




# Bayesian vine copulas for modelling dependence in data breach losses

Jia Liu<sup>1\*</sup>, Jackie Li<sup>2</sup> and Kevin Daly<sup>3</sup>

<sup>1</sup>Research School of Finance, Actuarial Studies & Statistics, College of Business and Economics, Australian National University, Canberra, Australian Capital Territory, Australia; <sup>2</sup>Department of Econometrics and Business Statistics, Monash Business School, Monash University, Melbourne, Victoria, Australia; and <sup>3</sup>Department of Economics, Finance and Property, School of Business, Western Sydney University, Sydney, New South Wales, Australia

\*Corresponding author. E-mail: [jia.liu3@anu.edu.au](mailto:jia.liu3@anu.edu.au)

(Received 08 March 2021; revised 04 January 2022; accepted 05 January 2022; first published online 3 February 2022)

## Abstract

Potential data breach losses represent a significant part of operational risk and can be a serious concern for risk managers and insurers. In this paper, we employ the vine copulas under a Bayesian framework to co-model incidences from different data breach types. A full Bayesian approach can allow one to select both the copulas and margins and estimate their parameters in a coherent fashion. In particular, it can incorporate process, parameter, and model uncertainties, and this is very important for applications in risk management under current regulations. We also conduct a series of sensitivity tests on the Bayesian modelling results. Using two public data sets of data breach losses, we find that the overall dependency structure and tail dependence vary significantly between different types of data breaches. The optimally selected vine structure and pairwise copulas suggest more conservative value-at-risk estimates when compared to the other suboptimal copula models.

**Keywords:** Bayesian vine copula; Tail dependence; Data breach losses

## 1. Introduction

The recent decade has seen phenomenal growth in the market of cyber insurance globally, which offers protection against privacy breaches and adverse cyber incidents. The surge is mainly driven by increasing number of corporates demanding coverage for potential financial losses as a result of various data privacy violations (Herath & Herath, 2011; Kshetri, 2020; Cole & Fier, 2021). Despite a burgeoning market, relatively little is known about the actual loss patterns of such data risks, which is unsettling from an actuarial perspective. Fahrenwaldt et al. (2018) underscored the lack of data as a serious impediment to adequate pricing and reserving for cyber insurance products, and this point was echoed by Xu & Hua (2019). They worked around the issue by developing purely theoretical mathematical models to simulate cyber losses. Nonetheless, it is always paramount to directly analyse the trends and relationships in actual data breach losses as more data become available. In this paper, we examine empirically the distributions of real data breaches of different breach types and the dependency structure between them through a Bayesian vine copula framework.

Regardless of the modelling approach, a consensus amongst the previous studies on cyber losses is that it is imperative to incorporate the dependency structure between different sources of risks or losses for the purpose of proper loss aggregation (e.g. Böhme & Schwartz, 2010; Ögüt et al., 2011; Fahrenwaldt et al., 2018; Eling & Jung, 2018; Xu & Hua, 2019). The dependency modelling

techniques adopted for cyber insurance have evolved from linear correlation such as in Böhme & Kataria (2006) to multivariate copulas (e.g. Xu & Hua, 2019) in recent years. Furthermore, the majority of these studies consider dependencies between different mechanisms underlying the cyber-attacks in a theoretical setup without empirical verification (e.g. Mukhopadhyay *et al.*, 2006, 2013; Xu & Hua, 2019). Eling and Jung (2018) fitted a multi-dimensional vine copula to real-world breach data using a two-step approach of first estimating the margins and then fitting the copula via the R package “VineCopula.” The vine copula offers much more flexibility for describing the dependency structure between multiple random variables compared to traditional multivariate copulas. In the “VineCopula” package, the model identification process is based on the maximum spanning tree approach (for example, see Dißmann *et al.*, 2013) (Package “VineCopula,” <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>).

Building on Eling and Jung (2018)’s work, we further explore vine copula modelling of actual data breaches with a one-step, integrated Bayesian method that is user-friendly to actuaries and practitioners. This approach avoids the potential pitfalls present in the two-step inference function for margins and maximum spanning tree methods, and it can readily be applied in pricing and risk management decision-making as to be demonstrated later in the paper. While our results confirm the asymmetric tail dependence between different types of data breaches as reported by Eling and Jung (2018), we also illustrate the deficiency in using only the point estimates for model parameters when simulating data breaches from an actuarial perspective. Considering the industry’s limited, primal understanding about the joint loss distributions of actual cyber losses as well as reservations in sharing of such data, the Bayesian modelling approach is both more prudent in allowing for uncertainties and more convenient in incorporating proprietary information. As such, our approach provides an appropriate modelling tool for insurers and policymakers involved in the cyber insurance market. Moreover, we discover that the generalised beta type II (GB2) distribution, a distribution that is often used for modelling traditional general insurance claims (e.g. Shi & Yang, 2018), is also appropriate for modelling the marginal distributions of breach data. As a distribution family with almost all of the conventional claim size modelling distributions such as the lognormal, gamma, Weibull, Pareto, and Burr distributions nested under it, this finding carries significant implications for future modelling of data breach events and pricing of cyber insurance.

Joe (1996) first proposed the idea of expressing the joint density function as a product of conditional and unconditional copula densities and marginal densities. As the possible ways of decomposition grow more than exponentially with the number of dimensions, Bedford and Cooke (2002) introduced a graphical structure called R-vine to organise the construction. The word “vine” was used because the dependency structure can be displayed visually resembling the outlook of a grapevine. Aas *et al.* (2009) adopted a sequential maximum likelihood estimation procedure to compute the parameters of each pair copula separately. Dißmann *et al.* (2013) applied the concept of a maximum spanning tree to select the optimal vine structure. Since then, there have been a variety of vine copula applications in financial research (e.g. Brechmann *et al.*, 2013; Geidosch & Fischer, 2016; Fink *et al.*, 2017) and more recently actuarial research (Chuliá *et al.*, 2016; Eling & Jung, 2018; Zhou, 2019).

In this paper, we take an approach different to that in Eling and Jung (2018) and incorporate the vine copula structure into a Bayesian framework for co-modelling incidences from different data breach types. Previous applications of Bayesian vine copulas in other areas (e.g. finance, electricity, marketing) and simulation studies can be found in Smith *et al.* (2010), Min and Czado (2010, 2011), Smith (2011, 2015), Smith and Khaled (2012), and Gruber and Czado (2015, 2018). This Bayesian approach allows a simultaneous estimation of the margins and the copulas, which can help avoid the potential bias under the usual inference function for margins (IFM) method. Moreover, it can be used to integrate all of process, parameter, and model uncertainties in a coherent manner. In particular, we allow for the uncertainty in model selection by setting prior probabilities for the copula candidates and calculating the corresponding posterior probabilities.

This method of copula selection is different to the traditional statistical criteria (e.g. Akaike information criterion, AIC) commonly used in the actuarial literature regarding copulas and provides a new perspective to the conventional process of copula identification. Furthermore, we tailor the Bayesian framework to select between the few best vine copula structures for our breach data. To the best of our knowledge, this paper provides the first attempt to implement simultaneous Bayesian selections of both the vine copula tree structure and the pairwise copulas in an insurance valuation problem. In the following analysis, we also present an important empirical finding that the conventional maximum spanning tree method, despite being the default choice in vine copula applications, does not lead to the most optimal structure for data breach events. This result is new to the literature and has a significant implication on the continual search for the most efficient and effective way to identify the best copula structure in modelling dependencies in insurance and other data.

The paper is organised as follows. Section 2 covers the basics of vine copulas. Section 3 introduces our Bayesian framework for vine copula modelling. Section 4 provides a description of the data breach events collected from a public database. Section 5 presents the simulation results from the Bayesian modelling and an example of application to risk management. Section 6 sets forth the results of an extensive list of sensitivity tests. Section 7 applies the Bayesian models to another data set of data breach losses and discusses the results and implications. Section 8 gives the concluding remarks. The Appendix provides some key equations for the copulas adopted.

## 2. Vine Copulas

Sklar (1959) proved that for any multivariate distribution function, there exists a unique copula function  $C$  that links the marginal distribution functions, given that those marginal functions are continuous. This theory leads to the convenience that the dependency structure and the marginal distributions can be selected separately. There are three specific aspects broadly defining the dependency structure, namely the overall shape, level of association, and tail dependence. For the bivariate case, there are a large number of choices, including the common ones like the Gaussian copula,  $t$  copula, and Archimedean copulas (e.g. Joe, 1997; Nelsen, 1999). For the multivariate case with three or more dimensions, however, there are severe limitations in their initial extensions. Notably, the multivariate Gaussian and  $t$  copulas and the multi-dimensional Archimedean copulas give the same shape of dependence and tail dependence for each pair of random variables. Some recently proposed solutions to overcome this inflexibility include the hierarchical (fully or partially nested) Archimedean copulas (e.g. Okhrin et al., 2013) and the vine copulas (e.g. Joe, 2014; Czado, 2019), the latter of which is adopted here.

Let  $X_1, X_2, \dots, X_d$  be a set of  $d$ -dimensional multivariate random variables. For instance, when  $d = 4$ , the joint density function  $f_{1234}$  can be expressed in terms of conditional and unconditional densities:

$$f_{1234}(x_1, x_2, x_3, x_4) = f_{4|123}(x_4|x_1, x_2, x_3)f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f_1(x_1),$$

in which  $f_1$  is the marginal density of  $X_1$ ,  $f_{2|1}$  is the conditional density of  $X_2|X_1$ ,  $f_{3|12}$  is the conditional density of  $X_3|X_1, X_2$ , and  $f_{4|123}$  is the conditional density of  $X_4|X_1, X_2, X_3$ . Using the concept of conditional probability  $\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$  for two dependent events  $A$  and  $B$ , the conditional densities above can be expressed for example as the following:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2))f_2(x_2),$$

$$f_{3|12}(x_3|x_1, x_2) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)f_{3|2}(x_3|x_2),$$

$$f_{3|2}(x_3|x_2) = c_{23}(F_2(x_2), F_3(x_3))f_3(x_3),$$

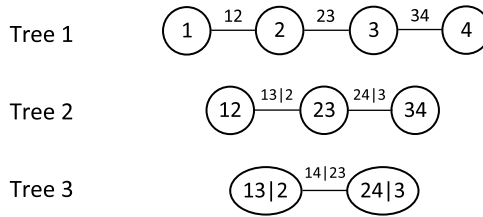


Figure 1. An example of a 4-dimensional D-vine.

$$f_{4|123}(x_4|x_1, x_2, x_3) = c_{14|23}(F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3)|x_2, x_3)f_{4|23}(x_4|x_2, x_3),$$

$$f_{4|23}(x_4|x_2, x_3) = c_{24|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)|x_3)f_{4|3}(x_4|x_3),$$

$$f_{4|3}(x_4|x_3) = c_{34}(F_3(x_3), F_4(x_4))f_4(x_4),$$

where  $c$  refers to a copula density,  $F$  is a distribution function, and the subscripts represent the dimensions and conditions involved. This process is called pair-copula construction (PCC). Note that the joint density function can also be expressed as  $f_{1234}(x_1, x_2, x_3, x_4) = c_{1234}(F_1(x_1), F_2(x_2), F_3(x_3), F_4(x_4))f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4)$ . It can then be deduced that a multivariate copula ( $c_{1234}$ ) can be decomposed into conditional and unconditional bivariate copulas ( $c_{12}, c_{13|2}, c_{23}, c_{14|23}, c_{24|3}, c_{34}$ ). As each bivariate copula can be selected separately from the others, there is a huge amount of flexibility for modelling the entire dependency structure. However, it is also clear that the derivation above is just one specific example, and there are actually numerous ways to decompose the multivariate copula. For higher dimensions, it is extremely tedious to deal with the different possible combinations.

Vines are a flexible graphical method that can be used for building multivariate copulas based on a cascade of bivariate copulas. A  $d$ -dimensional R-vine is a sequence of  $d - 1$  linked layers of trees. Tree 1 has  $d$  nodes  $N_1$  and edges  $E_1$ . Tree 2 has  $(d - 1)$  nodes  $N_2 = E_1$  and edges  $E_2$ , Tree 3 has  $(d - 2)$  nodes  $N_3 = E_2$  and edges  $E_3$ , and so on, i.e. an edge in a tree becomes a node in the next tree. There is a proximity condition such that for any two nodes connected by an edge in a tree, they must represent two edges connected to the same node in the previous tree. Figure 1 provides a graphical vine representation of the 4-dimensional example above. The edges 12, 23, and 34 represent the unconditional bivariate copula densities  $c_{12}, c_{23}$ , and  $c_{34}$ , respectively. The edges 13|2, 24|3, and 14|23 represent the conditional bivariate copula densities  $c_{13|2}, c_{24|3}$ , and  $c_{14|23}$ . In fact, it is a special case of R-vine called D-vine, where no node is connected to more than two edges (i.e. a path structure). Another special case, called C-vine, is shown in Figure 2, in which tree  $i$  has a unique node that is connected to  $(d - i)$  edges (i.e. a star structure). Including these two cases, there are indeed a total of  $d!2^{(d-2)!/(d-4)!/2-1} = 24$  possible combinations. This way of drawing trees of nodes and edges make the decomposition more manageable, when compared to using the first principles to break down the joint density function as above.

The construction of an appropriate R-vine requires making suitable choices on the overall vine structure, the bivariate copulas, and the copula parameters. The state-of-the-art approach in the literature is: (1) compute the copula data using the IFM method; (2) select the spanning tree that maximises the sum of the absolute values of sample Kendall’s tau values across all connected pairs in that tree for each layer; (3) select the bivariate copula with the lowest AIC value for each pair; and (4) use maximum likelihood to estimate the parameters of the selected copulas sequentially. The so-called simplifying assumption is usually taken such that all the bivariate copulas do not

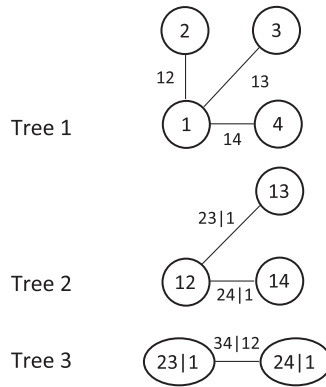


Figure 2. An example of a 4-dimensional C-vine.

depend on the conditioning values directly, though the arguments inside the copulas still depend on the conditioning values<sup>1</sup>. The conditional arguments can be calculated from the corresponding *h*-functions<sup>2</sup>. More details on the estimation process can be found in Aas et al. (2009) and Dißmann et al. (2013).

Despite its practical convenience, this current standard approach has a number of potential drawbacks. Firstly, the pre-estimation of the margins may induce bias in the estimation of the copulas. The underlying covariance between the copula parameter estimators and marginal parameter estimators cannot be calculated. Also, the copula parameters are estimated sequentially, which may cause more bias in the higher trees where the parameters are computed later in the sequence. Secondly, the maximum spanning tree serves as a heuristic method to identify the optimal vine structure amongst numerous possible combinations. The identified structure, however, may not represent the truly optimal structure in probabilistic terms. Thirdly, it would be difficult to incorporate parameter and model uncertainties, as well as other relevant information such as experts’ opinions. Omitting these uncertainties and failing to use prior information could lead to a serious underestimation of the aggregate risk. In the next section, we will introduce our Bayesian framework to implement vine copula modelling. This Bayesian approach provides a sound and practical way to address these potential issues.

### 3. Bayesian Framework

Suppose *X* represents the vector of observations of *d* multivariate random variables (monthly data breach events of the four breach types, i.e. *d* = 4), *X<sub>f</sub>* is the future values,  $\Theta$  denotes all the unknown parameters, and *M* refers to the model choice. Under the Bayesian setting, the joint distribution  $p(X_f, X, \Theta, M)$  is equal to  $p(X_f, X|\Theta, M)p(\Theta|M)p(M)$ . The primary goal is to derive the joint posterior distribution  $p(X_f, \Theta, M|X)$ , given the observations *X*. This joint posterior distribution can be used to deduce a variety of very useful Bayesian inferences via conditioning on and marginalising variables, such as the highest posterior probability of  $p(M|X)$  (i.e. the

<sup>1</sup>For example, the conditional bivariate copula density  $c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)$  can be simplified as  $c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))$ , i.e.  $c_{13|2}$  depends on the conditioning value  $x_2$  only through its arguments  $F_{1|2}(x_1|x_2)$  and  $F_{3|2}(x_3|x_2)$  but not directly through  $x_2$ . As noted in Hobæk Haff et al. (2010), this simplifying assumption is set for purely practical purposes such that inferencing is fast, flexible, tractable, and robust. Though not all multivariate distributions can be represented by such a simplified PCC, it can still provide a rather good approximation.

<sup>2</sup>Two examples of the *h*-functions are  $F_{1|2}(x_1|x_2) = \frac{\partial}{\partial F_2(x_2)} C_{12}(F_1(x_1), F_2(x_2))$  and  $F_{1|23}(x_1|x_2, x_3) = \frac{\partial}{\partial F_{3|2}(x_3|x_2)} C_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))$ , under the simplifying assumption. We have derived some results in the Appendix.

“best” model), the posterior distribution of the unknown parameters  $p(\Theta|X, M)$  given a model (or  $p(\Theta|X)$  averaged across all the model candidates), and the predictive distribution of the future values  $p(X_f|X, M)$  given a model (or  $p(X_f|X)$  averaged across all the model candidates). A comparison between two models  $M_1$  and  $M_2$  can be conducted by examining the posterior odds  $p(M_1|X)/p(M_2|X)$ , which is equal to the Bayes factor  $p(X|M_1)/p(X|M_2)$  times the prior odds  $p(M_1)/p(M_2)$ .

For complex models, it is generally difficult to derive a closed-form expression for the joint posterior distribution. With the rapidly advancing computing power nowadays, one can use Markov chain Monte Carlo (MCMC) simulation to approximate the joint posterior distribution, in which random samples are simulated from a Markov chain with its stationary distribution being equal to the posterior distribution. In this paper, we use the software JAGS (Plummer, 2017) to perform MCMC simulations. They use the Gibbs sampling (GS) method, which involves successive simulations from the full conditional posterior distribution of each variable in turn. It reduces the multi-dimensional simulation from the joint posterior distribution into a sequence of one-dimensional simulations. The programming language is user-friendly and is highly suitable for building various Bayesian models<sup>3</sup>.

In the literature, there is a long list of bivariate copulas to choose from. We consider the Clayton copula and its rotations, Frank, and Gaussian copulas as the major model choices. The Clayton copula function is  $C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$  for  $\theta > 0$ ,  $u_1 = F_1(x_1)$ , and  $u_2 = F_2(x_2)$ , and the 90°, 180°, and 270° rotations can be obtained from  $u_2 - C(1 - u_1, u_2)$ ,  $u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$ , and  $u_1 - C(u_1, 1 - u_2)$ , respectively. It has lower tail dependence<sup>4</sup> (i.e. in the lower-left quadrant), and its rotations have tail dependence in the lower-right, upper-right, and upper-left quadrants. The Frank copula function is given as  $C(u_1, u_2) = -\ln(1 + \frac{\exp(-\theta u_1) - 1}{\exp(-\theta) - 1} \frac{\exp(-\theta u_2) - 1}{\exp(-\theta) - 1})/\theta$  for  $-\infty < \theta < \infty$ . It has no tail dependence and is symmetric over all the four quadrants. The Gaussian copula function is expressed as  $C(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ , in which  $\Phi_2$  is the distribution function of the standard bivariate normal distribution with association parameter  $-1 < \theta < 1$  and  $\Phi^{-1}$  is the inverse distribution function of the standard normal. Like the Frank copula, it has no tail dependence and is symmetric. This list of copulas covers a variety of features which would be suitable for different dependency structures in the data.

To make a Bayesian selection between the model candidates for each pair of random variables, we express the effective copula density as a weighted average of those of the individual models, i.e.  $c(u_1, u_2) = \sum w_j c_j(u_1, u_2)$ , in which the weight  $w_j$  is equal to one when the corresponding copula density  $c_j$  is chosen and is zero otherwise. The highest posterior probability of  $p(M|X)$  (reflected by the highest  $E(w_j|X)$ ) would then indicate the most optimal model from a Bayesian perspective. The predictive distribution of the future values can be obtained from either  $p(X_f|X, M)$  or  $p(X_f|X)$ , depending on the purpose of the analysis. The former (Bayesian model selection, BMS) can be used when the purpose is to identify a single model which provides key insights into the nature of the data or to heuristically collect a few promising models for further examination, while the latter (Bayesian model averaging, BMA) can be adopted if one wants to incorporate more model uncertainty into the prediction. Note that when the data patterns are complex, the model with the highest posterior probability from a finite list of models may not sufficiently represent the true underlying mechanism, and accordingly, one may prefer to use the model averaging approach (over all the models covered or just a few selected models).

Regarding the prior  $p(M)$ , a popular choice for model selection is the discrete uniform prior. It is an uninformative prior in the sense that all the model candidates are being favoured equally

<sup>3</sup>The JAGS code for implementing a Bayesian vine copula structure can be provided upon request. If greater flexibility and more specific features are needed, one may develop his or her own MCMC algorithm (e.g. Min & Czado, 2010, 2011).

<sup>4</sup>Lower tail dependence is defined as  $\lambda_L = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}$ . It measures the association in the lower-left-quadrant tail. There is lower tail dependence if  $0 < \lambda_L \leq 1$  and no lower tail dependence if  $\lambda_L = 0$ .



**Table 1.** Summary statistics of PRC monthly data breach events from 2005 to 2019.

Type of Data Breach	Average	Standard Deviation
H	49,321,500	280,911,769
P	300,773	779,670
E	1,403,474	7,327,805
D	1,401,309	7,855,950

before the analysis. Since we do not have other information relevant for the data breach events, we employ this uninformative prior in our study. For the prior  $p(\Theta|M)$ , we assume uninformative uniform priors for the copula parameters and uninformative gamma priors for the marginal parameters<sup>5</sup>. In the absence of subjective inputs or experts' opinions, setting uninformative priors would minimise the influence of the prior assumptions. Moreover, we assume that the priors are independent with one another, as it is often difficult to understand how they are associated a priori.

#### 4. Data

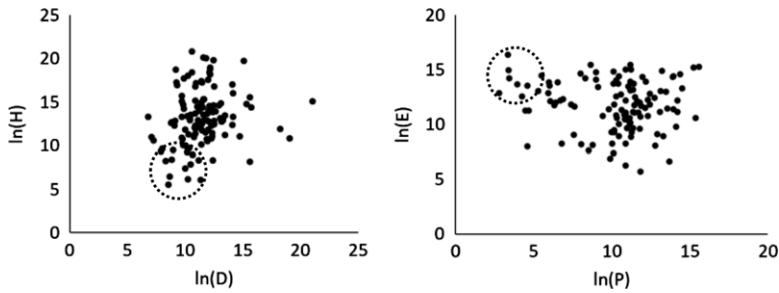
We collect 9,015 data breach events from 2005 to 2019 from the Privacy Rights Clearinghouse (PRC, <https://privacyrights.org/>) and focus on the non-zero monthly events of four breach types including hacking/malware (H), loss of physical records of data (P), loss of electronic devices containing sensitive data (E), and other unintended disclosure (D)<sup>6</sup>. A frequently used database for empirical cyber risk studies (e.g. Edwards et al., 2016; Eling & Loperfido, 2017; Eling & Jung, 2018), the PRC database stores information of publicly verifiable breach events since 2005 that encroach on the privacy of citizens of the United States. The loss resulting from such incidents is measured by the number of personal records being breached. The breach may occur in any type of organisation or business entity ranging from government bodies to listed companies<sup>7</sup>. Currently, it is the most comprehensive and biggest data breach database that is free of charge and accessible to the general public. For the purpose of this analysis, we sum up the losses arising from different incidents that are under the same breach type and have happened in the same month and fit our Bayesian model to these monthly aggregated losses. As shown in Table 1, the standard deviations of the losses far exceed the means of the losses. It highlights the critical importance of assessing adequately the level of uncertainty of data breach losses for risk management purposes.

Figure 3 displays the relationships between the D and H losses and between the P and E losses in logarithmic scale. There is a positive relationship ( $\tau = 0.18$ ) between the H and D losses, which

<sup>5</sup>In Section 5, for all the copula and marginal parameters, the prior variances are set as large as possible, as long as the resulting range is valid theoretically and also the JAGS simulation process runs smoothly. The priors can be taken as, say, Uniform(0, 5) for the Clayton copula parameter, Uniform(-5, 5) for the Frank copula parameter, Uniform(-0.9, 0.9) for the Gaussian copula parameter, and Gamma( $\alpha, \beta$ ) for the marginal parameters, in which  $\alpha$  and  $\beta$  are chosen such that the prior standard deviation is several times larger than the prior mean. Note that if the prior variance is too large, however, the JAGS program may clash, and in those cases, one can reduce the prior variance accordingly and may also set the prior mean closer to certain preliminary estimate such as the method of moments estimate and maximum likelihood estimate in order to avoid this programming issue.

<sup>6</sup>These four categories constitute a refined categorization on the original five known data breach types in the database where portable devices and stationary computer loss are combined to form a new category focusing on loss of electronic devices (E). These data account for around 85% of all the non-zero loss data breach events covered in the database. Eling & Jung (2018) studied the hacking, loss of electronic devices, unintended disclosure, and insider attack (instead) breach types from 2005 to 2016 only. They applied the usual sequential estimation method mentioned in Section 2 when fitting the vine copulas and used standard statistical criteria to identify the bivariate copulas and the marginal distributions.

<sup>7</sup>The PRC database has grouped the organizations into seven categories, and Eling & Jung (2018) analyzed the breaches by industry but found generally weaker dependence compared to breach type.



**Figure 3.** The left graph plots the hacking/malware (H) losses against other unintended disclosure (D) losses; the right graph plots the losses of electronic devices containing sensitive data (E) against the losses of physical records of data (P). The losses are in logarithmic scale.

seem to have some form of lower tail dependence (as circled in the graph; non-parametric estimate of 0.34; e.g. Frahm *et al.*, 2005). By contrast, there is a negative relationship ( $\tau = -0.04$ ) between the P and E losses, with some tail dependence in the upper-left quadrant (non-parametric estimate of 0.18). These patterns clearly call for different copula functions to be used for the two pairs of random variables and highlight the importance of having the flexibility offered by the vine copula structure under which a specific bivariate copula can be selected for each pair of random variables. For instance, the association between the H and D losses looks like what the Clayton copula can describe, while the association between the P and E losses may be modelled by a suitable rotation of the Clayton copula. Note that these are only very preliminary observations, as the final optimal copula choices depend on formal Bayesian inferencing (Section 5) and also the tree level involved. The variables in the higher trees are conditioned on a larger number of other variables, which would then dilute or distort some of the original patterns in the raw data.

**4.1. GB2 as marginal distribution**

The losses of each breach type appear to have a right-skewed and long-tailed distribution. For the margins, we use the generalised beta type II (GB2) distribution (e.g. Chan *et al.*, 2018). The GB2 density function is

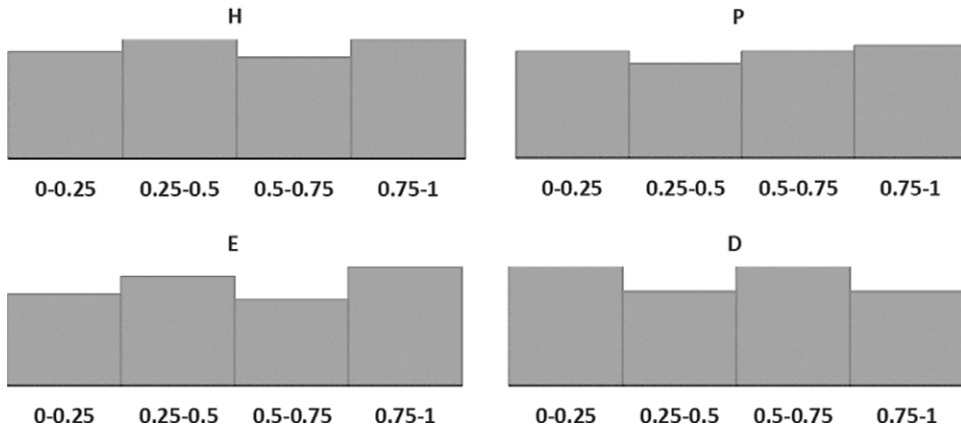
$$f(x) = \frac{\Gamma(p + q)}{\Gamma(p)\Gamma(q)} \frac{\left(\frac{a}{b}\left(\frac{x}{b}\right)\right)^{ap-1}}{\left(1 + \left(\frac{x}{b}\right)^a\right)^{p+q}},$$

in which  $x > 0$ ,  $-\infty < a < \infty$ ,  $b > 0$ ,  $p > 0$ , and  $q > 0$ . The distribution function can be expressed as

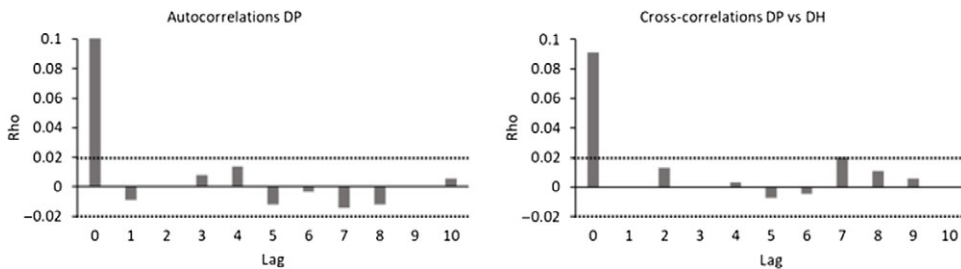
$$F(x) = F_{\text{Beta}}\left(\frac{\left(\frac{x}{b}\right)^a}{1 + \left(\frac{x}{b}\right)^a}\right),$$

where  $F_{\text{Beta}}$  is the distribution function of the Beta( $p, q$ ) distribution. The GB2 distribution nests a range of commonly used loss distributions as special cases, such as gamma ( $a = 1, b = \beta q, q \rightarrow \infty$ ), Weibull ( $b = \beta q^{1/a}, p = 1, q \rightarrow \infty$ ), lognormal ( $a \rightarrow 0, b = (a\sigma)^{2/a} q^{1/a}, p = \frac{a\mu + 1}{(a\sigma)^2}, q \rightarrow \infty$ ), and Pareto ( $a = 1, p = 1$ ). In contrast to the usual frequentist approach of testing a finite set of traditional loss distributions on the data, incorporating the GB2 distribution into the Bayesian framework can take into account the uncertainty in selecting the model distribution, as the GB2 distribution itself generalises a big family of important distributions. In this way, model uncertainty can be integrated coherently within the Bayesian framework and there is no need to choose one (amongst gamma, Weibull, lognormal, Pareto etc.) explicitly based on a certain goodness-of-fit test.





**Figure 4.** Histograms of copula data from fitted GB2 distributions under IFM method (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure). The parameter estimates are  $a = 0.65, b = 6.84, p = 1.61, q = 0.64$  for H,  $a = 1.00, b = 26.79, p = 0.44, q = 1.20$  for P,  $a = 0.28, b = 22.30, p = 4.78, q = 5.69$  for E, and  $a = 1.60, b = 9.40, p = 0.31, q = 0.37$  for D.

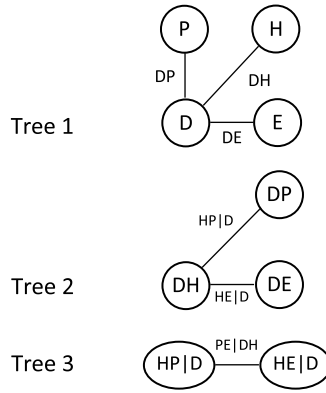


**Figure 5.** Autocorrelations of selected copula's parameter between D and P (left) and cross-correlations between selected copulas' parameters of DP and DH (right) (H: hacking/malware; P: loss of physical records of data; D: other unintended disclosure).

As a preliminary analysis, we fit the GB2 distribution to each breach type separately (using maximum likelihood as a starting point). All the resulting chi-square test values for testing the goodness-of-fit are found to be not statistically significant ( $p$ -values of 0.13 (H), 0.10 (P), 0.57 (E), 0.29 (D)) and so the null hypothesis is not rejected. Moreover, Figure 4 shows that the histograms of the copula data (between 0 and 1) from the fitted GB2 distributions under the IFM method look fairly like the uniform distribution, which again suggests that the GB2 is a reasonable distribution for the data.

## 5. MCMC Simulation Results

For each chain of our MCMC simulations, we omit the first 60,000 iterations and apply a thinning of 40 (i.e. sampling every 40th iteration) when collecting the posterior samples. The resulting 10,000 samples are used for obtaining the required inferences. We notice that the autocorrelations and cross-correlations over successive iterations are negligible and there is no issue in convergence to the stationary distribution. For instance, Figure 5 illustrates that the autocorrelations of the selected copula's parameter between the D and P losses and also the cross-correlations between the selected copulas' parameters of the edges DP and DH are insignificant. All the estimated Monte Carlo errors are also within 2% of the sample standard deviations, which further indicate that the



**Figure 6.** The most optimal R-vine (C-vine) in terms of DIC (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

level of convergence is satisfactory. Moreover, the results given by different chains with reasonably spaced initial values are very close. It means that the simulations are generally robust to the selection of initial values.

Figure 6 demonstrates the vine copula structure with the lowest Deviance Information Criterion (DIC) value. The DIC is calculated as the posterior mean of  $-2 \ln p(X_i|\Theta)$  plus the effective number of parameters. The most optimal structure, amongst all the 24 possible R-vines, turns out to be a C-vine (DIC = 6,941). In Tree 1, D is the unique node connected to the other three nodes. In Tree 2, the edge DH (bivariate copula between D and H) from Tree 1 becomes the unique node. Table 2 gives the posterior mean of the weight  $w_j$  of bivariate copula density  $c_j$  for each edge in each tree. In Tree 1, the highest weights (italic figures in Table 2) of the edges DP and DH come from the Clayton copula, and the highest weight of DE comes from the  $90^\circ$  rotation. In Tree 2, the highest weights of HP|D and HE|D belong to the Clayton copula and the  $90^\circ$  rotation, respectively. In Tree 3, the highest weight of PE|DH arises from the  $270^\circ$  rotation. These copulas with the most weights can be seen as the “best” models under Bayesian inferencing. This approach of Bayesian copula selection is more unified with the entire estimation process, when compared to the ad hoc approach of using a standard statistical measure like the AIC to choose the pairwise copulas consecutively.

There are a number of interesting observations on the posterior estimates. First, the final choices for DH in Tree 1 (Clayton) and PE|DH in Tree 3 ( $270^\circ$  rotation) are broadly in agreement with the graphs in Figure 3. The former has lower tail dependence while the latter has tail dependence in the upper-left quadrant. Moreover, in Tree 1, the choices are quite certain as the highest weights are 0.78 to 0.86, but in Tree 3, the highest weight is only 0.51 and there is a second highest weight of 0.26. There are a few potential reasons behind the less certain choice in Tree 3. Generally speaking, at the higher trees, the variables are conditioned on a larger number of other variables, which may mix up or dilute the patterns. They tend to have lower dependence and so the effects between fitting different copulas would become less distinct. Furthermore, as shown in Figure 3 (right graph), there seems to be some form of upper tail dependence (opposite to lower tail dependence) as well, which would also make the  $180^\circ$  rotation a competitive candidate. The tail dependence properties of the copula choices for DP, DH, and DE are broadly in line with the coefficients of tail dependence (lower tail of 0.22, lower tail of 0.34, bottom-right of 0.28, respectively) estimated non-parametrically (e.g. Schmidt & Stadtmüller, 2006). Table 3 provides the posterior means of the parameters of the selected copulas from  $p(\Theta|X, M)$  (i.e. BMS). It is worth noting that while the lower trees appear to have stronger relationships (larger implied  $\tau$ 's from calculated  $\theta$ 's), in line with the rationale of the maximum spanning tree in pulling together

**Table 2.** Posterior means of weights of bivariate copula candidates for each edge in each tree in the most optimal vine copula structure (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

$w_j$	Tree 1			Tree 2		Tree 3
Copula	DP	DH	DE	HP D	HE D	PE DH
Clayton	0.78	0.80	0.00	0.61	0.00	0.03
90°	0.00	0.00	0.86	0.01	0.78	0.02
180°	0.02	0.00	0.00	0.07	0.00	0.26
270°	0.00	0.00	0.01	0.01	0.00	0.51
Frank	0.05	0.12	0.03	0.09	0.17	0.07
Gaussian	0.14	0.09	0.09	0.22	0.05	0.09

**Table 3.** Posterior estimates (mean, median, standard deviation, tail quantile, correlation, Kendall's tau) of parameters of selected copulas in the most optimal vine copula structure (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

Copula	Mean	Median	SD	2.5%	97.5%	MLE
DP	0.36	0.34	0.14	0.12	0.66	0.25
DH	0.52	0.51	0.15	0.26	0.84	0.32
DE	0.29	0.28	0.12	0.09	0.55	0.17
HP D	0.28	0.27	0.13	0.06	0.57	0.19
HE D	0.67	0.67	0.15	0.38	0.96	0.50
PE DH	0.21	0.20	0.11	0.02	0.45	0.13
Correlation	DP	DH	DE	HP D	HE D	PE DH
DP	1.00	0.25	0.11	-0.01	-0.05	0.14
DH	0.25	1.00	0.23	0.01	-0.03	0.05
DE	0.11	0.23	1.00	-0.01	-0.02	0.02
HP D	-0.01	0.01	-0.01	1.00	0.07	-0.06
HE D	-0.05	-0.03	-0.02	0.07	1.00	-0.08
PE DH	0.14	0.05	0.02	-0.06	-0.08	1.00
Tau	DP	DH	DE	HP D	HE D	PE DH
DP	1.00	0.17	0.08	-0.01	-0.03	0.10
DH	0.17	1.00	0.14	0.01	-0.02	0.04
DE	0.08	0.14	1.00	-0.01	-0.02	0.02
HP D	-0.01	0.01	-0.01	1.00	0.04	-0.03
HE D	-0.03	-0.02	-0.02	0.04	1.00	-0.05
PE DH	0.10	0.04	0.02	-0.03	-0.05	1.00

those with stronger associations first, there is an exception of HE|D in Tree 2, which has the largest implied  $\tau$  ( $= -0.25$ ) in magnitude. This result pinpoints that though the maximum spanning tree is a convenient and practical method to find a reasonable vine structure, the resulting choice may not be the truly optimal structure overall, at least in terms of the DIC being used here. Dißmann et al. (2013) also commented that the maximum spanning tree method does not necessarily lead to the global optimum (e.g. likelihood, AIC). Further evidence on this issue using Bayesian inferencing is provided in Section 6.

**Table 4.** Goodness-of-fit test statistics (Kolmogorov–Smirnov and AIC; rankings amongst six copula candidates in brackets) of selected copulas in the most optimal vine copula structure (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

Copula	KS	AIC
DP (Clayton)	0.12 (2)	−4.01 (1)
DH (Clayton)	0.09 (1)	−7.58 (1)
DE (90°)	0.14 (6)	−1.35 (1)
HP D (Clayton)	0.09 (2)	−3.20 (1)
HE D (90°)	0.10 (1)	−24.71 (1)
PE DH (270°)	0.14 (5)	−0.28 (1)

Table 3 also contains the posterior medians, standard deviations, and tail quantiles of the copula parameters, and the corresponding correlations and Kendall’s tau estimates between the parameters. The posterior distributions of the copula parameters are largely symmetric. All the parameters are significantly different from zero (based on the *t* test at 5% significance level). Since the maximum likelihood estimates correspond roughly to the posterior modes under uniform priors, they are also included in the table for comparison. While their relative sizes are similar to those of the posterior means and medians, they are all smaller by about 0.1 or more. As these maximum likelihood estimates are obtained from the IFM method and then the usual sequential estimation method, the extent that they are smaller would be a reflection of the biases being introduced in the multi-step estimation process. From a Bayesian perspective, it is also desirable to estimate all the parameters jointly such that the associations amongst the parameters can also be assessed. In Table 3, a few posterior correlations are significantly different from zero. This kind of dependence should then be taken into account when using the calibrated model to simulate the future values.

Table 4 shows some additional test results on the goodness-of-fit of the selected copulas in Table 2 above, based on the posterior means of the parameters (BMS), but outside the Bayesian system. The tests include a bivariate version of the Kolmogorov–Smirnov test and the AIC, as follows:

$$KS = \max_j |\tilde{Pr}(X_1 \leq x_{1,j}, X_2 \leq x_{2,j}) - \tilde{C}(u_{1,j}, u_{2,j})|$$

$$AIC = -2\hat{l} + 2,$$

where  $\tilde{Pr}(\cdot)$  denotes the empirical probability,  $\tilde{C}$  represents the selected copula,  $u_{1,j} = F_1(x_{1,j})$ ,  $u_{2,j} = F_2(x_{2,j})$ , and  $\hat{l}$  is the computed log-likelihood. The ranking of the test statistic value of each selected copula amongst the initial list of copula choices is given in brackets. The AIC values agree that the selected copulas under the Bayesian framework are the best choices. On the other hand, while the Kolmogorov–Smirnov test values indicate that the selected copulas under the Bayesian framework are one of the two best choices for four of the pairs, the values for DE and PE|DH suggest otherwise. These differences may be caused by the fact that these two pairs have weaker relationships than the other pairs and so their patterns may not be as clear-cut as the others, leading to less consistent conclusions between different statistical criteria. Moreover, this test requires the copula data obtained from the IFM method, which means that the consideration of whether a copula is suitable or not is separate from the consideration of the margins. This view is different to the rationale of coherent selection and estimation within our Bayesian framework. Nevertheless, all the test values are statistically insignificant and none of the selected copulas are rejected under the Kolmogorov–Smirnov test.

We now examine an application to risk management to see the impact of using an appropriate vine copula. Taking reference of the Solvency II regulations in Europe, we calculate the

**Table 5.** Posterior estimates of monthly value-at-risk for a portfolio aggregating events of all four breach types under five different dependence assumptions, including BMS, BMA, multivariate  $t$ , multivariate Gaussian, and independence copulas.

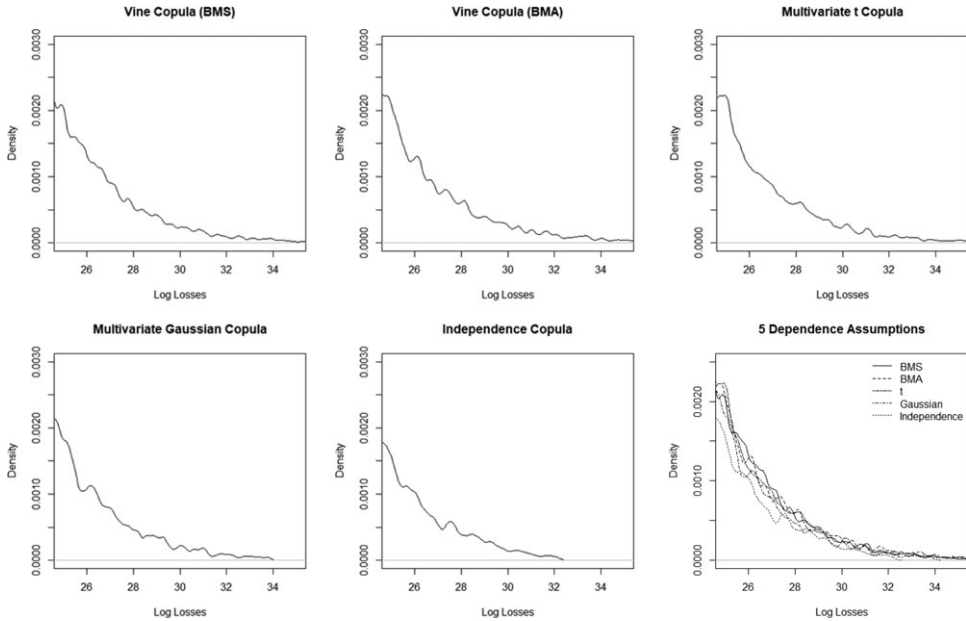
Copula	Value-at-Risk (in 10,000 s)			
	90%	95%	99%	99.5%
BMS	67	366	18,140	99,304
BMA	67	338	18,338	126,610
$t$	66	331	14,875	80,701
Gaussian	64	297	12,142	66,120
Ind.	58	244	8,202	37,028

value-at-risk (VaR) of the monthly total data breach losses (aggregating events from all the four breach types into one portfolio)<sup>8</sup>. While the selected copulas in Table 2 are fairly dominating and so using  $p(X_f|X, M)$  (i.e. BMS) appears to be reasonable, the results from using  $p(X_f|X)$  (i.e. BMA) are also obtained for comparison. Moreover, a few other different assumptions are also included in the analysis. The first is simply assuming the four breach types are independent of one another. The second is adopting a multivariate Gaussian copula as well as a multivariate  $t$  copula<sup>9</sup>, which are restrictive in the sense that they impose the same shape of dependence and tail dependence across all pairs of breach types. Table 5 sets forth the posterior estimates of the risk measures under these five assumptions. First, assuming independent data breach types ignores the underlying relationships and clearly underestimates the VaR. While the multivariate Gaussian copula allows for the relationships between the losses, it does not take tail dependence into account and so the resulting risk measures are still smaller than the others. Compared to the multivariate Gaussian copula, the multivariate  $t$  copula, with tail dependence, gives larger estimates only by some extent, probably because it still lacks the flexibility of dealing with each pair differently and the degree of freedom is 14 on average which is quite large. The 99.5% VaR estimates under the BMS produced by our Bayesian model are significantly larger than the three sets of estimates above, highlighting the critical importance of catering for the dependency structure properly when assessing the extreme tail. Further, it is interesting to see that the Bayesian model under the BMA gives the largest 99% and 99.5% VaR estimates. This difference reflects a substantial impact of incorporating copula model uncertainty into the framework<sup>10</sup>. It suggests that the practice of choosing a copula based on the AIC or similar measures and then relying solely on its results could lead to a significant underestimation of the aggregate risk. Figure 7 plots the extreme right tails of the simulated posterior (kernel) densities of the monthly total data breaches (in logarithmic scale) under the five assumptions. All the five posterior distributions appear to be skewed to the right. The BMA and BMS ones have the thickest tails, followed by those from the multivariate  $t$ , multivariate Gaussian, and then independence copulas.

<sup>8</sup>If data were sufficient, it would be interesting to compare our approach against using extreme-value copulas with extreme-value distributions for the purpose of capital assessment. However, we only have about 123 monthly data points, which make the latter approach practically infeasible. Moreover, it often becomes rather arbitrary when choosing the data thresholds for fitting extreme-value kind of models, due to the usual difficulty in striking a balance between the degree of compliance to the extreme value theory and a sufficient sample size.

<sup>9</sup>The 4-dimensional Gaussian and  $t$  copula functions are  $\Phi_4(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3), \Phi^{-1}(u_4))$  and  $t_{\nu,4}(t_{\nu}^{-1}(u_1), t_{\nu}^{-1}(u_2), t_{\nu}^{-1}(u_3), t_{\nu}^{-1}(u_4))$ , in which  $t_{\nu}^{-1}$  is the Student  $t$  inverse distribution function,  $t_{\nu,4}$  is the distribution function of the standard 4-dimensional  $t$  distribution,  $D$  is the dispersion matrix, and  $\nu$  is the degree of freedom. We assume an uninformative uniform prior for the non-zero non-diagonal components in  $C$ , in which  $D = CBC'$ ,  $C$  is a lower triangular matrix, and  $B$  is a diagonal matrix with positive entries, such that  $D$  is positive definite. The latter copula converges to the former when  $\nu$  approaches infinity. Both are implemented in a Bayesian setting together with the margins.

<sup>10</sup>Note that the losses here are expressed in terms of the number of records. Without detailed information about the monetary sizes of the different data breaches, it would be difficult to assess the financial impact further. As a supplementary analysis, Section 7 analyses another data set of data breach losses and provides a further discussion on the financial impact.



**Figure 7.** Simulated posterior densities of monthly total data breaches under five different dependence assumptions, including BMS, BMA, multivariate *t*, multivariate Gaussian, and independence copulas.

Under Solvency II in Europe, all of process error, parameter error, and model error must be taken into account sufficiently. For instance, QIS5 Technical Specifications (CEIOPS, 2010) stated that “it is important to assess the model error that results from the use of a given valuation technique.” It described model error and parameter error as the “uncertainty which is related to the measurement of the risk and is not an intrinsic property of the risk.” It would be inadequate to simply use the empirical or fitted distribution to assess the aggregate tail risk, which then includes only the inherent risk (process error). This full allowance for all uncertainties is particularly important when data are scarce – in our analysis, not just that there are merely about 123 monthly data points, but also that data breaches and cyber losses represent a new kind of insurance opportunities, the development of which is still at its infant stage. Ignoring the potential impact of model uncertainty and parameter uncertainty for a new market is a significant omission in risk management and capital assessment under the current regulations in Europe. Over time, as more data can be collected and more other information can be obtained about data breaches and cyber losses, more confidence can be given to the modelling process and the extent of model error and parameter error would then naturally be reduced.

### 6. Sensitivity Testing

In this section, we perform a range of sensitivity tests on the Bayesian modelling results discussed in the previous section. Firstly, we experiment with a number of different priors for the copula parameters and the copula selection. While the uniform distribution is often a reasonable default choice, it would be informative to explore the consequences of taking other possible alternatives. Table 6 compares the posterior means and standard deviations of the copula parameters (BMS) obtained from using uninformative logistic, normal, or Cauchy priors<sup>11</sup> instead against the

<sup>11</sup>For the Clayton copula parameter, the logistic, normal, and Cauchy priors are truncated from below at 0. For the Gaussian copula parameter, the three priors are truncated from above at 1 and from below at -1. For Kendall’s tau, the priors



**Table 6.** Posterior estimates (mean, standard deviation) of parameters of selected copulas from eight different priors, using the most optimal vine copula structure (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

Mean	Uniform	Logistic	Normal	Cauchy	Uniform	Logistic	Normal	Cauchy
Copula	on copula parameter				on Kendall's tau			
DP	0.36	0.39	0.38	0.37	0.33	0.36	0.36	0.34
DH	0.52	0.51	0.52	0.51	0.49	0.49	0.49	0.49
DE	0.29	0.31	0.30	0.3	0.27	0.29	0.29	0.28
HP D	0.28	0.29	0.28	0.29	0.26	0.28	0.27	0.26
HE D	0.67	0.69	0.71	0.74	0.64	0.69	0.70	0.72
PE DH	0.21	0.23	0.22	0.21	0.20	0.22	0.21	0.20
SD	Uniform	Logistic	Normal	Cauchy	Uniform	Logistic	Normal	Cauchy
Copula	on copula parameter				on Kendall's tau			
DP	0.14	0.13	0.13	0.14	0.14	0.13	0.13	0.14
DH	0.15	0.14	0.15	0.15	0.15	0.14	0.15	0.15
DE	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11
HP D	0.13	0.13	0.12	0.12	0.12	0.12	0.12	0.12
HE D	0.15	0.17	0.17	0.2	0.16	0.18	0.18	0.19
PE DH	0.11	0.11	0.12	0.11	0.11	0.11	0.11	0.11

previous estimates. Another alternative of assuming uninformative uniform, logistic, normal, or Cauchy priors for Kendall's tau and then converting it into the copula parameter is also tested.

It can be seen from Table 6 that the posterior means and standard deviations are fairly consistent across different prior assumptions. It means that our Bayesian framework is reasonably robust to the prior settings and that the priors used are adequately vague priors. This result is also reasonable in the sense that as the selected copulas are fairly dominating, the MCMC simulation would give similar results for any reasonable prior choices.

Moreover, we examine the effects of using unequal prior probabilities of  $p(M)$  rather than the discrete uniform distribution. Table 7 compares the posterior means of the weights under different settings of prior probabilities. It is observed that when much larger prior probabilities are given to the other copula candidates, the final posterior weights of those pairs with stronger relationships (all of Tree 1 and most of Tree 2 cases) still indicate the same optimal choices as previously, though the weights become a little lower. Despite the reasonable robustness of our results, it is important to note that certain prior information can be incorporated into a more informative prior if it is available and relevant. For instance, a high prior probability can be allocated to a particular copula if some other related data of similar nature but with a bigger size demonstrate the kind of dependence possessed by that copula.

We then compare the situations when model uncertainty and parameter uncertainty do not exist. The former can be removed by using the selected copulas directly without allowing for Bayesian copula selection. The latter condition can be created by removing the prior distributions and treating the parameter estimates as fixed values. Table 8 shows the simulated estimates of the above-mentioned risk measures under the three ways of allowing for uncertainties. Parameter uncertainty clearly makes the most significant contribution to the aggregate risk, as revealed by

are truncated from above at 1 and from below at  $-1$  in general; in particular, the priors are further truncated from below at 0 for the Clayton copula and the  $180^\circ$  rotation or from above at 0 for the  $90^\circ$  and  $270^\circ$  rotations. Again, the scale parameters of the priors are set as large as possible, given that the range is valid and that there is no interruption in the JAGS simulation process. For the copula parameters, the mean parameters are taken as 0.5 and the scale parameters as 5 in the priors. For Kendall's tau, the mean parameters are set as 0.2 and the scale parameters as 2 in the priors.

**Table 7.** Posterior means of weights of selected copulas (in brackets) for each edge in each tree in the most optimal vine copula structure, using different sets of prior probabilities (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

$w_j$	Tree 1			Tree 2		Tree 3
Priors	DP	DH	DE	HP D	HE D	PE DH
Uniform	0.78 (Clayton)	0.80 (Clayton)	0.86 (90°)	0.61 (Clayton)	0.78 (90°)	0.51 (270°)
50% 180°; 10% Others	0.68 (Clayton)	0.85 (Clayton)	0.87 (90°)	0.45 (Clayton)	0.90 (90°)	0.67 (180°)
50% Frank; 10% Others	0.60 (Clayton)	0.66 (Clayton)	0.80 (90°)	0.44 (Clayton)	0.65 (90°)	0.33 (270°)
50% Gauss; 10% Others	0.49 (Clayton)	0.59 (Clayton)	0.69 (90°)	0.56 (Gauss)	0.65 (90°)	0.38 (Gauss)

**Table 8.** Simulated estimates of monthly value-at-risk for a portfolio aggregating events from all four breach types under three different ways of allowing for uncertainties, including BMS and BMA, having model error removed, and having both parameter error and model error removed.

	Value-at-Risk (in 10,000 s)			
Copula	90%	95%	99%	99.5%
BMS	67	366	18,140	99,304
BMA	67	338	18,338	126,610
No Model Error	64	335	13,307	67,234
No Model & Parameter Errors	23	71	1,139	3,693

**Table 9.** Posterior means of weights of bivariate copula candidates for each edge in each tree in the most optimal vine copula structure, using mixtures of copulas (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

$w_j$	Tree 1			Tree 2		Tree 3
Copula	DP	DH	DE	HP D	HE D	PE DH
Clayton	<i>0.23</i>	<i>0.23</i>	0.15	<i>0.21</i>	0.09	0.13
90°	0.15	0.13	<i>0.22</i>	0.15	<i>0.25</i>	0.14
180°	0.15	0.15	0.15	0.17	0.09	<i>0.21</i>
270°	0.13	0.12	0.15	0.13	0.17	0.21
Frank	0.17	0.19	0.16	0.17	0.22	0.15
Gaussian	0.17	0.18	0.17	0.18	0.18	0.16

the much smaller calculated risk measures when it is excluded. Model uncertainty also plays an important part, though to a lesser extent. Insurance loss data are usually limited in size, and it is highly inappropriate to overlook the substantial impact of parameter and model uncertainties.

We also allow the prior weights of the copulas to be a continuous variable between zero and one instead (with the sum of all the prior weights being one). In this way, the effective copula density becomes a “mixture” of different copula densities, which can offer more flexibility for modelling the dependency structure. Table 9 gives the posterior means of the weights under the new prior assumption. It is interesting to see that while the copulas with the largest weights (italic) are the same as before (except Tree 3, marginally), the weights are more spread out across different candidates. It is a clear sign of this alternative setting being able to exploit and mingle the various features of the different copulas to cope with the data complexity. However, the resulting DIC value increases to 6,956, which implies that the enhanced model sophistication cannot be justified by the better fitting.

**Table 10.** Posterior means of weights of marginal distribution candidates based on dummy prior weights and continuous prior weights, using the most optimal vine copula structure (H: hacking/malware; P: loss of physical records of data; E: loss of electronic devices containing sensitive data; D: other unintended disclosure).

$w_j$	Dummy Prior Weights				Continuous Prior Weights			
	H	P	E	D	H	P	E	D
Marginal Distribution								
Gamma	0.00	0.29	0.01	0.00	0.21	0.27	0.40	0.04
Weibull	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
Lognormal	0.99	0.71	0.99	0.00	0.35	0.46	0.30	0.42
Pareto	0.01	0.00	0.00	1.00	0.43	0.26	0.29	0.52

We further incorporate the selection of vine structure directly into the Bayesian framework. First, we choose the vine structures with the two lowest DIC values and also the optimal structure from the maximum spanning tree. Then, we express the effective joint density as a weighted average of those of the three structures, i.e.  $f(x_1, x_2, x_3, x_4) = \sum w_j f_j(x_1, x_2, x_3, x_4)$ , where the weight  $w_j$  is equal to one if the corresponding joint density  $f_j$  is selected and is zero otherwise. From a Bayesian perspective, the best vine structure would have the highest posterior probability. The posterior means of the weights of the three vine structures (from two lowest DICs and maximum spanning tree) included are 0.54, 0.46, and 0, respectively, and the results of which are consistent with those from the use of DIC values (6,941, 6,942, 6,946). (Based on maximum likelihood, given the margins, their corresponding log-likelihood values are 17.69, 16.77, 15.33; the AIC values are  $-23.37$ ,  $-21.54$ ,  $-18.65$ ; and the BIC values are  $-6.50$ ,  $-4.67$ ,  $-1.78$ , respectively.) This test reemphasises the fact that the maximum spanning tree may not identify the truly optimal structure, despite being a convenient heuristic method which is widely applied.

We also replace the GB2 distribution with a mixture of the gamma, Weibull, lognormal, and Pareto distributions, using the same rationale in the copula selection and vine structure selection above. Both dummy prior weights (0 or 1) and continuous prior weights (0 to 1) are tested. Correspondingly, Table 10 provides the resulting posterior means of the weights of different marginal distributions for each breach type. Amongst the four two-parameter distribution candidates, the Pareto or Lognormal distribution is clearly the optimal choice when the prior selection can only be one distinct distribution. However, when a mixture distribution is allowed in the prior, the posterior weights of the individual candidates are more dispersed (except for Weibull). Interestingly, the optimal choices under the two prior assumptions match with each other for the P and D losses but not for the H and E losses. Furthermore, we realise that the DIC value of the latter is a lot smaller than that of the former. These results imply that the two-parameter distributions may be insufficient to capture all the salient features of the data breach losses. Practitioners trying to model cyber losses with two-parameter distributions should thus exercise extra caution, especially for pricing and reserving purposes, as the losses may then be substantially underestimated. The consequent effects on the vine copula model are also examined. Although the best copula choices remain the same as before, the copula parameter estimates (BMS) are significantly, systematically smaller under the dummy prior weights. This deviation further confirms the insufficiency of using the two-parameter distributions for analysing the data breach events in this paper.

Furthermore, we incorporate a time trend into the GB2 distribution in order to allow for possible changes in data breach sizes over time. Since the mean of the GB2 distribution is equal to  $b \frac{\Gamma(p+\frac{1}{a})\Gamma(q-\frac{1}{a})}{\Gamma(p)\Gamma(q)}$ , we set  $b(t) = b(t-1) + \mu + \varepsilon(t)$ , in which  $b(t)$  becomes a time-varying parameter,  $\mu$  is the drift term, and  $\varepsilon(t)$  is the normal error term (with zero mean) at month  $t$ . Uninformative uniform, normal, and inverse gamma priors are used for  $b(1)$ ,  $\mu$ , and the error term's variance, respectively. Approximate  $t$  test statistics are then calculated on  $\mu$  for the four breach types via dividing the posterior means by the posterior standard deviations. It turns out

that the time drifts of the P and E losses are statistically significant while those of the D and H losses are not. At the same time, the posterior means of some copula parameters are somewhat smaller than previously. It appears that some of the potential relationships between the data breach losses across time are now captured by the newly added temporal trends and accordingly the levels of association reflected in some of the copulas are reduced.

We also experiment with adding a time trend into the dependency structure to allow for possible changes in dependency over time. As an illustration, we assume a time-varying  $\tau = \frac{\exp(2s)-1}{\exp(2s)+1}$  for HE|D in Tree 2 (with the largest copula parameter estimate in magnitude in Table 3), where  $s(t) = s(t-1) + \mu + \varepsilon(t)$ ,  $\mu$  is the drift term, and  $\varepsilon(t)$  is the normal error term at month  $t$ . The priors are set in the same way as above. Based on the  $t$  test statistic calculated ( $-1.37$ ), the time drift is not statistically significant, which means that there is no need to use a dynamic copula in this case.

Finally, we test two simplified vine structures (i.e. truncated vines), one of which has the variables set simply as independent in Tree 3, and the other having all the variables in Trees 2 and 3 being independent with one another (e.g. Brechmann *et al.*, 2012). Their DIC values are 6,942 and 7,051, which are higher than that of the optimal full model above (6,941), suggesting that these two vine structures are rather over-simplified and comparatively not as suitable for modelling the breach data.

## 7. Further Analysis on Department of Health and Human Services (HHS) Data set

In this section, we study a new data set containing healthcare data breaches from 2009 to 2020, obtained from the ongoing public database maintained by the U.S. Department of Health and Human Services (HHS). Recently, this database has been analysed by McLeod and Dolezel (2018) in the field of cyber-analytics. There are 3,241 data breach incidents reported during the investigation period with the incident description, number of individuals affected, and type of breach made known to the public. We examine the severity of the monthly data breach losses in terms of the number of affected individuals from all the incidents reported in the month by four major breach types, as well as the dependency structure between the different breach types. The four major breach categories include the losses that are mainly caused by theft (T), hacking (H), unauthorised access (U), and mishandling of documents (M). The category for mishandling of documents includes incidences that can be mainly described as either loss of documents or improper disposal of documents. Essentially, the first three data breach categories are concerned with malicious breach attempts, whereas the last category covers negligence in handling confidential or sensitive information. Categorisation of the incidents depends on both the type of breach as specified in the database and incident description provided in the database. Altogether, the number of incidents covered by the four categories accounts for 95% of all the data breach events reported during the study period. Table 11 provides a summary of the descriptive statistics of the monthly losses by breach type. Again, the standard deviations of the losses are much larger than the means, which suggest a high level of uncertainty in the future outcomes. Compared to the summary statistics in Table 1, however, the coefficients of variation of the losses in the HHS data tend to be smaller than those of the PRC data. For both data sets, the hacking breach type contributes to a larger number of losses than the other types of losses, indicating that hacking is a very serious issue in data security.

The most optimal structure for the HHS data set is a C-vine (DIC = 1,593). In Tree 1, T is the unique node linked to the others. In Tree 2, the edge TU from Tree 1 becomes the unique node. Table 12 provides the posterior weight of bivariate copula density for each edge in each tree. In Tree 1, the highest weights (italic figures in Table 12) of TH, TU, and TM come from the Frank copula, Gaussian copula, and 90° Clayton rotation. In Tree 2, the highest weights of HU|T and UM|T belong to the Clayton copula and the 270° rotation. In Tree 3, the highest weight of

**Table 11.** Summary statistics of HHS monthly data breach events from 2009 to 2020 (T: theft; H: hacking; U: unauthorised access; M: mishandling of sensitive/confidential documents).

Type of Data Breach	Average	Standard Deviation
T	213,599	715,842
H	975,148	2,444,078
U	128,943	338,098
M	103,820	543,303

**Table 12.** Posterior means of weights of bivariate copula candidates for each edge in each tree in the most optimal vine copula structure (T: theft; H: hacking; U: unauthorised access; M: mishandling of sensitive/confidential documents).

$w_j$	Tree 1			Tree 2		Tree 3
Copula	TH	TU	TM	HU T	UM T	HM TU
Clayton	0.00	0.04	0.03	0.40	0.18	0.14
90°	0.17	0.17	0.45	0.05	0.07	0.12
180°	0.00	0.08	0.03	0.15	0.18	0.08
270°	0.02	0.24	0.10	0.04	0.25	0.19
Frank	0.70	0.22	0.18	0.14	0.15	0.21
Gaussian	0.11	0.26	0.22	0.22	0.16	0.26

HM|TU arises from the Gaussian copula. It can be seen that different pairs of random variables call for different copula choices under the Bayesian setting.

There are some similarities between the HHS results here and the PRC results in Section 5. First, the optimal structure is a C-vine in both cases, while the other structures like a D-vine are suboptimal. Second, the hacking breach type, despite having the largest number of losses, is not chosen as the unique node in Tree 1 under the Bayesian framework for both data sets. It reflects that whether a breach type would be selected as the unique node depends mainly on how it is associated with the other breach types but less on its size. Nevertheless, there are significant differences in the components of the two vines. Firstly, all the selected copulas for the PRC data have tail dependence, but only half of the selected copulas for the HHS data have tail dependence. Moreover, the posterior weights of the selected copulas are lower for the HHS data than for the PRC data. These results are in line with the fact that the HHS data have comparatively fewer extreme losses than the PRC data do. For instance, as noted earlier, the coefficients of variation of the losses in the HHS data are smaller. So, using the HHS data, those copulas with tail dependence would be less needed while the other copulas without tail dependence would become more competitive in the Bayesian selection process.

Table 13 gives the posterior means, medians, standard deviations, and tail quantiles of the parameters of the selected copulas from the BMS. The copula parameters are significantly different from zero, except for TM and HM|TU. As a comparison, the corresponding maximum likelihood estimates are quite close to the posterior means and medians. The lower trees tend to have stronger relationships (larger implied  $\tau$ 's from calculated  $\theta$ 's) than the higher ones.

Table 14 (top panel) reports the posterior estimates of the VaR of the monthly total data breach losses under the five assumptions of BMS, BMA, multivariate  $t$ , multivariate Gaussian, and independence copulas for the HHS data. We adopt the method in Jacobs (2014) and convert the number of events into dollar terms. It is a regression model given as  $dollar\ loss = \exp[7.68 + 0.76 \times \ln(events)]$ . This empirically deduced equation can serve as a rough approximation of the real

**Table 13.** Posterior estimates (mean, median, standard deviation, tail quantile) of parameters of selected copulas in the most optimal vine copula structure (T: theft; H: hacking; U: unauthorised access; M: mishandling of sensitive/ confidential documents).

Copula	Mean	Median	SD	2.5%	97.5%	MLE
TH	-2.13	-2.11	0.55	-2.95	-1.34	-2.41
TU	-0.13	-0.14	0.05	-0.19	-0.07	-0.11
TM	0.22	0.16	0.16	0.07	0.49	0.22
HU T	0.16	0.13	0.06	0.10	0.25	0.16
UM T	0.12	0.15	0.06	0.03	0.18	0.08
HM TU	-0.01	0.01	0.13	-0.21	0.12	-0.03

**Table 14.** Posterior estimates of monthly value-at-risk for a portfolio aggregating events of all four breach types under five different dependence assumptions, including BMS, BMA, multivariate *t*, multivariate Gaussian, and independence copulas.

HHS		Value-at-Risk (in \$1 m)			
Copula	90%	95%	99%	99.5%	
BMS	9	30	570	2,164	
BMA	9	31	596	2,267	
<i>t</i>	9	30	474	1,545	
Gaussian	9	28	429	1,481	
Ind.	8	24	298	874	
PRC		Value-at-Risk (in \$1 m)			
Copula	90%	95%	99%	99.5%	
BMS	58	211	4,092	14,896	
BMA	58	198	4,126	17,917	
<i>t</i>	57	195	3,519	12,724	
Gaussian	56	180	3,016	10,935	
Ind.	52	155	2,238	7,038	

loss, as there are no available data in monetary terms and there is no general method to estimate the loss. As in the previous section, treating different data breach types as independent omits the underlying relationships and produces the smallest VaR. For the 90% and 95% VaR, assuming any of the copula models considered leads to about 20% increase in the tail estimate, when compared to the independence copula. For the 99% and 99.5% VaR, the multivariate Gaussian and *t* copulas leads to about 60% increase in the tail estimate. Under the BMS and BMA, the 99% and 99.5% VaR estimates are larger by about 90% or more, in which the BMA still produces the largest tail estimates. Table 14 (bottom panel) also presents the posterior VaR estimates for the PRC data used in Section 5. Similarly, for the 90% and 95% VaR, using any of the copula models brings about an 20% increase in the tail estimate, and for the 99% and 99.5% VaR, the multivariate Gaussian and *t* copulas generates a 60% increase in the tail estimate. Under the BMS and BMA, the 99% and 99.5% VaR estimates increase by around 80% or more from those under the independence copula. These results illustrate that both allowing for the dependency structure adequately and incorporating model uncertainty can play a vital role in the assessment of capital. Without addressing these uncertainties properly, there could be significant underestimation of the aggregate risk level and insufficient capital for covering adverse events.



## 8. Concluding Remarks

In this paper, we have adopted a Bayesian framework to implement the vine copula structure for modelling data breach events jointly. We have observed varying levels of tail dependence between different breach types, based on both the optimal copulas selected and non-parametric estimation of tail dependence coefficients. An insurer providing cover on these types of losses must make a sufficient allowance for the dependency structure when calculating premiums and reserves. As demonstrated earlier, while the estimated severities of data breaches at 90% confidence level are not too different (within 20%) from one another regardless of whether they are simulated by the vine copula, or the multivariate elliptical copulas, or the independence model, the differences are drastic for the 99% and 99.5% VaR. They range from about 50% when comparing between the vine and multivariate  $t$  copulas to being more than double between the vine copula and the independence model. Therefore, the assumptions made on the underlying dependencies between risks could have huge impacts on the required capital and other risk management considerations. Regulators and insurers should be keenly aware of model risk and take proper measures to prevent underestimation of risks, especially given the short history of development in the line of cyber insurance business.

With the continual advancement in technology in our fast-paced society, the cyber insurance market is destined for rapid expansion in the foreseeable future, and more knowledge and data in this arena would become accessible over time. Considering our evolving understanding and the dynamic nature of cyber claims, the proposed Bayesian modelling framework is naturally tailored for businesses interested in cyber risk estimation. Besides being capable of incorporating exclusive information provided by subject matter experts, where available, this Bayesian approach allows for a simultaneous determination of the vine structure, copulas, and margins and estimation of all the unknown parameters in a coherent way. It not only avoids introducing biases as under the IFM and sequential estimation methods but also integrates process, parameter, and model uncertainties in one unified setting. A proper allowance for all uncertainties is of utmost importance in the area of risk management under current regulations. Moreover, we also note that the numerical results are quite robust to the prior specifications and that a mixture of copulas can provide more modelling flexibility.

It is often quite difficult to use the maximum likelihood to estimate the copula and marginal parameters jointly. The Bayesian approach does not have this problem and can generate all the estimates coherently from MCMC simulation. Furthermore, by taking all kinds of uncertainties into account, the proposed Bayesian model provides an opportunity to improve the forecasting performance and the computation of risk measures. Particularly, failing to assess model uncertainty appropriately may lead to a serious miscalculation of such regulatory reporting items as reserves and capital.

We have shown how to implement the selection between a few “best” vine structures in the Bayesian framework. However, as the number of possible combinations grows rapidly with the number of dimensions, the scalability of our approach is limited. Our suggestions are that the proposed approach is more suitable for lower dimensions and that for higher dimensions it can be applied in conjunction with other ways of identifying the vine structure such as the standard information criteria, maximum spanning tree, and subjective inputs of grouping the variables in the order of importance. Recently, Gruber and Czado (2018) developed a reversible jump MCMC algorithm for moving within a large model candidate space and showed that it can achieve quick convergence to desirable models. They claimed that their algorithm is suitable for up to around 10 dimensions, and so future research for more scalable Bayesian methods is called for. Moreover, further work is needed to make this kind of advanced algorithms more accessible to industry practitioners and other academics, like the user-friendly platforms offered by WinBUGS (Spiegelhalter et al., 2003) and JAGS, for wider use in actuarial practice and research.

**Acknowledgements.** The authors would like to thank the editor and the referees for their valuable comments and suggestions which greatly enhance the content and presentation of this paper.

## References

- Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**, 182–198.
- Bedford, T. & Cooke, R.M. (2002). Vines – a new graphical model for dependent random variables. *Annals of Statistics*, **30**, 1031–1068.
- Böhme, R. & 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. & Schwartz, G. (2010). Modeling cyber-insurance: towards a unifying framework. Workshop on the Economics and Insurance Security (WEIS). Harvard University, US.
- Brechmann, E.C., Czado, C. & Aas, K. (2012). Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics*, **40**, 68–85.
- Brechmann, E.C., Hendrich, K. & Czado, C. (2013). Conditional copula simulation for systemic risk stress testing. *Insurance: Mathematics and Economics*, **53**, 722–732.
- CEIOPS. (2010). QIS5 Technical Specifications. [https://www.bafin.de/SharedDocs/Downloads/DE/Versicherer\\_Pensionsfonds/QIS/dl\\_qis5\\_technical\\_specifications.html](https://www.bafin.de/SharedDocs/Downloads/DE/Versicherer_Pensionsfonds/QIS/dl_qis5_technical_specifications.html)
- Chan, J.S.K., Choy, S.T.B., Makov, U.E. & Landsman, Z. (2018). Modelling insurance losses using contaminated generalised beta type-II distribution. *ASTIN Bulletin*, **48**, 871–904.
- Chuliá, H., Guillén, M. & Uribe, J.M. (2016). Modeling longevity risk with generalized dynamic factor models and vine-copulae. *ASTIN Bulletin*, **46**, 165–190.
- Cole, C.R. & Fier, S.G. (2021). An empirical analysis of insurer participation in the U.S. cyber insurance market. *North American Actuarial Journal*, **25**, 232–254.
- Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas*. Springer International Publishing, Cham.
- Dißmann, J., Brechmann, E.C., Czado, C. & 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. & Forrest, S. (2016). Hype and heavy tails: a closer look at data breaches. *Journal of Cybersecurity*, **2**, 3–14.
- Eling, M. & Jung, K. (2018). Copula approaches for modeling cross-sectional dependence of data breach losses. *Insurance: Mathematics and Economics*, **82**, 167–180.
- Eling, M. & Loperfido, N. (2017). Data breaches: goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, **75**, 126–136.
- Fahrenwaldt, M.A., Weber, S. & Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin*, **48**, 1175–1218.
- Fink, H., Klimova, Y., Czado, C. & Stöber, J. (2017). Regime switching vine copula models for global equity and volatility indices. *Econometrics*, **5**, 1–38.
- Frahm, G., Junker, M. & Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance: Mathematics and Economics*, **37**, 80–100.
- Geidosch, M. & Fischer, M. (2016). Application of vine copulas to credit portfolio risk modeling. *Journal of Risk and Financial Management*, **9**, 1–15.
- Gruber, L. & Czado, C. (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis*, **10**, 937–963.
- Gruber, L.F. & Czado, C. (2018). Bayesian model selection of regular vine copulas. *Bayesian Analysis*, **13**, 1111–1135.
- Herath, H.S.B. & Herath, T. (2011). Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies: Analyses and Actuarial Computations*, **2**, 7–20.
- Hobæk Haff, I., Aas, K. & Frigessi, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic? *Journal of Multivariate Analysis*, **101**, 1296–1310.
- Jacobs, J. (2014). *Analyzing Ponemon Cost of Data Breach*. <https://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. *Lecture Notes-Monograph Series*, **28**, 120–141.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, New York.
- Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall, New York.
- Kshetri, N. (2020). The evolution of cyber-insurance industry and market: an institutional analysis. *Telecommunications Policy*, **44**, 1–14.
- McLeod, A. & Dolezel, D. (2018). Cyber-analytics: modeling factors associated with healthcare data breaches. *Decision Support Systems*, **108**, 57–68.
- Min, A. & Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, **8**, 511–546.

- Min, A. & Czado, C.** (2011). Bayesian model selection for D-vine pair-copula constructions. *The Canadian Journal of Statistics*, **39**, 239–258.
- Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A. & Sadhukhan, S.K.** (2006). e-risk management with insurance: a framework using copula aided Bayesian belief networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, **6**, 126.
- Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A. & Sadhukhan, S.K.** (2013). Cyber-risk decision models: to insure IT or not? *Decision Support Systems*, **56**, 11–26.
- Nelsen, R.B.** (1999). *An Introduction to Copulas*. Springer-Verlag, Heidelberg.
- Ögüt, 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**, 497–512.
- Okhrin, O., Okhrin, Y. & Schmid, W.** (2013). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, **173**, 189–204.
- Plummer, M.** (2017). *JAGS Version 4.3.0 User Manual*. <https://sourceforge.net/projects/mcmc-jags/>
- Schmidt, R. & Stadtmüller, U.** (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, **33**, 307–335.
- Shi, P. & Yang, L.** (2018). Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, **113**, 122–133.
- Sklar, A.** (1959). Fonctions de répartition a n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8**, 229–231.
- Smith, M.S.** (2011). Bayesian approaches to copula modelling. In P. Damien, P. Dellaportas, N. Polson, and D. Stephens (Eds.), *Hierarchical Models and MCMC: A Tribute to Adrian Smith* (pp. 395–402). Oxford University Press, Oxford, Bayesian Statistics.
- Smith, M.S.** (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, **31**, 815–833.
- Smith, M.S. & Khaled, M.A.** (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association*, **107**, 290–303.
- Smith, M., Min, A., Almeida, C. & Czado, C.** (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, **105**, 1467–1479.
- Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D.** (2003). *WinBUGS User Manual*. <http://www.mrc-bsu.cam.ac.uk/bugs>
- Xu, M. & Hua, L.** (2019). Cybersecurity insurance: modeling and pricing. *North American Actuarial Journal*, **23**, 220–249.
- Zhou, R.** (2019). Modelling mortality dependence with regime-switching copulas. *ASTIN Bulletin*, **49**, 373–407.

## Appendix

Let  $X$  and  $Y$  be two bivariate random variables linked by the copula function  $C(u, v)$  for  $u = F_X(x)$  and  $v = F_Y(y)$ . If it is the Clayton copula  $C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$  for  $\theta > 0$ , the copula density is

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) = (1 + \theta)(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} u^{-\theta - 1} v^{-\theta - 1},$$

and the corresponding  $h$ -functions are derived as

$$\frac{\partial}{\partial u} C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 1} u^{-\theta - 1},$$

$$\frac{\partial}{\partial v} C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 1} v^{-\theta - 1}.$$

If it is the 90° rotation of the Clayton copula  $C(u, v) = v - ((1 - u)^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ , the copula density is

$$c(u, v) = (1 + \theta)((1 - u)^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} (1 - u)^{-\theta - 1} v^{-\theta - 1},$$

and the  $h$ -functions are given as

$$\frac{\partial}{\partial u} C(u, v) = ((1 - u)^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 1} (1 - u)^{-\theta - 1},$$

$$\frac{\partial}{\partial v} C(u, v) = 1 - ((1 - u)^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 1} v^{-\theta - 1}.$$

If it is the 180° rotation  $C(u, v) = u + v - 1 + ((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta}$ , the copula density is

$$c(u, v) = (1 + \theta)((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 2}(1 - u)^{-\theta - 1}(1 - v)^{-\theta - 1},$$

and the  $h$ -functions are

$$\frac{\partial}{\partial u} C(u, v) = 1 - ((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 1}(1 - u)^{-\theta - 1},$$

$$\frac{\partial}{\partial v} C(u, v) = 1 - ((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 1}(1 - v)^{-\theta - 1}.$$

If it is the 270° rotation  $C(u, v) = u - (u^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta}$ , the copula density is

$$c(u, v) = (1 + \theta)(u^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 2}u^{-\theta - 1}(1 - v)^{-\theta - 1},$$

and the  $h$ -functions are

$$\frac{\partial}{\partial u} C(u, v) = 1 - (u^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 1}u^{-\theta - 1},$$

$$\frac{\partial}{\partial v} C(u, v) = (u^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta - 1}(1 - v)^{-\theta - 1}.$$

For the Frank copula  $C(u, v) = -\ln(1 + \frac{\exp(-\theta u) - 1}{\exp(-\theta) - 1} \frac{\exp(-\theta v) - 1}{\exp(-\theta) - 1}) / \theta$  given  $-\infty < \theta < \infty$ , the copula density is

$$c(u, v) = -\frac{\theta \exp(-\theta u) \exp(-\theta v)}{\exp(-\theta) - 1 + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)} + \frac{\theta \exp(-\theta u) \exp(-\theta v)(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{(\exp(-\theta) - 1 + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1))^2},$$

and the corresponding  $h$ -functions are derived as

$$\frac{\partial}{\partial u} C(u, v) = \frac{\exp(-\theta u)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1 + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)},$$

$$\frac{\partial}{\partial v} C(u, v) = \frac{\exp(-\theta v)(\exp(-\theta u) - 1)}{\exp(-\theta) - 1 + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}.$$

For the Gaussian copula  $C(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v))$  given  $-1 < \theta < 1$ , the copula density is

$$c(u, v) = \frac{1}{\sqrt{|D|}} \exp(-\frac{1}{2}(\Phi^{-1}(u), \Phi^{-1}(v))(D^{-1} - I)(\Phi^{-1}(u), \Phi^{-1}(v))'),$$

where  $D = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$  and  $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ , and the  $h$ -functions are given as

$$\frac{\partial}{\partial u} C(u, v) = \Phi\left(\frac{\Phi^{-1}(v) - \theta\Phi^{-1}(u)}{\sqrt{1 - \theta^2}}\right),$$

$$\frac{\partial}{\partial v} C(u, v) = \Phi\left(\frac{\Phi^{-1}(u) - \theta\Phi^{-1}(v)}{\sqrt{1 - \theta^2}}\right).$$