j.jebo.2012.04.005>.
- Tavabi, Nazgol, Andres Abeliuk, Negar Mokhberian, Jeremy Abramson, and Kristina Lerman. 2020. Challenges in forecasting malicious events from incomplete data. In *Companion proceedings of the web conference 2020*, edited by Amal E. F. Seghrouchni, 603–10. ACM Digital Library. New York: Association for Computing Machinery.
- Wangen, Gaute, Christoffer Hallstensen, and Einar Snekkenes. 2018. A framework for estimating information security risk assessment method completeness. *International Journal of Information Security* 17 (6): 681–699. <https://doi.org/10.1007/s10207-017-0382-0>.
- Wheatley, Spencer, Annette Hofmann, and Didier Sornette. 2021. Addressing insurance of data breach cyber risks in the catastrophe framework. *The Geneva Papers on Risk and Insurance — Issues and Practice* 46 (1): 53–78. <https://doi.org/10.1057/s41288-020-00163-w>.
- Wheatley, Spencer, Thomas Maillart, and Didier Sornette. 2016. The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*. <https://doi.org/10.1140/epjb/e2015-60754-4>.
- Woods, Daniel W., and Rainer Böhme. 2021. Systematization of knowledge: quantifying cyber risk. *IEEE Symposium on Security & Privacy*. https://informationsecurity.uibk.ac.at/pdfs/WB2020_sok_cyber_risk_snp.pdf. Accessed April 19, 2021.
- World Economic Forum (WEF). 2021. The Global Risks Report 2021: 16th edition. Insight report. http://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf. Accessed May 10, 2021.
- Wrede, Dirk, Thorben Freers, Graf von der Schulenburg, and Johann-Matthias. 2018. Herausforderungen Und Implikationen Für Das Cyber-Risikomanagement Sowie Die Versicherung Von Cyberrisiken



- Eine Empirische Analyse. *Zeitschrift Für Die Gesamte Versicherungswissenschaft* 107 (4): 405–434. <https://doi.org/10.1007/s12297-018-0425-2>.
- Wu, Mingyue Zhang, Jinzhu Luo, Xing Fang, Xu. Maochao, and Peng Zhao. 2021. Modeling multivariate cyber risks: deep learning dating extreme value theory. *Journal of Applied Statistics*. <https://doi.org/10.1080/02664763.2021.1936468>.
- Xie, Xiaoying, Charles Lee, and Martin Eling. 2020. Cyber insurance offering and performance: an analysis of the U.S. cyber insurance market. *The Geneva Papers on Risk and Insurance — Issues and Practice* 45 (4): 690–736. <https://doi.org/10.1057/s41288-020-00176-5>.
- Xu, Maochao, Kristin M. Schweitzer, Raymond M. Bateman, and Xu. Shouhuai. 2018. Modeling and predicting cyber hacking breaches. *IEEE Transactions on Information Forensics and Security* 13 (11): 2856–2871. <https://doi.org/10.1109/TIFS.2018.2834227>.
- Zängerle, Daniel, and Dirk Schiereck. 2022. Cyber risks—from a maze of terms to a uniform terminology. *HMD Praxis Der Wirtschaftsinformatik*. <https://doi.org/10.1365/s40702-022-00888-3>.
- Zeller, Gabriela, and Matthias Scherer. 2021. A comprehensive model for cyber risk based on marked point processes and its application to insurance. *European Actuarial Journal*. <https://doi.org/10.1007/s13385-021-00290-1>.
- Zhao, Zifeng, Peng Shi, and Zhengjun Zhang. 2020. Modeling multivariate time series with copula-linked univariate D-vines. *Journal of Business & Economic Statistics*. <https://doi.org/10.1080/07350015.2020.1859381>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

About the authors

Daniel Zängerle is a doctoral candidate at the Department of Corporate Finance at the Technical University of Darmstadt. He holds Bachelor's and Master's degrees in Mathematics and Management as well as a second Master's in Economics. His research focus lies on cyber risk and cyber risk management.

Dirk Schiereck is Professor and Head of the Corporate Finance department at the Technical University of Darmstadt. His current research focuses on corporate finance, asset management and the digitalisation of the financial services industry, including cyber risk. With his academic experience, he is appointed to several supervisory boards and is a member of the scientific advisory board of the German Investor Relations Association, the German Credit Market Standard e.V. and the German Derivatives Association (DDV).

