ESTIMATING COPULAS FOR INSURANCE FROM SCARCE OBSERVATIONS, EXPERT OPINION AND PRIOR INFORMATION: A BAYESIAN APPROACH

BY

PHILIPP ARBENZ AND DAVIDE CANESTRARO

ABSTRACT

A prudent assessment of dependence is crucial in many stochastic models for insurance risks. Copulas have become popular to model such dependencies. However, estimation procedures for copulas often lead to large parameter uncertainty when observations are scarce. In this paper, we propose a Bayesian method which combines prior information (e.g. from regulators), observations and expert opinion in order to estimate copula parameters and determine the estimation uncertainty. The combination of different sources of information can significantly reduce the parameter uncertainty compared to the use of only one source. The model can also account for uncertainty in the marginal distributions. Furthermore, we describe the methodology for obtaining expert opinion and explain involved psychological effects and popular fallacies. We exemplify the approach in a case study.

KEYWORDS

Copulas, Expert judgment, Insurance, Dependence measure, Bayesian inference, Correlation, Risk management.

1. INTRODUCTION

In insurance, it is crucial to take into account the effects of dependence when modeling the joint distribution of risks. Estimating the dependence structure is relatively easy if many joint observations are available, see McNeil et al. (2005). In a (re)insurance setting, often only very few *joint* observations are available, which may be the case even when plenty of information is available on the marginal distributions. In such a case, it is often considered adequate to make an assumption of independence or impose simple assumptions on correlations. However, these approaches have been shown to contain several pitfalls when used in risk management, see Embrechts et al. (2002).

Models for financial and insurance risks which account for dependence in a more comprehensive way than a variance-covariance approach have gained

Astin Bulletin 42(1), 271-290. doi: 10.2143/AST.42.1.2160743

much interest in recent years, see e.g. McNeil et al. (2005). As the credit crisis showed, dependence, particularly in the tails, must be accounted for correctly, see Donnelly and Embrechts (2010).

One possibility to model dependence between random variables are copula functions. On some probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, the joint cumulative distribution function of a random vector $(X_1, \ldots, X_d) \in \mathbb{R}^d$ with margins $F_i(x) = \mathbb{P}[X_i \leq x]$, $i = 1, \ldots, d$, can be written as

$$\mathbb{P}[X_1 \leq x_1, \ldots, X_d \leq x_d] = C(F_1(x_1), \ldots, F_d(x_d)), \text{ for all } (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

where $C : [0,1]^d \to [0,1]$ is a so-called copula. We refer to Nelsen (2006) for a detailed introduction to copulas and an overview on parametric copula families. Copulas allow to separate the dependence structure from the margins and have gained widespread use in risk management and financial modeling, see McNeil et al. (2005) and Genest et al. (2009), respectively.

Suppose an insurance company uses copulas to model dependence. If joint observations are scarce, the actuary may decide to use also other sources of information, such as expert judgment or regulatory guidelines, in order to find a good estimate of the copula parameters. For instance, experts may predict certain yet unobserved joint extreme events, which would lead to a higher degree of dependence than what is implied by the observations. We are not aware of an existing sound mathematical framework to combine different sources of information in order to estimate copula parameters. This paper fills this gap by using a Bayesian framework within a parametric copula model and provides a robust method that is applicable even if observations are scarce. Our method is based on Lambrigger et al. (2007) who apply Bayesian inference to combine three sources of information in order to estimate regulatory capital for operational risk.

Decisions that involve expert judgement should be rational and be perceived as such. In particular if this judgement has to be defended in front of auditors, regulators or rating agencies. Furthermore, it is often tricky to avoid psychological traps. Hence, certain expert elicitation principles must be adhered to, which we will outline later.

The paper is organized as follows. In Sections 2 to 5, we concentrate on bivariate copula parameter estimation. Section 2 describes the Bayesian inference approach, Section 3 outlines the methodology to set the prior density and Section 4 describes psychological and procedural aspects of expert judgement. Section 5 discusses the Bayesian modeling of expert assessments. Section 6 extends the model to a multivariate setting, including uncertain marginal distributions. We give an application in Section 7 and conclude in Section 8.

2. Two dimensional Bayesian copula inference

This section introduces the Bayesian inference approach to estimate a copula parameter. For didactic reasons, we concentrate first on the bivariate case with known margins, i.e. we introduce a method for estimation of the copula parameter of a random vector (X_1, X_2), where the margins F_1 and F_2 of X_1 and X_2 are continuous and known. The multidimensional case with unknown margins will be described in Section 6.

For the remainder of the paper, we denote with $\rho(\cdot,\cdot)$ a fixed dependence measure, i.e. ρ maps a pair of random variables to a value in $\mathbb{R}$ (in most practical cases the interval of possible values is $[-1, 1]$ or $[0, 1]$), which is then called their "degree of dependence". We assume that ρ is independent of the margins, i.e. $\rho(X_1, X_2) = \rho(t_1(X_1), t_2(X_2))$ for all random vectors (X_1, X_2) and strictly increasing transformations t_1 and t_2. Commonly used dependence measures satisfying this condition are for example Kendall's tau or asymptotic tail dependence. Other dependence measures can be found in McNeil et al. (2005).

Our method allows statistical inference in the following situation.

Situation 2.1. *The following three sources of information are given.*

(1) *A set $\mathcal{O}$ of N independent observations* $(X_{1,n}, X_{2,n})$, $n = 1, \ldots, N$, *of* (X_1, X_2).
(2) *From K experts, a set $\mathcal{E}$ of point estimates φ_k* of $\rho(X_1, X_2)$, $k = 1, \ldots, K$.
(3) *An additional prior source of information (e.g. regulatory guidelines) which provides an estimate of $\rho(X_1, X_2)$.*

Let $\theta = \rho(X_1, X_2)$ be the unknown value of the dependence measure applied to (X_1, X_2). The direct elicitation of the value of the canonical copula parameter is not feasible as this quantity is not familiar to the expert in terms of the way he collects and evokes his knowledge. However, a question that asks for the value of a dependence measure can be formulated in a way such that substantial answers can be given even if experts are unfamiliar with probability theory. For that reason, we will parameterize the copula through the dependence measure and ask experts to estimate this dependence measure.

Definition 2.2. *Let $\mathcal{C}_2 = \{C_\theta : \theta \in \Theta\}$ denote a family of absolutely continuous bivariate copulas $C_\theta : [0, 1]^2 \to [0, 1]$, parameterized through the dependence measure ρ. I.e. $\rho(U_1, U_2) = \theta$ for vectors (U_1, U_2) with $\mathbb{P}[U_1 \le u_1, U_2 \le u_2] = C_\theta(u_1, u_2)$. We denote with $c(\cdot|\theta)$ the density and with $\Theta \subset \mathbb{R}$ the set of admissible parameters.*

Most combinations of commonly used dependence measures (Kendall's tau, Spearman's rho, asymptotic tail dependence) and bivariate copula classes that are indexed by a real-valued parameter (Clayton, Gumbel, Frank, Gaussian, t with fixed degrees of freedom) satisfy the requirements of Definition 2.2. Note that the parametrization in terms of the dependence measure also guarantees that the copulas are identifiable.

The following proposition allows statistical inference on θ, where the Bayesian approach combines all available information given in Situation 2.1. We will assume that the copula of (X_1, X_2) is contained in $\mathcal{C}_2$ and the experts give estimates of $\rho(X_1, X_2)$. As experts often base their judgement on common knowledge, we allow experts to be dependent, too, and model their dependence structure through a copula. We will fully describe the modeling of expert assessments in Section 5.

For a detailed introduction to Bayesian inference we refer the reader to Bernardo and Smith (1994). For ease of notation, we denote with $p(\cdot)$ all (un)conditional densities of those random variables with no specifically designated density.

Proposition 2.3. *Suppose Situation 2.1 holds along with the following assumptions.*

A1 *Conditionally on the degree of dependence θ, (X_1, X_2) has a distribution given by the copula $C = C_\theta \in \mathcal{C}_2$ and the fixed, known margins F_1* and *F_2. I.e.,*

$$\mathbb{P}[X_1 \le x_1, X_2 \le x_2 \,|\, \theta] = C_\theta(F_1(x_1), F_2(x_2)).$$

A2 *Conditionally on θ, experts point estimates φ_k have a joint distribution given by*

$$\mathbb{P}[\varphi_1 \le x_1, \dots, \varphi_k \le x_K | \theta] = C^E(G_1(x_1|\theta), \dots, G_K(x_K|\theta)),$$

*where $G_k(\cdot|\theta)$ is the conditional distribution of the k-*th *expert and C^E describes the dependence between experts. Assume both C^E and G_k have a density, denoted by c^E and $g_k(\cdot|\theta)$, respectively.*

A3 *Conditionally on θ, $\mathcal{O}$ and $\mathcal{E}$ are independent.*

A4 *The prior source of information can be translated into a prior density $p(\theta)$: $\Theta \to [0, \infty)$.*

Then the posterior density $p(\theta|\mathcal{O},\mathcal{E})$ of θ given $\mathcal{O}$ and $\mathcal{E}$ satisfies

$$\begin{aligned} p(\theta \,|\, \mathcal{O},\mathcal{E}) \propto p(\theta) \prod_{n=1}^{N} c\big(F_1(X_{1,n}), F_2(X_{2,n})|\theta\big) \\ c^E\big(G_1(\varphi_1|\theta), \dots, G_K(\varphi_K|\theta)\big) \prod_{k=1}^{K} g_k(\varphi_k|\theta), \end{aligned} \tag{2.1}$$

where the symbol $\propto$ denotes proportionality with respect to θ.

Proof. Bayes' Theorem leads to $p(\theta|\mathcal{O},\mathcal{E})\, p(\mathcal{O},\mathcal{E}) = p(\mathcal{O},\mathcal{E}|\theta)\, p(\theta)$, which we can write as $p(\theta|\mathcal{O},\mathcal{E}) \propto p(\mathcal{O},\mathcal{E}|\theta)\, p(\theta)$. Due to A3, we get $p(\mathcal{O},\mathcal{E}|\theta) = p(\mathcal{O}|\theta)$ $p(\mathcal{E}|\theta)$. As a consequence of A1 and A2 the quantities $p(\mathcal{O}|\theta)$ and $p(\mathcal{E}|\theta)$, i.e. the likelihoods of $\mathcal{O}$ and $\mathcal{E}$, given θ, can be written as

$$\begin{aligned} p(\mathcal{O}|\theta) &= \prod_{n=1}^{N} c\Big(F_1(X_{1,n}), F_2(X_{2,n})|\theta\Big), \\ p(\mathcal{E}|\theta) &= c^E\Big(G_1(\varphi_1|\theta), \dots, G_K(\varphi_K|\theta)\Big) \prod_{k=1}^{K} g_k(\varphi_K|\theta). \end{aligned}$$ □

We make some remarks on Proposition 2.3 and its assumptions.

Any Bayesian point estimator (such as mean or median) can now be used to calculate a point estimate $\hat{\theta}$ of the copula parameter from $p(\theta|\mathcal{O},\mathcal{E})$. We propose to use the posterior mean $\mathbb{E}[\theta|\mathcal{O},\mathcal{E}]$ in order to be consistent with the modeling of the expert assessments, which will be based on matching conditional moments. The uncertainty of $\hat{\theta}$ can then be assessed through the variance $\mathrm{var}(\theta|\mathcal{O},\mathcal{E})$.

More details on the practical transformation of the prior information into the prior density are given in Section 3. If such a prior source of information is unavailable, an uninformative prior can be used. The modeling of the experts' marginal conditional distribution (the $G_k(\cdot|\theta)$) and the dependence between the experts (through the copula C^E) will be addressed in Section 4. If experts are conditionally independent, then the c^E term in (2.1) drops out. Dependence between experts and observations is difficult to avoid, but it is unclear how this influences the expert assessments. An expert could tend to underestimate due to inexistent joint observations or to overestimate by overcorrecting through unrealistically extreme scenarios.

For sensitivity analyses or in case no expert opinion or observations are available, we can calculate $p(\theta|\mathcal{O})$ and $p(\theta|\mathcal{E})$, the posterior of θ given either $\mathcal{O}$ or $\mathcal{E}$, in which case $p(\mathcal{O}|\theta)$ or $p(\mathcal{E}|\theta)$ drops out of (2.1).

In other applications of Bayesian inference, conjugate priors are often used, in which case both $p(\theta)$ and $p(\theta|\mathcal{O},\mathcal{E})$ belong to the same parametric class of distributions, see Bernardo and Smith (1994). We are not aware of any conjugate priors for copulas.

The normalization factor $p(\mathcal{O},\mathcal{E})$ is not explicitly known in most cases. It is however possible to sample from $\theta|\mathcal{O},\mathcal{E}$ through Markov Chain Monte Carlo methods (MCMC), see Robert and Casella (2005). If Θ is low dimensional (as it is the case in the present bivariate setting), direct grid discretizations may also be feasible.

As our method is intended to be used mainly with small N and K, we refrain from proving the following asymptotic results on the convergence of $\theta|\mathcal{O},\mathcal{E}$. Proofs and more details concerning the following statements can be found in Section 10 of Van der Vaart (1998). For $N\to\infty$ the information contained in $\mathcal{O}$ increases, the influence of $p(\theta)p(\mathcal{E}|\theta)$ is diminished, and the posterior is driven by the observations. The Bernstein-von Mises Theorem states asymptotic normality of $\theta|\mathcal{O},\mathcal{E}$ for $N\to\infty$ if $p(\theta)p(\mathcal{E}|\theta)$ is smooth and positive in a neighborhood of the true parameter. Therefore, $\theta|\mathcal{O},\mathcal{E}$ and thus also $\hat{\theta}$ converge in probability with rate $1/\sqrt{N}$. Furthermore, Bayesian point estimators are asymptotically efficient and asymptotically equivalent to maximum likelihood estimators.

Modelling the prior with a shifted Beta distribution yields a smooth and positve prior $p(\theta)$ on Θ. Parametrizing the Beta distribution in terms of its mean to get the conditional expert's density $g_k(\varphi_k|\theta)$ also satisfies the smoothness conditions. We will later choose C^E as a Frank copula, which implies that $p(\theta)p(\mathcal{E}|\theta)$ satisfies the conditions of the Bernstein-von Mises Theorem.

3. Assessing the Prior Distribution

This section describes the methodology to set the prior density $p(\theta)$ used in (2.1). In all relevant cases considered in the paper we have that Θ is an interval.

Suppose we can infer a point estimate $\hat{\theta}_p$ of θ (p for prior) from the prior source of information, e.g. regulatory guidelines. We then propose to model $p(\theta)$ with a shifted Beta distribution with mean $\mathbb{E}[\theta] = \hat{\theta}_p$ and support Θ. The shifted Beta distribution is very suitable as it can take on a wide range of shapes and means, yet with only two parameters. The variance $\mathrm{var}(\theta)$ determines the credibility which is given to $\hat{\theta}_p$. We propose to estimate $\mathrm{var}(\theta)$ from the prior source of information or, alternatively, through assessing the subjective confidence that is given to the estimate $\hat{\theta}_p$.

In case no prior belief is available then $p(\theta)$ can be set as uninformative, i.e., uniform on Θ. For an uninformative prior, the posterior distribution depends mainly on $\mathcal{O}$ and $\mathcal{E}$. See Price and Manson (2002) for an introduction to uninformative priors.

The following list gives three possibilities for the prior source of information.

(1) **Regulatory guidelines.** Some insurance regulators publish reference values for the correlation between certain risk types. See for instance 9.2 in CEIOPS (2010) for the proposals of the European Union regulators or Section 8.4 in FOPI (2006) for directives from the Swiss regulators.

(2) **Physically similar situations.** Analogous to the proposal in Lambrigger et al. (2007), $\hat{\theta}_p$ can be taken as the known degree of dependence $\rho(X_1^*, X_2^*)$ of two random variables X_1^* and X_2^* whose dependence is similar (at least in nature) to the dependence between X_1 and X_2. This is related to credibility theory, where collective data are used as a starting point to estimate individual parameters. For instance, Schedule P data from US insurers could be used to calculate a prior estimate of the dependence between claims reserves in different lines of business.

(3) **Expert Judgement.** An expert can either estimate $\hat{\theta}_p$ according to his personal belief or in order to incorporate an artificial bias, for instance by putting weight on high degrees of dependence in order to avoid underestimating dependence.

4. The Elicitation of Expert Opinion

As observations are lacking or sparse in many statistical problems, expert opinion is increasingly recognized as an important source of additional information. In an insurance context, it is important to guarantee a reliable and robust expert judgement process which is credible from the point of view of insurance, regulator and rating agencies. Therefore, this section outlines some psychological and procedural principles which are necessary to turn expert opinion into scientifically meaningful statements. For an overview of the recent literature on the use of expert opinion see Meyer and Booker (2001), Ouchi (2004) or Clemen and Winkler (1999).

The task of assessing dependence using experts has received little attention. Böcker et al. (2010) model the uncertainty in correlation matrices using expert judgement. The approach to directly elicit the value of a dependence measure has been subjected to criticism as well as appraisal, see Morgan and Henrion (1992) and Clemen et al. (2000), respectively.

Our approach of asking experts for estimates of θ and applying Bayesian inference to combine the estimates represents a so-called mathematical approach. In contrast, in behavioral approaches, experts interact and agree on a common conclusion by means of discussions and other forms of interaction. However, behavioral approaches have the disadvantage that they are prone to be influenced by dominant personalities, they can suffer from the limited participation of less confident experts, and there is a general tendency to reach a conclusion too fast, see Mosleh et al. (1988) and Daneshkhah (2004).

In order to comprehensively understand a process involving expert judgment, one must understand the psychological effects involved when experts assess probabilistic quantities and make judgements under uncertainty. For instance, experts may not be familiar with describing their beliefs in terms of probabilities. Two large research streams tried to describe these effects: the *cognitive models* and the *heuristics and biases* approach, see Kynn (2008) and the references therein for an overview and critique.

We refrain from giving a review on abstract psychological aspects. Instead, we provide the following examples in order to illustrate some of the psychological effects that can influence experts in the assessment of probabilistic quantities.

- In a study described by Kahneman and Tversky (1982), people were asked: "*Linda is 31 years old, single, outspoken, bright and majored in philosophy. She is deeply concerned with issues of discrimination and social justice. Which is more likely? (i) Linda is a bank teller or (ii) Linda is a bank teller who is active in the feminist movement.*" Most answers rank (ii) to be more probable than (i), not considering that (ii) must have a probability less or equal than (i) because (ii) is a subset of (i).
- According to Eddy (1982), doctors tend to confuse $\mathbb{P}(\textit{positive test}\,|\,\textit{disease})$ (the test sensitivity) with $\mathbb{P}(\textit{disease}\,|\,\textit{positive test})$ (the power of the test).
- Kahneman and Tversky (1973) find that people tend to ignore prior probabilities when judging conditional probabilities. The question "*Is a meticulous, introverted, meek and solemn person more likely to be engaged as a librarian or as a salesman?*" was mostly answered with librarian as the stereotype of a librarian better suits the characteristics of this person. However, this answer ignores the fact that there are many more salesmen than librarians.
- The probability assigned to an event increases with the amount of details that are given to describe it. For instance, the estimated probability that a

person dies due to a natural cause is usually smaller than the sum of the separately estimated probabilities for heart disease, diabetes and other natural causes, see O'Hagan et al. (2006).

The general goal of applying expert elicitation procedures is to allow decisions to be taken in a rational manner and to be perceived to be as such. To that end, Cooke (1991) recommends to adhere to the following five principles.

(1) **Reproducibility.** All data must be open to qualified reviewers and results must be reproducible in order to allow revision from auditors or regulators.
(2) **Accountability.** Questionnaires are stored and each opinion can be linked to the corresponding expert.
(3) **Empirical control.** There should be in principle the possibility to verify expert opinion on the basis of measurable observations.
(4) **Neutrality.** There must not exist any incentives (such as a change in reputation or salary) for the experts to give answers different from their true honest opinion.
(5) **Fairness.** Experts are not discriminated or given smaller weights due to reasons that cannot be justified through the mathematical model.

More detailed suggestions and guidelines are given in Cooke and Goossens (2000).

5. The Bayesian modeling of the expert assessments

We will now address the mathematical modeling of the expert assessments $\mathcal{E}$. By the assumptions in Proposition 2.3, $p(\mathcal{E}|\theta)$ reads as

$$p(\mathcal{E}|\theta) = c^E\big(G_1(\varphi_1|\theta), \ldots, G_K(\varphi_K|\theta)\big) \prod_{k=1}^{K} g_k(\varphi_k|\theta),$$

where the $G_k(\cdot|\theta)$ describe the experts conditional distribution and c^E is the the density of the copula that represents the dependence between experts.

As we believe our experts to be correct, on average, we model the expert estimates to be conditionally unbiased, i.e. $\mathbb{E}[\varphi_k|\theta] = \theta$ for all $\theta \in \Theta$, as it is also done in Lambrigger et al. (2007). To reflect experts' uncertainty we assign each expert a variance σ_k^2, $k = 1, \ldots, K$, which is assumed to be independent of θ : $\mathrm{var}(\varphi_k|\theta) = \sigma_k^2$ for all $\theta \in \Theta$. It remains to fit a conditional density which attains these moments. Due to the versatility and the explicit expressions for moments, and because Θ is assumed to be an interval, we use the (shifted) Beta distribution, see Appendix A. Other distributions on intervals such as the Kumaraswamy, triangular and raised cosine distributions were tested. However, these have complicated expressions for moments and/or give zero weight to large parts of Θ, hence, from a modeling point of view they were found to be inferior to the Beta distribution.

As proposed in Jouini and Clemen (1996), we use the Frank copula for C^E:

$$C^E(G_1(\varphi_1|\theta), \ldots, G_K(\varphi_k|\theta)) = \frac{-1}{\varphi} \ln\left(1 + (e^{-\varphi} - 1)^{1-K} \prod_{k=1}^{K} (e^{-\varphi G_k(\varphi_k|\theta)} - 1)\right).$$

The Frank copula has a single parameter $\phi \in (0, \infty)$ and an analytic density. It is radially symmetric, which means that joint high expert assessments behave like joint low expert assessments. The parameter ϕ can be set by fixing Kendall's tau ρ_τ, which is equal for all pairs of experts assessments φ_i and φ_j. The relation between ρ_τ and θ is given by $\rho_\tau(\varphi_i, \varphi_j|\theta) = 1 - 4\phi^{-1} + 4\phi^{-2}\int_0^\phi t/(\exp(t) - 1)\,\mathrm{d}t$.

Note that the assumptions on variance and copula of the expert assessments determine the amount of information that is contained in $\mathcal{E}$, hence we do not have to further specify the relative weights between the observations and the experts.

It remains to estimate ϕ and the σ_k^2. We first show three possible approaches to calculate estimates $\widehat{\sigma_k^2}$ of σ_k^2, where the most suitable approach (or a combination of those) must be chosen according to the situation at hand.

(1) **Homogeneous experts.** If experts are assumed to have equal uncertainty, i.e. $\sigma_k^2 = \sigma^2$ for $k = 1, \ldots, K$, we may estimate

$$\widehat{\sigma^2} = \frac{1}{K-1} \sum_{k=1}^{K} (\varphi_k - \bar{\varphi})^2,$$

where $\bar{\varphi} = \frac{1}{K}\sum_{k=1}^{K} \varphi_k$. This approach is also advocated by Lambrigger et al. (2007).

(2) **Seed variables.** Suppose we have a number H of seed variables. Seed variables are values $\kappa_0^{(h)}$, $h = 1, \ldots, H$, which are known to the person doing the elicitation but not known to the experts. The experts are then asked to provide estimates $\kappa_k^{(h)}$, $k = 1, \ldots, K$ of $\kappa_0^{(h)}$. By again assuming that the experts are unbiased and that their uncertainty in estimating the seed variables is the same as the uncertainty in estimating θ, we can estimate σ_k^2 by

$$\widehat{\sigma_k^2} = \frac{1}{H} \sum_{h=1}^{H} \left(\kappa_k^{(h)} - \kappa_0^{(h)}\right)^2, \quad k = 1, \ldots, K.$$

Ideal seed variables for our situation are given by $\kappa_0^{(h)} = \rho(Y_1^h, Y_2^h)$ for some random vectors (Y_1^h, Y_2^h), $h = 1, \ldots, H$, which also lie in the field of expertise of the expert. For instance, the (Y_1^h, Y_2^h) could be random vectors that are similar in nature to (X_1, X_2), but for which much more observations are available to estimate $\rho(Y_1^h, Y_2^h)$.

(3) **Subjective variances.** The σ_k^2 can be estimated through any technique deemed feasible, e.g. through the number of years of the experts' experience or through subjective judgment, see Gokhale and Press (1982). This includes a weighted average between approaches 1, 2 and 3. Winkler (1968) proposes

to let experts provide an estimate of their uncertainty themselves, which is however criticized in Cooke et al. (1988), as experts tend to be too optimistic.

Finally, we also have to determine ϕ, the copula parameter that determines the dependence between experts. In most cases, sufficient statistical data is not available to estimate ϕ and the actuary has to resort to his personal judgement, analogous to approach 3 above for the estimation of the σ_k^2. The following aspects increase the dependence between experts and should be considered:

- Experts often come from similar professional environments and share sources of information. They may know the same historic data and predictions of the future.
- With respect to probability theory and related concepts, experts may have been exposed to the same learning methods, terminology, and misconceptions.
- The use of identical elicitation procedures and questionnaires for several experts can increase the risk of common misunderstandings.

However, according to Kallen and Cooke (2002), experts are in general less dependent than implied by the heuristics above.

6. Multivariate copula inference

This section extends our method to a multivariate setting including uncertain marginal distributions. We assume that the (possibly multidimensional) copula parameter $\boldsymbol{\theta}$ as well as the parameters of the marginal distributions X_i, $i = 1, \ldots, d$, are uncertain. In order to make Bayesian inference feasible, we make similar assumptions on the available information as in Section 2.

We will assume that the copula of interest lies in a class of copulas which can be parameterized through the value of the dependence measure ρ for a given set of pairs of margins.

Definition 6.1. *Let $C_d = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ denote a family of absolutely continous copulas $C_{\boldsymbol{\theta}} : [0,1]^d \to [0,1]$ with a parameter vector $\boldsymbol{\theta} = \{\theta^{i,j} \in \mathbb{R} : (i,j) \in \mathbf{I}\}$, where $\mathbf{I} \subset \{(i,j) : 1 \leq i < j \leq d\}$. The copulas $C_{\boldsymbol{\theta}}$ are parametrized such that $\theta^{i,j}$ is equal to the degree of dependence between margin i and j, i.e.*

$$\mathbb{P}[U_1 \leq u_1, \ldots, U_d \leq u_d] = C_{\boldsymbol{\theta}}(u_1, \ldots, u_d) \Leftrightarrow \rho(U_i, U_j) = \theta^{i,j} \textit{ for all } (i,j) \in \mathbf{I}.$$

We denote with $c(\cdot|\boldsymbol{\theta})$ the density and with $\Theta \subset \mathbb{R}^{|\mathbf{I}|}$ the set of admissible parameters $\boldsymbol{\theta}$.

The set $\mathbf{I}$ denotes the pairs of margins whose degree of dependence can be directly controlled by $\boldsymbol{\theta}$. A natural example for C_d are all elliptic copulas for which the characteristic generator is fixed, such as the Gaussian copula or the t-copula with fixed degrees of freedom. Note that also one-parameter Archimedean copulas (Clayton, Gumbel etc.) are contained in $\mathcal{C}_d$, indeed in this case it is sufficient to know the dependency between one pair of margins, which then fully defines the multivariate copula, thus $\mathbf{I} = \{(1,2)\}$. Also more exotic families like nested Archimedean copulas are contained in $\mathcal{C}_d$.

Situation 6.2. *The following three sources of information are given.*

(1) *A set $\mathcal{O}$ of N independent observations $(X_{1,n}, \ldots, X_{d,n})$, $n = 1, \ldots, N$, of $(X_1, \ldots, X_d)$.*
(2) *From K experts, a set $\mathcal{E}$ of point estimates $\varphi_k^{i,j}$ of $\rho(X_i, X_j)$, $k = 1, \ldots, K$, $(i,j) \in \mathbf{I}$.*
(3) *An additional, prior, source of information on the joint distribution of $(X_1, \ldots, X_d)$.*

We will assume that, conditionally, the distribution of $(X_{1,n}, \ldots, X_{d,n})$ is given through a copula $C \in \mathcal{C}_d$ and margins $F_{\psi_i} \in \mathcal{F}_i$, where the $\mathcal{F}_i = \{F_{\psi_i} : \psi_i \in \Psi_i\}$, $i = 1, \ldots, d$ denote families of univariate absolutely continuous distributions with a parameter set $\Psi_i \subset \mathbb{R}^r$, $r \in \mathbb{N}$ and density $f_i(\cdot \mid \psi_i)$. We will also model the parameters ψ_i of the marginals in a Bayesian framework, which allows to incorporate the uncertainty on the marginal distributions.

The following proposition extends Proposition 2.3 and gives a method to estimate the distribution of $(X_{1,n}, \ldots, X_{d,n})$ by calculating a joint posterior distribution of copula and marginal parameters. This Bayesian method uses all information given in Situation 6.2.

Proposition 6.3. *Suppose Situation 6.2 holds along with the following assumptions.*

B1 *Conditionally on the parameters $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d) \in \Psi_1 \times \cdots \times \Psi_d$, $(X_1, \ldots, X_d)$ has a distribution given by the copula $C = C_{\boldsymbol{\theta}} \in \mathcal{C}_d$ and margins $F_{\psi_i} \in \mathcal{F}_i$, $i = 1, \ldots, d$:*

$$\mathbb{P}[X_1 \leq x_1, \ldots, X_d \leq x_d \mid \boldsymbol{\theta}, \boldsymbol{\psi}] = C_{\boldsymbol{\theta}}(F_{\psi_1}(x_1), \ldots, F_{\psi_d}(x_d)).$$

B2 *The prior source of information can be translated into a prior density $p(\boldsymbol{\theta}, \boldsymbol{\psi}) : \Theta \times \Psi_1 \times \cdots \times \Psi_d \to [0, \infty)$. The vector $\boldsymbol{\theta}$ and the $\psi_1, \ldots, \psi_d$ are unconditionally independent with a density*

$$(\boldsymbol{\theta}, \psi_1, \ldots, \psi_d) \sim p(\boldsymbol{\theta}) \prod_{i=1}^{d} p(\psi_i).$$

B3 *Conditionally on $\boldsymbol{\theta}$, expert point estimates $\varphi_k^{i,j}$ have a joint distribution given through the copula model*

$$\mathbb{P}\big[\varphi_k^{i,j} \leq x_k^{i,j},\ 1 \leq k \leq K, (i,j) \in \mathbf{I} \,\big|\, \boldsymbol{\theta}\big] = C^E\big(G_k^{i,j}(x_k^{i,j} \mid \boldsymbol{\theta}),\ 1 \leq k \leq K, (i,j) \in \mathbf{I}\big),$$

with a absolutely continuous copula $C^E : [0,1]^{K|\mathbf{I}|} \to [0,1]$ and margins $G_k^{i,j}(\cdot \mid \boldsymbol{\theta})$, where the densities are denoted by c^E and $g_k^{i,j}(\cdot \mid \boldsymbol{\theta})$, respectively. $\mathcal{E}$ is independent of $\boldsymbol{\psi}$.

B4 *Conditionally on $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, $\mathcal{O}$ and $\mathcal{E}$ are independent.*

Then the posterior density $p(\boldsymbol{\theta},\boldsymbol{\psi}\,|\,\mathcal{O},\mathcal{E})$ of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ given $\mathcal{O}$ and $\mathcal{E}$ satisfies

$$p(\boldsymbol{\theta},\boldsymbol{\psi}\,|\,\mathcal{O},\mathcal{E}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{d} p(\varphi_i) \prod_{n=1}^{N} \left(c\Big(F_{\psi_1}(X_{1,n}), \ldots, F_{\psi_d}(X_{d,n})\,|\,\boldsymbol{\theta}\Big) \prod_{i=1}^{d} f_i(X_{i,n}\,|\,\psi_i) \right)$$
$$\times\; c^E\Big(\big(G_k^{i,j}(\varphi_k^{i,j}\,|\,\boldsymbol{\theta})\big)_{\substack{1\le k\le K\\(i,j)\in\mathbf{I}}}\Big) \prod_{\substack{1\le k\le K\\(i,j)\in\mathbf{I}}} g_k^{i,j}(\varphi_k^{i,j}\,|\,\boldsymbol{\theta}). \qquad (6.1)$$

Proof. The proof is analogous to the proof of Proposition 2.3. By Bayes' Theorem, we have $p(\boldsymbol{\theta},\boldsymbol{\psi}\,|\,\mathcal{O},\mathcal{E}) \propto p(\boldsymbol{\theta},\boldsymbol{\psi})p(\mathcal{O},\mathcal{E}\,|\,\boldsymbol{\theta},\boldsymbol{\psi})$. Recalling B2 and B4, we get

$$p(\boldsymbol{\theta},\boldsymbol{\psi}\,|\,\mathcal{O},\mathcal{E}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{d} p(\psi_i) p(\mathcal{O}\,|\,\boldsymbol{\theta},\boldsymbol{\psi}) p(\mathcal{E}\,|\,\boldsymbol{\theta},\boldsymbol{\psi}).$$

For the conditional densities of $\mathcal{O}$ and $\mathcal{E}$, we have by B1 and B3 that

$$p(\mathcal{O}\,|\,\boldsymbol{\theta},\boldsymbol{\psi}) = \prod_{n=1}^{N} \left(c\Big(F_{\psi_1}(X_{1,n}), \ldots, F_{\psi_d}(X_{d,n})\,|\,\boldsymbol{\theta}\Big) \prod_{i=1}^{d} f_i(X_{i,n}\,|\,\psi_i) \right)$$
$$p(\mathcal{E}\,|\,\boldsymbol{\theta},\boldsymbol{\psi}) = c^E\Big(\big(G_k^{i,j}(\varphi_k^{i,j}\,|\,\boldsymbol{\theta})\big)_{\substack{1\le k\le K\\(i,j)\in\mathbf{I}}}\Big) \prod_{\substack{1\le k\le K\\(i,j)\in\mathbf{I}}} g_k^{i,j}(\varphi_k^{i,j}\,|\,\boldsymbol{\theta}). \qquad \square$$

We make the following remarks about Proposition 6.3.

To set $p(\mathcal{E}\,|\,\boldsymbol{\theta},\boldsymbol{\psi})$, we use a similar modeling approach as proposed in Section 5. Conditionally, $\varphi_k^{i,j}\,|\,\boldsymbol{\theta}$ is modeled with a shifted Beta distribution with mean $\mathbb{E}[\varphi_k^{i,j}\,|\,\boldsymbol{\theta}] = \theta^{i,j}$. For the variance, we assume $\mathrm{var}(\varphi_k^{i,j}\,|\,\boldsymbol{\theta}) = \sigma_k^2$, which implies that the uncertainty for different assessments of a specific expert are equal.

The copula c^E describes the dependency between all experts assessments $\varphi_k^{i,j}$ for $(i,j)\in\mathbf{I}$, $k = 1, \ldots, K$. Hence, compared to the situation in Proposition 2.3, c^E does not only describe the dependence between experts, but also the dependence between the assessments of one specific expert. Again, we propose to model c^E with a Frank copula, calibrated according to the proposals in Section 5. If the dependence between experts is deemed different to the dependence between assessments of one single expert, also more complex dependence structures can be used, for instance through nested Archimedean copulas, see Hofert (2010).

We suggest to set the prior $p(\boldsymbol{\theta})$ as done in Böcker et al. (2010),

$$p(\boldsymbol{\theta}) = \prod_{(i,j)\in\mathbf{I}} p^{i,j}(\theta^{i,j}) 1_{\{\boldsymbol{\theta}\in\Theta\}},$$

where the marginal priors $p^{i,j}(\theta^{i,j})$ are modeled as described in Section 3 and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. In most cases where $|\mathbf{I}| > 1$, the parameter set Θ is not a product space, hence the term $\mathbb{1}_{\{\boldsymbol{\theta}\in\Theta\}}$. For instance, in the case of elliptic copulas, $\boldsymbol{\theta}$ must induce a positive definite correlation matrix.

The prior densities $p(\psi_i)$ for the margins can be set using the classical methods from univariate Bayesian inference, see Bühlmann and Gisler (2005). Of course, also the information on the marginal parameters ψ_i can be complemented with expert judgement, as done in Lambrigger et al. (2007). We refrain from doing so here in order to keep notation simple.

The following corollary covers the simpler case where the marginal distributions are certain, in the sense that the true marginal parameters $\psi_0 = (\psi_{1,0}, \ldots, \psi_{d,0})$ are known, thus $\mathbb{P}[X_i \le x_i] = F_{\psi_{i,0}}(x_i)$.

Corollary 6.4. *Under the assumptions of Proposition 6.3, and with* $\mathbb{P}[\boldsymbol{\psi} = \boldsymbol{\psi}_0] = 1$, *we have*

$$\begin{aligned} p(\boldsymbol{\theta}|\mathcal{O},\mathcal{E}) \propto\; & p(\boldsymbol{\theta}) \prod_{n=1}^{N} c\Big(F_{\psi_1}(X_{1,n}), \ldots, F_{\psi_d}(X_{d,n})|\boldsymbol{\theta}\Big) \\ & c^E\Big(\big(G_k^{i,j}(\varphi_k^{i,j}|\boldsymbol{\theta})\big)_{\substack{1\le k\le K \\ (i,j)\in\mathbf{I}}}\Big) \prod_{\substack{1\le k\le K \\ (i,j)\in\mathbf{I}}} g_k^{i,j}(\varphi_k^{i,j}|\boldsymbol{\theta}). \end{aligned} \tag{6.2}$$

Proof. Analogous to the proof of Proposition 6.3. □

7. Case study: An analysis of the Danish fire dataset

In this section, we give a case study investigating the dependence between monthly losses to buildings and losses to tenants due to industrial fire. We use Proposition 6.3 within an empirical Bayesian approach to infer the parameters of copula and margins.

To obtain observations $\mathcal{O}$, we use the well known multivariate dataset of Danish industrial fire insurance losses[1], which has been analyzed in several actuarial papers. The dataset contains 2167 single fire insurance losses over the period from 1980 to 1990. The losses can be split into the three parts 'building', 'contents' and 'profits'. The values are in millions of Danish Krone and inflated to 1985 values. From a risk management perspective, insurance companies are mainly interested in the aggregate losses per time period, thus we investigate monthly losses instead of single losses. More than 70% losses in the category profits are zero and of smaller magnitude than the category contents. Other publications using this dataset circumvent this problem by

[1] Available at http://www.ma.hw.ac.uk/~mcneil/data.html, retrieved on October 17, 2011.

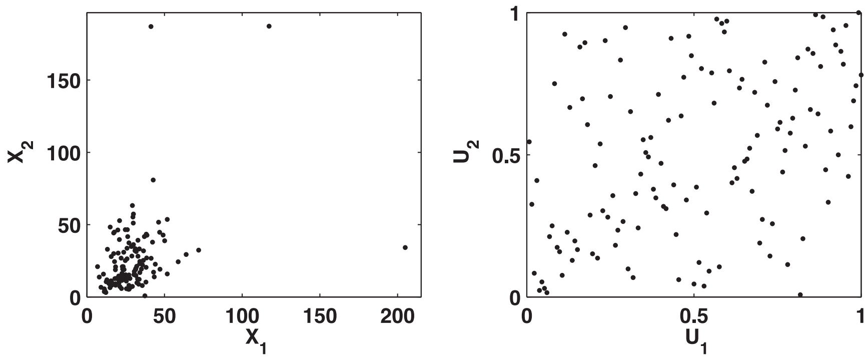

FIGURE 1: The left plot shows the $\mathcal{E}$, the 132 observations of (X_1, X_2). The right plot shows the associated empirical copula pseudo-observations, i.e., the rescaled rank order statistics of $\mathcal{E}$.

entirely removing the datapoints with zero entries, see Haug et al. (2011). Instead, we proceed by forming a bivariate dataset from the original data. The first component (X_1) represents aggregate monthly losses to buildings and the second component (X_2) represents aggregate monthly losses to tenants, which is the sum of losses to contents and profits. The resulting dataset has size $N = 132$ and does not have any zero components. The observations as well as the associated empirical copula pseudo-observations are shown in Figure 1.

As a source for prior information we use Hall (2010), which is a study on fire losses in the US. It estimates the correlation between losses to buildings and losses due to business interruption by 20%.

As we are interested in joint extreme large events, we use the upper asymptotic tail dependence $\rho(X_1, X_2) = \lim_{u\uparrow 1} \mathbb{P}[F_2(X_2) \geq u \,|\, F_1(X_1) \geq u]$ as dependence measure. The elicitation of $\rho(X_1, X_2)$ is indeed feasible. Experts can estimate the "non-asymptotic tail dependence" $\mathbb{P}[X_2$ is extremely large $|\, X_1$ is extremely large] as an approximation of $\rho(X_1, X_2)$ as follows:

(1) Predict all non-negligible causes for X_1 to be extremely large, denoted by $event_j$, $j = 1, \ldots, J$. Of course, also causes without historic evidence need to be considered.
(2) Estimate $\mathbb{P}[event_j \,|\, X_1$ is extremely large] ($j = 1, \ldots, J$), i.e. the likelihood that $event_j$ is the cause if X_1 is known to be extremely large. Roughly speaking, these likelihoods are merely weights that sum to one and indicate the importance for each $event_j$.
(3) Estimate $\mathbb{P}[X_2$ is extremely large $|\, event_j]$ ($j = 1, \ldots, J$), i.e. the likelihood that X_2 is also strongly affected given that one knows that $event_j$ happens.

Finally, by the law of total probability, the expert's answer to the question "*Given that an extremely bad outcome is observed in X_1, what is your estimate of the probability that X_2 will experience an extremely bad outcome?*" is given by

$$\varphi_k = \sum_{j=1}^{J} \mathbb{P}[X_2 \text{ is extremely large} \,|\, event_j] \cdot \mathbb{P}[event_j \,|\, X_1 \text{ is extremely large}] \approx \rho(X_1, X_2).$$

Note that this approach does not require the potentially very difficult task to estimate the probabilities $\mathbb{P}[event_j]$, $\mathbb{P}[X_1$ is extremely large] or $\mathbb{P}[X_2$ is extremely large].

We have asked four actuaries with experience in industrial fire insurance to estimate $\rho(X_1, X_2)$ through the above approach. They identified five ($J = 5$) possible causes for X_1 being extremely large:

- $event_1$: A single, accidentially caused very large fire.
- $event_2$: A large number of small, unrelated fires.
- $event_3$: A sequence of arsons by a pyromaniac.
- $event_4$: Terrorism causing either one large loss or several smaller losses.
- $event_5$: A large number of fires due to riot and civil unrest.

Let $A_{k,j}$ and $B_{k,j}$ denote the estimates of the probabilities $\mathbb{P}[event_j \,|\, X_1$ is extremely large] and $\mathbb{P}[X_2$ is extremely large $|\, event_j]$, respectively, by the k-th expert. These estimates, as well as the resulting estimates $\varphi_k = \sum_{j=1}^{5} A_{k,j} B_{k,j}$ of $\rho(X_1, X_2)$ are given in Table 1. In order to estimate the experts variance, we use the assumption of homogeneous experts and get an estimate $\widehat{\sigma^2} = 0.02354$ through method (1), as shown in Section 5. As proposed in Jouini and Clemen (1996) we use a Frank copula to capture the dependence between experts. We calibrate the copula to a Kendall's tau of 0.32 ($\phi = 3.1477$), as Kallen and Cooke (2002) find in one of their datasets an average correlation of 0.32 between experts.

TABLE 1

THE ESTIMATED PROBABILITIES $A_{k,j} = \mathbb{P}[EVENT_j \,|\, X_1$ IS EXTREMELY LARGE] AND $B_{k,j} = \mathbb{P}[X_2$ IS EXTREMELY LARGE $|\, EVENT_j]$ FOR EACH EXPERT $k = 1, 2, 3, 4$.
THE ESTIMATE φ_j OF $\rho(X_1, X_2)$ BY THE j-TH EXPERT IS GIVEN BY $\varphi_k = \sum_{j=1}^{5} A_{k,j} B_{k,j}$.

	Expert 1		Expert 2		Expert 3		Expert 4	
	$A_{1,j}$	$B_{1,j}$	$A_{2,j}$	$B_{2,j}$	$A_{3,j}$	$B_{3,j}$	$A_{4,j}$	$B_{4,j}$
$event_1$ (single large fire)	40%	80%	30%	60%	70%	95%	45%	55%
$event_2$ (extreme frequency)	30%	40%	10%	20%	20%	30%	15%	25%
$event_3$ (arson)	15%	40%	10%	35%	0%	20%	10%	40%
$event_4$ (terrorism)	5%	50%	25%	40%	5%	10%	15%	35%
$event_5$ (riot and civil unrest)	10%	30%	25%	40%	5%	20%	15%	15%
Estimate of $\rho(X_1, X_2)$	$\varphi_1 = 0.555$		$\varphi_2 = 0.435$		$\varphi_3 = 0.740$		$\varphi_4 = 0.400$	

As we want a dependence structure with upper tail dependence, but without lower tail dependence, we chose the family of Gumbel copulas, which is also used in Haug et al. (2011) and Blum et al. (2002). We examine the marginal samples through QQ-plots for different distributional families, which leads to the selection of lognormal distributions for the margins. Thus, we assume that, conditionally on the five parameters θ, μ_1, μ_2, σ_1, and σ_2, the random vector (X_1, X_2) has distribution

$$\mathbb{P}[X_1 \leq x_1, X_2 \leq x_2 \,|\, \theta, \mu_1, \mu_2, \sigma_1, \sigma_2] = C_\theta\Big(F_{\mu_1,\sigma_1}(x_1), F_{\mu_2,\sigma_2}(x_2)\Big).$$

Parametrized in terms of the upper tail dependence $\theta = \rho(X_1, X_2) \in \Theta = [0, 1]$, the Gumbel copula is given by

$$C_\theta(u_1, u_2) = \exp\left(-\left((-\ln(u_1))^{\frac{\ln(2)}{\ln(2-\theta)}} + (-\ln(u_2))^{\frac{\ln(2)}{\ln(2-\theta)}}\right)^{\frac{\ln(2-\theta)}{\ln(2)}}\right).$$

The margins are conditionally lognormal, $F_{\mu_1,\sigma_1}(x) = \mathbb{P}[X_i \leq x \,|\, \mu_i, \sigma_i] = \Phi((\ln(x) - \mu_i)/\sigma_i)$ for $i = 1, 2$, where Φ is the standard normal cdf.

Even though correlation is not the same dependence measure as $\rho(\cdot,\cdot)$, we translate the 20% correlation estimate given in Hall (2010) into a prior point estimate $\hat{\theta}_p = 0.2$. We set the prior $p(\theta)$ to be beta distributed, calibrated to a mean $\mathbb{E}[\theta] = \hat{\theta}_p = 0.2$. As prior variance we use the same estimate as for the experts, $\text{var}(\theta) = \widehat{\sigma^2} = 0.02354$, which gives the prior roughly the same weight as one expert. For the marginal parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ we use uninformative priors.

To summarize, we employ an empirical Bayes approach, in which Bayesian inference is used for the distributional parameters of (X_1, X_2). For the other parameters (variance of prior $\text{var}(\theta)$, variance of experts $\text{var}(\varphi_k \,|\, \theta)$, parameter ϕ of C^E), we use point estimates obtained from $\mathcal{O}$, $\mathcal{E}$ or other publications.

Under the assumptions of Proposition 6.3, $p(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)\, p(\mathcal{O}, \mathcal{E} \,|\, \theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$ can be calculated, which allows to simulate from $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2) \,|\, \mathcal{O}, \mathcal{E}$ through the use of the Metropolis-Hastings algorithm, as described in Robert and Casella (2005). Using a sample of size 10,000,000, we estimate posterior densities of θ, μ_i, and σ_i which are shown in Figure 2.

Suppose we are also interested in estimating the risk measure $\text{VaR}_{0.99}(X_1 + X_2)$, i.e. the 99% Value-at-Risk of the aggregate $X_1 + X_2$. The uncertainty in the

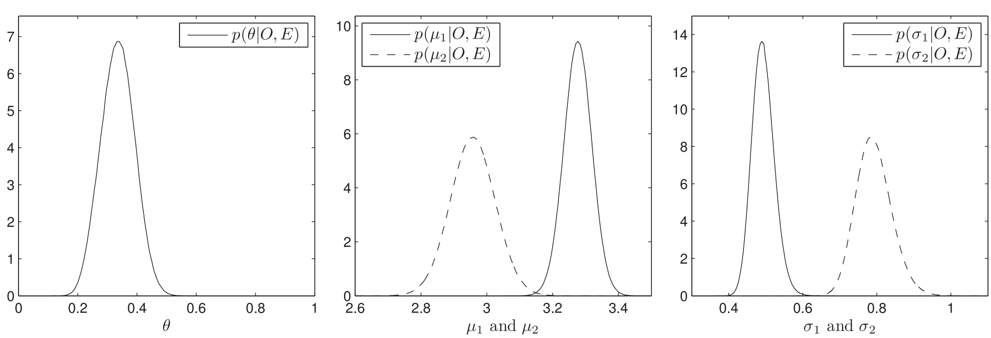

FIGURE 2: Posterior densities of $\theta, \mu_1, \mu_2, \sigma_1$, and σ_2, obtained through the Metropolis-Hastings MCMC algorithm.

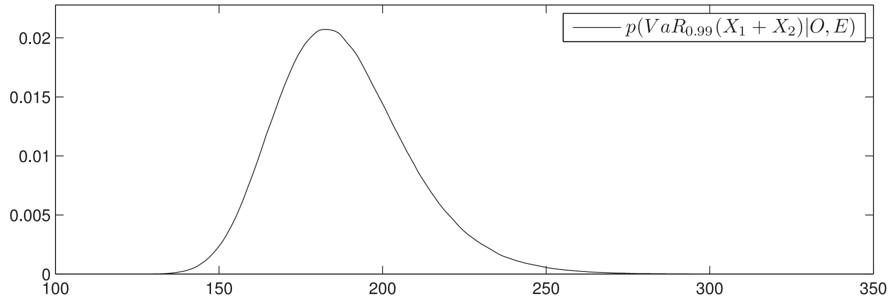

FIGURE 3: Posterior density of $\mathrm{VaR}_{0.99}(X_1 + X_2)$, estimated through nested Monte Carlo simulation.

parameters of the distribution of (X_1, X_2) naturally carries over to the uncertainty of a functional of (X_1, X_2), such as $\mathrm{VaR}_{0.99}(X_1 + X_2)$. We can estimate the posterior density $p(\mathrm{VaR}_{0.99}(X_1 + X_2)|\mathcal{O}, \mathcal{E})$ though nested Monte Carlo simulations. For each simulated parameter vector $(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$, drawn from the posterior distribution, we draw a sample of size 10,000 from $(X_1, X_2)|(\theta, \mu_1, \mu_2, \sigma_1, \sigma_2)$. This sample allows to estimate $\mathrm{VaR}_{0.99}(X_1 + X_2)$, which provides one realization of the posterior distribution of $\mathrm{VaR}_{0.99}(X_1 + X_2)|\mathcal{O}, \mathcal{E}$. Figure 3 shows the estimated density $p(\mathrm{VaR}_{0.99}(X_1 + X_2)|\mathcal{O}, \mathcal{E})$.

Table 2 summarizes point estimates and associated uncertainties of the posterior distribution of the parameters and $\mathrm{VaR}_{0.99}(X_1 + X_2)$. We also give the 90% credible interval, defined as the 5% and 95% quantile of the posterior distribution.

TABLE 2

STATISTICS OF THE POSTERIOR DISTRIBUTIONS OF θ, μ_1, μ_2, σ_1, σ_2 AND $\mathrm{VaR}_{0.99}(X_1 + X_2)$. WE GIVE POSTERIOR MEAN, STANDARD DEVIATION AND THE 90% CREDIBLE INTERVAL.

	$\mathbb{E}[\cdot|\mathcal{O}, \mathcal{E}]$	$\sqrt{\mathrm{var}(\cdot|\mathcal{O}, \mathcal{E})}$	90% credible interval
θ	0.335	0.056	[0.242,0.428]
μ_1	3.277	0.043	[3.207,3.347]
μ_2	2.957	0.068	[2.845,3.069]
σ_1	0.493	0.030	[0.447,0.544]
σ_2	0.792	0.048	[0.718,0.877]
$\mathrm{VaR}_{0.99}(X_1 + X_2)$	188.3	20.33	[158.3,224.6]

In practice, quantiles/VaRs are mostly estimated through parametric models (calibrated with maximum likelihood) or through techniques stemming from extreme value theory, such as the Peaks-over-Threshold method (POT), see McNeil et al. (2005). However, these techniques do not allow to incorporate expert opinion and confidence intervals derived from them may be of dubious quality for a small sample size. On the other hand, Bayesian inference allows a natural representation of parameter uncertainty.

8. Conclusion

Based on Bayesian inference, we propose a method to estimate the joint distribution of a random vector $(X_1, \ldots, X_d)$ by combining three sources of information, namely prior information, observations, and expert opinion. The model is based on a copula approach, which separates the models and parameters for marginals and dependence structure. Through the Bayesian approach, uncertainties in the parameters can easily be accounted for. The same holds for functionals of $(X_1, \ldots, X_d)$, for instance for a risk measure applied to $X_1 + \cdots + X_d$. The model can also be used if not all of the three sources of information are available. For instance, the model allows to estimate model parameters and their uncertainty also if no observations ($N = 0$) or no experts ($K = 0$) are present.

Our method is most helpful in situations where observations are scarce, i.e. in cases where standard methods like maximum-likelihood usually exhibit severe parameter uncertainties.

Asymptotic normality holds under mild smoothness and positivity assumptions on the prior and the conditional distribution of experts assessments. If the number of observations tends to infinity, the posterior will converge to the true value and point estimates are asymptotically as efficient as maximum likelihood estimates.

We investigated the challenging process of turning expert opinion into quantitative information. Certain principles deduced from psychological and statistical research must be adhered to in order to get reliable results. We propose procedures to assess the accuracy of the expert assessments through estimating their variance, which controls their weight in the final estimate. The Bayesian approach allows for natural interpretation of expert opinion.

Acknowledgement

The authors would like to thank Hans Bühlmann, Michel Dacorogna, Bikramjit Das, Raffaele Dell'Amore, Paul Embrechts, Christian Genest, Marius Hofert, Dominik Lambrigger, Johanna Nešlehová and the anonymous referees for their insightful comments, which led to significant improvements of the paper. As SCOR Fellow (P. Arbenz) and SCOR employee (D. Canestraro), the authors thank SCOR for financial support.

Appendix A
Definition and Properties of the Beta distribution

A random variable $Z \in (0,1)$ is *Beta distributed*, if its density is given by $f_Z(x) = x^{\alpha-1}(1-x)^{\beta-1}/\mathrm{B}(\alpha,\beta)$ for $x \in (0,1)$, where $\alpha, \beta > 0$ and $\mathrm{B}(\cdot,\cdot)$ denotes the Beta function. Mean and variance are given by $\mathbb{E}[Z] = \alpha/(\alpha+\beta)$ and $\operatorname{var}(Z) = \alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$. The parameters can be inferred from the moments through $\alpha = \mathbb{E}[Z]^2(1 - \mathbb{E}[Z])/\mathrm{var}(Z) - \mathbb{E}[Z]$ and $\beta = \alpha(\mathbb{E}[Z]^{-1} - 1)$. The beta distribution is unimodal if $\alpha, \beta \geq 1$ or, equivalently, if the variance satisfies $\mathrm{var}(Z) \leq \min\{\mathbb{E}[Z]^2(1 - \mathbb{E}[Z])/(1 + \mathbb{E}[Z]), (1 - \mathbb{E}[Z])^2\, \mathbb{E}[Z]/(2 - \mathbb{E}[Z])\}$.

The random variable $a + (b - a)\, Z \in [a, b]$ for some $a < b$ is said to have a *shifted Beta distribution* with endpoints a and b. For θ close to the boundary of Θ, a fixed conditional variance $\mathrm{var}(\varphi_k | \theta) = \sigma_k^2$ can be infeasible. For these θ, we propose to reduce $\mathrm{var}(\varphi_k | \theta)$ to the point where $\varphi_k | \theta$ is unimodal.

References

Bernardo, J. and Smith, A. (1994) *Bayesian Theory.* Wiley, Chichester.

Blum, P., Dias, A. and Embrechts, P. (2002) The art of dependence modelling: the latest advances in correlation analysis. In Lane, M., editor, *Alternative Risk Strategies*, pages 339-356. Risk Books, London.

Böcker, K., Crimmi, A. and Fink, H. (2010) Bayesian risk aggregation: Correlation uncertainty and expert judgement. In Böcker, K., editor, *Rethinking Risk Measurement and Reporting Uncertainty.* RiskBooks, London.

Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications.* Springer, Berlin.

CEIOPS (2010) QIS5 Technical Specifications. Technical report, Committee of European Insurance and Occupational Pensions Supervisors.

Clemen, R., Fischer, G. and Winkler, R. (2000) Assessing dependence: Some experimental results. *Management Science*, **46(8)**, 1100-1115.

Clemen, R. and Winkler, R. (1999) Combining probability distributions from experts in risk analysis. *Risk Analysis*, **19(2)**, 187-203.

Cooke, R. (1991) *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford University Press, New York.

Cooke, R. and Goossens, L. (2000) Procedures guide for structural expert judgement in accident consequence modelling. *Radiation Protection Dosimetry*, **90(3)**, 303-309.

Cooke, R., Mendel, M. and Thijs, W. (1988) Calibration and information in expert resolution. *Auto-matica*, **24(1)**, 87-94.

Daneshkhah, A. (2004) Uncertainty in probabilistic risk assessment: A review. Working paper, The University Of Sheffield. www.sheffield.ac.uk/content/1/c6/03/09/33/risk.pdf.

Donnelly, C. and Embrechts, P. (2010) The devil is in the tails: actuarial mathematics and the sub-prime mortgage crisis. *ASTIN Bulletin*, **40(1)**, 1-33.

Eddy, D. (1982) Probabilistic reasoning in clinical medicine: Problems and opportunities. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment under uncertainty: Heuristics and biases*, pages 249-267. Cambridge University Press, Cambridge.

Embrechts, P., McNeil, A. and Straumann, D. (2002) Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond*, pages 176-223. Cambridge University Press, Cambridge.

FOPI (2006) Technical document on the Swiss Solvency Test. Technical report, Swiss Federal Office of Private Insurance.

Genest, C., Gendron, M. and Bourdeau-Brien, M. (2009) The advent of copulas in finance. *The European Journal of Finance*, **15(7-8)**, 609-618.

Gokhale, D. and Press, S. (1982) Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Journal of the Royal Statistical Society, Series A*, **145**, 237-249.

Hall, J. (2010) *The Total Cost of Fire in the United States.* National Fire Protection Association, Quincy, Massachusetts.

Haug, S., Klüppelberg, C. and Peng, L. (2011) Statistical models and methods for dependence in insurance data. *Journal of the Korean Statistical Society*, **40**, 125-139.

Hofert, M. (2010) *Sampling Nested Archimedean Copulas with Applications to CDO Pricing.* PhD thesis, University of Ulm.

Jouini, M. and Clemen, R. (1996) Copula models for aggregating expert opinions. *Operations Research*, **44(3)**, 444-57.
Kahneman, D. and Tversky, A. (1973) On the psychology of prediction. *Psychological review*, **80(4)**, 237-251.
Kahneman, D. and Tversky, A. (1982) On the study of statistical intuitions. *Cognition*, **11**, 123-141.
Kallen, M. and Cooke, R. (2002) Expert aggregation with dependence. In *Probabilistic Safety Assessment and Management,* pages 1287-1294. International Association for Probabilistic Safety Assessment and Management.
Kynn, M. (2008) The "heuristics and biases" bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **171(1)**, 239-264.
Lambrigger, D., Shevchenko, P. and Wüthrich, M. (2007) The quantification of operational risk using internal data, relevant external data and expert opinions. *Journal of Operational Risk*, **2(3)**, 3-27.
McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, Tools.* Princeton University Press, Princeton.
Meyer, M. and Booker, J. (2001) *Eliciting and Analyzing Expert Judgment: A Practical Guide.* Society for Industrial and Applied Mathematics, Philadelphia.
Morgan, M. and Henrion, M. (1992) *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge.
Mosleh, A., Bier, V. and Apostolakis, G. (1988) A critique of current practice for the use of expert opinions in probabilistic risk assessment. *Reliability Engineering and system safety*, **20(1)**, 63-85.
Nelsen, R. (2006) *An Introduction to Copulas.* Springer, New York, 2nd edition.
O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J. and Rakow, T. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities.* Wiley, Chichester.
Ouchi, F. (2004) A literature review on the use of expert opinion in probabilistic risk analysis. Working Paper 3201, World Bank Policy Research.
Price, H. and Manson, A. (2002) Uninformative priors for Bayes' theorem. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, **617(1)**, 379-391.
Robert, C. and Casella, G. (2005) *Monte Carlo Statistical Methods.* Springer, New York.
Van der Vaart, A. (1998) *Asymptotic Statistics.* Cambridge University Press, Cambridge.
Winkler, R. (1968) The consensus of subjective probability distributions. *Management Science*, **15**, B61-B75.

Philipp Arbenz (corresponding author)
RiskLab, Department of Mathematics
Rämistrasse 101
ETH Zurich
CH-8092 Zürich, Switzerland
E-Mail: philipp.arbenz@math.ethz.ch

Davide Canestraro
SCOR SE, Zurich Branch
General Guisan – Quai 26
CH-8022 Zürich, Switzerland
E-Mail: dcanestraro@scor.com