

# Simulation Methods for Robust Risk Assessment and the Distorted Mix Approach

Sojung Kim      Stefan Weber

*Leibniz Universität Hannover*

May 14, 2021\*

## Abstract

Uncertainty requires suitable techniques for risk assessment. Combining stochastic approximation and stochastic average approximation, we propose an efficient algorithm to compute the worst case average value at risk in the face of tail uncertainty. Dependence is modelled by the distorted mix method that flexibly assigns different copulas to different regions of multivariate distributions. We illustrate the application of our approach in the context of financial markets and cyber risk.

**Keywords:** Risk management; robustness; risk measures; simulation; cyber risk.

## 1 Introduction

Capital requirements are an instrument to limit the downside risk of financial companies. They constitute an important part of banking and insurance regulation, for example, in the context of Basel III, Solvency II, and the Swiss Solvency test. Their purpose is to provide a buffer to protect policy holders, customers, and creditors. Within complex financial networks, capital requirements also mitigate systemic risks.

The quantitative assessment of the downside risk of financial portfolios is a fundamental, but arduous task. The difficulty of estimating downside risk stems from the fact that extreme events are rare; in addition, portfolio downside risk is largely governed by the tail dependence of positions which can hardly be estimated from data and is typically unknown. Tail dependence is a major source of model uncertainty when assessing the downside risk.

In practice, when extracting information from data, various statistical tools are applied for fitting both the marginals and the copulas – either (semi-)parametrically or empirically. The selection of a copula is frequently made upon mathematical convenience; typical examples include Archimedean copulas, meta-elliptical copulas, extreme value copulas, or the empirical copula, see e.g. McNeil, Frey & Embrechts (2015). The statistical analysis and verification is based on the available data and is center-focused due to limited observations from tail events. This approach is necessarily associated with substantial uncertainty. The induced model risk thus affects the computation of monetary risk measures, the mathematical basis of capital requirements. These functionals are highly sensitive to tail events by their nature – leading to substantial misspecification errors of unknown size.

---

\*Institute of Actuarial and Financial Mathematics & House of Insurance, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany. e-mail: [sojung.kim@insurance.uni-hannover.de](mailto:sojung.kim@insurance.uni-hannover.de), [stefan.weber@insurance.uni-hannover.de](mailto:stefan.weber@insurance.uni-hannover.de).

In this paper, we suggest a novel approach to deal with this problem. We focus on the downside risk of portfolios. Realistically, we assume that the marginal distributions of individual positions and their copula in the central area can be estimated sufficiently well. We suppose, however, that a satisfactory estimation of the dependence structure in the tail area is infeasible. Instead, we assume that practitioners who deal with the estimation problem share viewpoints on a collection of copulas that potentially capture extremal dependence. However, practitioners are uncertain about the appropriate choice among the available candidates.

The family of copulas that describes tail dependence translates into a family of joint distributions of all positions and thus a collection of portfolio distributions. To combine the ingredients to joint distributions, we take a particularly elegant approach: The Distorted Mix (DM) method developed by Li, Yuen & Yang (2014) constructs a family of joint distributions from the marginal distributions, the copula in the central area and several candidate tail copulas. A DM copula is capable of handling the dependence in the center and in the tail separately. We use the DM method as the starting point for a construction of a convex family of copulas and a corresponding set of joint distributions.<sup>1</sup>

Once a family of joint distributions of the positions is given, downside risk in the face of uncertainty can be computed employing a classical worst case approach. To quantify downside risk, we focus on robust average value at risk (AV@R). The risk measure AV@R is the basis for the computation of capital requirements in both the Swiss solvency test and Basel III. As revealed by the axiomatic theory of risk measures, AV@R has many desirable properties such as coherence and sensitivity to tail events, see Föllmer & Schied (2004). In addition, AV@R is m-concave on the level of distributions, see Bellini & Bignozzi (2015), and admits the application of well-known optimization techniques as described in Rockafellar & Uryasev (2000) and Rockafellar & Uryasev (2002).

To be more specific, we consider a  $d$ -dimensional random vector  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  with given marginals and a copula  $C$  in a set of distorted mix copulas  $\mathcal{DM}_{\tilde{c}}$ . Considering aggregate losses  $X = \Psi(X_1, \dots, X_d)$  for some measurable function  $\Psi$ , we study the worst case risk

$$\max_{C \in \mathcal{DM}_{\tilde{c}}} \rho(X)$$

where  $\rho$  signifies AV@R at some fixed level. The exact problem will be described in Section 2.2.

Our model setup leads to a continuous stochastic optimization problem to which we apply a combination of stochastic approximation and sample average approximation. We explain how these techniques may be used to reduce the dimension of the mixture space of copulas. We discuss the solution technique in detail and illustrate its applicability in several examples.

The main contributions of the paper are:

- (a) For a given family of copulas modeling tail dependence, we describe a DM framework that conveniently allows worst case risk assessment.
- (b) We provide an efficient algorithm that numerically computes the worst case risk and identifies worst case copulas in a lower-dimensional mixture space.
- (c) We successfully apply our framework in selected case studies. The considered examples are financial markets and cyber risk.

The paper is structured as follows. Section 2 explains the DM approach to model uncertainty and formulates the optimization problem associated to the computation of robust AV@R. In Section 3, we develop an optimization solver combining stochastic approximation (i.e., the projected stochastic gradient method) and sample average approximation: stochastic approximation identifies candidate copulas and a good approximation of the worst-case risk; in many

---

<sup>1</sup>Simple mixture approaches were previously considered in the literature, but are less flexible than the DM approach. An early contribution with an application to financial markets is Hu (2006).

cases, risk is insensitive to certain directions in the mixture space of copulas, enabling us to use sample average approximation to identify worst-case solutions in lower dimensions. Section 4 discusses two applications of our framework, namely to financial markets and cyber risk. Section 5 concludes with a discussion of potential future research directions.

## Literature

The concept of model uncertainty or robustness is a challenging topic in practice that has also been intensively discussed in the academic literature. Risk refers to situations in which possible scenarios and their associated probabilities are known; uncertainty in contrast describes circumstances in which the probabilities of events are not known, but are characterized by a nontrivial set of probability measures. This paper considers model uncertainty of this type.

The underlying key assumption concerns the structure of the considered probability measures. We assume that marginal distributions and the dependence structure of typical events is known; uncertainty arises from tail dependence. Risk is measured on the basis of a worst-case approach, a suitable algorithm is suggested and implemented in case studies. These structural assumptions on the class of probability measures are motivated by typical considerations in practice and distinguish our analysis from previous approaches in the literature.

Our methodology parallels the one chosen by Glasserman & Xu (2014) in the sense that both papers study the worst-case in a given class of models. However, Glasserman & Xu (2014) consider probability measures that are in some neighborhood of a given benchmark model in terms of relative entropy or another divergence measure. Similarly, Hu & Hong (2013) and Breuer & Csiszár (2016) study ambiguity sets defined by relative entropy. While divergence measures are not proper metrics, optimal transport costs such as Wasserstein distances are metrics under regularity conditions on the cost function; Blanchet & Murthy (2019) investigate model uncertainty in such a framework. Bartl, Drapeau & Tangpi (2020) use a similar approach and study robust optimized certainty equivalent risk measures in the context of optimal transport costs.

Another, complementary perspective on robustness comes from statistics, as suggested by Hampel (1971). By Hampel's famous theorem, the classical notion of robustness can be characterized by continuity properties of functionals with respect to the weak topology. Distribution-based convex risk measures such as average value of risk are not continuous in this sense and thus not Hampel-robust, see Cont, Deguest & Scandolo (2010) and Kou, Peng & Heyde (2013). A refined notion of Hampel robustness that corresponds to finer topologies on sets of probability measures is suggested in the seminal paper Krätschmer, Schied & Zähle (2014) and applied to risk measures; the proposal in Krätschmer et al. (2014) lifts the corresponding issues of robust statistics to a higher level that permits a comprehensive analysis.

In the current paper, we focus on worst-case AV@R in a multi-factor model. We are interested in the worst-case risk if the dependence structure is uncertain. This is closely related to papers that derive bounds in the face of partial information about dependence, cf. Embrechts, Puccetti & Rüschendorf (2013), Bernard, Jiang & Wang (2014), Bernard & Vanduffel (2015), Rüschendorf (2017), Puccetti, Rüschendorf, Small & Vanduffel (2017), Li, Shao, Wang & Yang (2018), Embrechts, Liu & Wang (2018), Weber (2018), and Hamm, Knispel & Weber (2020). In contrast to these contribution, we propose an algorithmic DM approach that is based on candidate copulas; this setting is very flexible in terms of the marginal distributions and copulas that are considered. Our framework is suitable for capturing uncertainty about the dependence in specific regions of a joint distribution. A simpler setting of mixture distributions is studied in Zhu & Fukushima (2009) and Kakouris & Rustem (2014).

Our algorithm builds on sampling-based stochastic optimization techniques. Applications of stochastic approximation and stochastic average approximation to the evaluation of risk measures were investigated by Rockafellar & Uryasev (2000), Rockafellar & Uryasev (2002), Dunkel & Weber (2007), Bardou, Frikha & Pagès (2009), Dunkel & Weber (2010), Meng, Sun

& Goh (2010), Sun, Xu & Wang (2014), and Bardou, Frikha & Pagès (2016). The contributions discuss different risk measures including AV@R and utility-based shortfall risk, efficient estimation including variance reduction, portfolio optimization and hedging but do not concentrate on model uncertainty. Without a specific focus on risk measures, the relevant simulation techniques are also discussed in Kushner & Yin (2003), Shapiro (2003), Fu (2006), Bhatnagar, Prasad & Prashanth (2013), and Kim, Pasupathy & Henderson (2015). Ghosh & Lam (2019) study worst-case approximations for performance measures in the face of uncertainty, based on stochastic approximation; their analysis does, however, not specifically consider uncertainty about dependence in different regions of multivariate distributions as captured by the DM approach in this paper.

## 2 The Distorted Mix Approach to Model Uncertainty

### 2.1 Distorted Mix Copula

Letting  $(\Omega, \mathcal{F}, P)$  be an atomless probability space, we consider the family of random variables  $\mathcal{X} = L^1(\Omega, \mathcal{F}, P)$ . The task consists in computing the risk  $\rho(X)$  of an aggregate loss random variable  $X \in \mathcal{X}$  for a risk measure  $\rho$ . A finite distribution-based monetary risk measure  $\rho : \mathcal{X} \rightarrow \mathbb{R}$  is a functional with the following three properties:

- **Monotonicity:**  $X \leq Y \Rightarrow \rho(X) \leq \rho(Y) \quad \forall X, Y \in \mathcal{X}$
- **Cash-invariance:**  $\rho(X + m) = \rho(X) + m \quad \forall X \in \mathcal{X}, m \in \mathbb{R}$
- **Distribution-invariance:**  $P \circ X^{-1} = P \circ Y^{-1} \Rightarrow \rho(X) = \rho(Y) \quad \forall X, Y \in \mathcal{X}$

We consider a specific factor structure of aggregate losses. We assume that

$$X = \Psi(X_1, \dots, X_d) \in \mathcal{X}$$

where  $\mathbf{X} = (X_1, \dots, X_d)$  is a  $d$ -dimensional random vector and  $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$  is some measurable function. The individual components  $X_i$  may depict different business lines, risk factors, or sub-portfolios, and the function  $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$  summarizes the quantity of interests. Frequently used aggregations are the total loss  $X = \sum_{i=1}^d X_i$  and the excess of loss treaty  $X = \sum_{i=1}^d (X_i - k_i)^+$  for thresholds  $k_i \in \mathbb{R}^+$ .

Computing the risk measure  $\rho(X)$  requires a complete model of the random vector  $\mathbf{X} = (X_1, \dots, X_d)$ . Let  $F(x_1, \dots, x_d)$  be its unknown  $d$ -dimensional joint distribution which we aim to understand. By Sklar's theorem, any multivariate distribution  $F$  can be written as the composition of a copula  $C$  and the marginal distributions  $F_i$  of its components:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

The typical situation in practice is as follows:

- The marginals  $F_1(x_1), \dots, F_d(x_d)$  and the dependence structure in the central area, denoted by the copula  $C_0$ , can be estimated from available data. Typical examples of  $C_0$  may include the Gaussian copula, the t-copula, or the empirical copula.
- However, due to limited observations in the tail, the copula  $C_0$  might not capture the characteristics of the extreme area very well. Instead, in the face of tail uncertainty, extreme dependence should be captured by a collection of copulas instead of a single copula. This will be explained in Section 2.2.

Before we describe our approach to model uncertainty in the next section, we introduce an important tool for combining different copulas in order to handle the central and tail parts separately, the *Distorted Mix (DM) method*, see Li et al. (2014). A DM copula  $C$  is constructed from  $m + 1$  component copulas:  $C_0$  for the typical area, and  $C_1, \dots, C_m$  for the extreme area.

**Definition 1** (Distorted mix copula)

Let  $D_{ij} : [0, 1] \rightarrow [0, 1]$  be continuous distortion functions, i.e., continuous, increasing functions with  $D_{ij}(0) = 0, D_{ij}(1) = 1$ , and  $\alpha_i \geq 0, i = 0, \dots, m, j = 1, \dots, d$ , such that

$$\sum_{i=0}^m \alpha_i = 1, \quad \sum_{i=0}^m \alpha_i D_{ij}(v) = v \quad \forall v \in [0, 1], j = 1, \dots, d. \quad (1)$$

For any collection of copulas  $C_0, \dots, C_m$ , the corresponding distorted mix copula  $C : [0, 1]^d \rightarrow [0, 1]$  is defined by

$$C(u_1, \dots, u_d) = \sum_{i=0}^m \alpha_i C_i(D_{i1}(u_1), \dots, D_{id}(u_d)). \quad (2)$$

**Remark 1**

A copula captures the dependence structure of a multivariate random vector with marginal distribution functions  $F_1, F_2, \dots, F_d$  as a function of  $u_1 = F_1(x_1), u_2 = F_2(x_2), \dots, u_d = F_d(x_d)$  with  $x_1, x_2, \dots, x_d \in \mathbb{R}$ . The argument  $x_j$  is a quantile of  $F_j$  at level  $u_j, j = 1, 2, \dots, d$ : Levels close to 0 correspond to the lower tail of  $(X_1, X_2, \dots, X_d)$ , levels close to 1 to the upper tail, and other levels to the center of the distribution.

In equation (2), for  $i = 0, \dots, m$ , the parameter  $\alpha_i$  defines the probability fraction of the total dependence that is governed by copula  $C_i$  which is distorted by the distortion functions  $D_{i1}, D_{i2}, \dots, D_{id}$ . These distortion functions describe how the arguments (or levels) of copula  $C$  are mapped to the arguments (or levels) of the ingredient copulas  $C_i$ . We illustrate these features in the following example.

**Example 1**

Let  $d = m = 2$  and  $\alpha_0 = \alpha_1 = \alpha_2 = 1/3$ . We suppose that  $C_1$  and  $C_2$  are the comonotonic copulas, i.e.,  $C_1(u_1, u_2) = C_2(u_1, u_2) = \min(u_1, u_2)$ , and that  $C_0$  is the countermonotonic copula, i.e.,  $C_0(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ . We set  $D_{ij}(u_j) = \max\{3 \cdot (u_j - a_i), 0\} \wedge 1, a_1 = 0, a_2 = 2/3, a_0 = 1/3, j = 1, 2$ . Obviously, the lower and upper tails are governed by the comonotonic copulas  $C_1$  and  $C_2$ , respectively, and the central part is countermonotonic according to  $C_0$ . In this particular example, the dependence structure in each part is exclusively controlled by one of the copulas  $C_0, C_1$ , and  $C_2$ .<sup>2</sup>

**2.2 Worst-Case Risk Assessment**

In this section, we explain our approach to risk assessment in the face of tail uncertainty. As described in the previous section, we assume that the marginals of the random vector  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  are given. Its copula is unknown, but possesses the following DM structure:

- Let  $\mathfrak{D} = \{D_{ik} : i = 0, \dots, m, k = 1, \dots, d\}$  be a collection of distortion functions and  $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_m) \in [0, 1]^{m+1}$  satisfying assumption (1).
- In addition, we fix a copula  $C_0$  and a set  $\tilde{\mathfrak{C}}$  of copulas.

We assume that the copula of  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  belongs to the following family:

$$\mathcal{DM}_{\tilde{\mathfrak{C}}} = \left\{ \alpha_0 C_0(D_{01}(u_1), \dots, D_{0d}(u_d)) + \sum_{i=1}^m \alpha_i \tilde{C}_i(D_{i1}(u_1), \dots, D_{id}(u_d)), \tilde{C}_i \in \tilde{\mathfrak{C}} \forall i = 1, \dots, m \right\}$$

The worst-case risk assessment over all feasible distributions of  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  is equal to

$$\max_{C \in \mathcal{DM}_{\tilde{\mathfrak{C}}}} \rho(X) \quad (3)$$

<sup>2</sup>The example focuses on piecewise linear distortion functions with disjoint support. This is a special case of the distorted mix model in equation 2 in which different parts of the dependence structure are each controlled by a single copula  $C_1, \dots, C_m$ . Other choices of distortion functions capture more general dependence structures.

where  $X = \Psi(X_1, \dots, X_d)$  and  $\mathbf{X} = (X_1, \dots, X_d)$  has a copula  $C \in \mathcal{DM}_{\tilde{\mathcal{C}}}$  with the given marginals.

**Remark 2**

*Our approach assumes that the marginals of  $\mathbf{X}$  and the copula  $C_0$  are known; if the distortion functions are suitably chosen,  $C_0$  could, for example, govern the central area. However, dependence in other regions, e.g. tail dependence, is uncertain and captured by some family  $\tilde{\mathcal{C}}$  of copulas. The key structural assumption is that  $\mathbf{X}$  possesses a DM copula and that the distortions and associated probability fractions are fixed. These determine the composition of the copula of  $\mathbf{X}$ . The distortions and probability fraction associated to the copula  $C_0$  cannot be varied; for all other distortions and associated probability fractions the corresponding copulas may flexibly be chosen from the collection  $\tilde{\mathcal{C}}$ .*

**Remark 3**

*One possible approach would be to choose  $\tilde{\mathcal{C}}$  as a finite collection of  $K \geq m$  candidate copulas. In this case, the number of the DM copulas is either  $\binom{K}{m} \times m!$  or  $K^m$  if we allow duplicate components. This approach has two disadvantages: First, from a technical point of view the corresponding discrete optimization problem involves a very high number of permutations. Computing the value function for each of them is expensive, and the Ranking and Selection (R&S) method would not be efficient in this case. Second, with finitely many candidate copulas also their mixtures seem to be plausible ingredients to the DM method and should not be excluded a priori.*

For a given collection  $\mathcal{C} = \{C_1, C_2, \dots, C_K\}$  of  $K$  candidate copulas, we consider the family  $\tilde{\mathcal{C}}$  of their mixtures. That is, any element of  $\tilde{\mathcal{C}}$  can be expressed as a convex combination of elements of  $\mathcal{C}$ :

$$\tilde{C}^\gamma = \sum_{j=1}^K \gamma_j C_j, \quad \gamma \in \Delta^{K-1} = \left\{ \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_K \end{pmatrix} \in \mathbb{R}^K \mid \sum_{j=1}^K \gamma_j = 1 \text{ and } \gamma_j \geq 0 \text{ for all } j \right\},$$

where  $\Delta^{K-1}$  is the standard  $K - 1$  simplex. The  $K$  vertices of the simplex are the points  $\mathbf{e}_i \in \mathbb{R}^K$ , where  $\mathbf{e}_1 = (1, 0, \dots, 0)^\top$ ,  $\mathbf{e}_2 = (0, 1, \dots, 0)^\top$ ,  $\dots$ ,  $\mathbf{e}_K = (0, 0, \dots, 1)^\top$ . With this notation, our  $K$  candidate copulas  $C_j \in \mathcal{C}$  can be written as  $C_j = \tilde{C}^{\mathbf{e}_j}$ .

Any element in  $\mathcal{DM}_{\tilde{\mathcal{C}}}$  can now be represented by some  $\bar{\gamma} = (\gamma^1, \dots, \gamma^m) \in \mathbb{R}^{K \times m}$  with  $\gamma^1, \dots, \gamma^m \in \Delta^{K-1}$  according to the following formula:

$$\hat{C}^{\gamma^1, \dots, \gamma^m}(u_1, \dots, u_d) = \alpha_0 C_0(D_{01}(u_1), \dots, D_{0d}(u_d)) + \sum_{i=1}^m \alpha_i \tilde{C}^{\gamma^i}(D_{i1}(u_1), \dots, D_{id}(u_d)). \quad (4)$$

With this notation, the optimization problem (3) can be rewritten as

$$\max_{\bar{\gamma} = (\gamma^1, \dots, \gamma^m) \in (\Delta^{K-1})^m} \rho(X^{\bar{\gamma}}) \quad (5)$$

where  $X^{\bar{\gamma}}$  represents the aggregate loss  $\Psi(X_1, \dots, X_d)$  with  $(X_1, \dots, X_d)$  having copula  $\hat{C}^{\gamma^1, \dots, \gamma^m}$  and the given marginals. We call  $\hat{C}^{\gamma^1, \dots, \gamma^m}$  in (4) a *robust DM copula* if it attains the optimal solution of (5). Optimization problem (5) enables us to search the solution inside of the multiple simplexes and paves a way to utilize the gradient approach.

We will now construct and explore a sampling-based optimization solver. For this purpose, we focus on one particular risk measure, the average value at risk (AV@R), also called conditional value at risk or expected shortfall. This risk measure forms the basis of Basel III and the Swiss

Solvency test. If  $p \in (0, 1)$  is the level of the AV@R, a number close to 1, the corresponding AV@R of the losses  $X^{\bar{\gamma}}$  is defined as

$$c_p(\bar{\gamma}) = \frac{1}{1-p} \int_p^1 q_t^-(F_{X^{\bar{\gamma}}}) dt$$

where  $q_t^-(F) = \inf\{x \in \mathbb{R} \mid F(x) \geq t\}$  for a distribution  $F$  and  $F_{X^{\bar{\gamma}}}$  stands for the distribution of  $X^{\bar{\gamma}}$ . Accordingly, we denote VaR by  $v_p(\bar{\gamma}) = q_p^-(F_{X^{\bar{\gamma}}})$ . With this notation, our optimization problem is

$$\begin{aligned} & \max_{\gamma^1, \dots, \gamma^m \in \Delta^{K-1}} c_p(\gamma^1, \dots, \gamma^m) \\ &= \max_{\bar{\gamma} \in (\Delta^{K-1})^m} c_p(\bar{\gamma}). \end{aligned} \quad (6)$$

## 2.3 Sampling Algorithm

### 2.3.1 Portfolio Vector

The factor structure of DM copulas provides the basis for adequate simulation methods (see Proposition 1, Li et al. (2014)). Samples of the copula

$$\hat{C}^{\gamma^1, \dots, \gamma^m}$$

defined in eq. (4) can be generated according to the following Algorithm 1.

---

**Algorithm 1** Sampling algorithm of the DM copula (4) generated by  $\bar{\gamma}$

---

- 1: **procedure** ROBUSTDMC( $\alpha, \bar{\gamma}, \mathcal{C}, \mathcal{D}$ )
  - 2:     **sample** a random variable  $Z^1$  distributing discretely as  $P(Z^1 = i) = \alpha_i$  for  $i = 0, \dots, m$
  - 3:     **if**  $Z^1 \neq 0$  **then**
  - 4:         **sample** a random variable  $Z^2$  distributing discretely as  $P(Z^2 = j | Z^1) = \gamma_j^{Z^1}$  for  $j = 1, \dots, K$
  - 5:     **else set**  $Z^2 = Z^1 = 0$
  - 6:     **sample** a random vector  $\mathbf{V} = (V_1, \dots, V_d)$  from the joint distribution  $C_{Z^2}$
  - 7:     **for**  $k = 1$  **to**  $d$  **do**
  - 8:          $U_k = D_{Z^1 k}^{-1}(V_k)$
  - 9:     **return**  $\mathbf{U} = (U_1, \dots, U_d)$
- 

Samples of

$$(X_1, \dots, X_d)$$

with copula  $\hat{C}^{\gamma^1, \dots, \gamma^m}$  and arbitrary marginal distributions  $F_1, F_2, \dots, F_d$  can be generated according to the quantile transformation

$$(X_1, \dots, X_d) =^d (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)).$$

### 2.3.2 Aggregate Loss

The simulation of the aggregate losses  $X^{\bar{\gamma}}$  is now based on a simple transformation. Setting

$$A(s) = \{(u_1, \dots, u_d) : \Psi(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \leq s\},$$

we define distribution functions

$$\begin{aligned} G_0(s) &= \int \mathbf{1}_{A(s)} dC_0(D_{01}(u_1), \dots, D_{0d}(u_d)), \\ G_{ij}(s) &= \int \mathbf{1}_{A(s)} dC_j(D_{i1}(u_1), \dots, D_{id}(u_d)), \quad i = 1, \dots, m, j = 1, \dots, K, \end{aligned}$$

and note that

$$F_{X^{\bar{\gamma}}}(s) := \mathbb{P}[X^{\bar{\gamma}} \leq s] = \alpha_0 G_0(s) + \sum_{i=1}^m \alpha_i \sum_{j=1}^K \gamma_j^i G_{ij}(s). \quad (7)$$

If  $\Psi^0$  and  $\Psi^{ij}$  are distributed according to  $G_0$  and  $G_{ij}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, K$ , and independent of  $Z_1$  and  $Z_2$  defined in Algorithm 1, then

$$X^{\bar{\gamma}} \stackrel{d}{=} \mathbf{1}_{[Z_1=0]} \Psi^0 + \sum_{i=1}^m \sum_{j=1}^K \mathbf{1}_{[Z_1=i, Z_2=j]} \Psi^{ij}.$$

This representation will be instrumental for our simulation algorithms. For later use, we denote the density functions of  $G_0(s)$ ,  $G_{ij}(s)$ , and  $F_{X^{\bar{\gamma}}}(s)$  by  $g_0(s)$ ,  $g_{ij}(s)$ , and  $f_{X^{\bar{\gamma}}}(s)$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, K$ , respectively, provided that they exist.

### 3 Optimization Solver

In this section, we develop an algorithm solving problem (6) that builds on two classical approaches: *Stochastic Approximation (SA)* and *Sample Average Approximation (SAA)*. While SA is an iterative optimization algorithm that is based on noisy observations, SAA first estimates the whole objective function and transforms the optimization into a deterministic problem. We combine both approaches.

The standard stochastic gradient algorithm of SA quickly approximates the worst-case risk, but the convergence to a worst-case copula is slow. It turns out that in many cases the risk is insensitive to certain directions in the mixture space of copulas. We exploit this observation in order to reduce the dimension of the problem and identify a suitable subset of  $\mathfrak{C}$  that excludes copulas whose contribution to the worst-case risk is small. We then determine a solution in the corresponding simplex, relying on SAA, which is computationally efficient in lower dimensions only, but provides a good global solution to optimization problems, even if stochastic gradient algorithms are noisy and slow.

Our method thus first applies SA to estimate worst-case risk together with a candidate mixture from which a lower-dimensional problem is constructed. Second, SAA is used, but only in the lower-dimensional mixture space – utilizing a large sample set that reduces noise.

**Step 1 – Sampling.** We generate  $N$  independent copies of the  $m \times K + 1$  random variables  $\Psi^0$  and  $\Psi^{ij}$  according to the distribution functions  $G_0$  and  $G_{ij}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, K$ , respectively.

**Step 2 – SA Algorithm.** The PSG-RobustAV@R Algorithm 2 discussed in Section 3.1 seeks a candidate solution and terminates after a small number of iterations. We design a stopping rule that determines when to move to the next step.

**Step 3 – SAA Algorithm.** From the solution obtained in Step 2 we construct a lower-dimensional simplex in which we search for a solution. We apply SAA on a suitable grid. The SAA-RobustAV@R Algorithm 4 is discussed in Section 3.2.

#### 3.1 Stochastic approximation: gradient approach

SA is a recursive procedure evaluating noisy observations of the objective function and its subgradient. The algorithm moves in the gradient direction approaching a local optimum by a first-order approach (minimization and maximization require, of course, opposite signs).



---

**Algorithm 2** The projected stochastic gradient algorithm for the robust AV@R
 

---

1: **procedure** PSG-ROBUSTAV@R

2:   **Input** the level  $p$  of AV@R, the step sizes  $\{\delta_t = t^{-a}\}_{t \geq 1}$ , the sample size sequences  $\{N_t\}_{t \geq 1}$ , the number of iterations  $M$ , the PDF of  $X^{\bar{\gamma}_t}$  at iteration  $t$  denoted by  $f_{X^{\bar{\gamma}_t}}(s)$ , and the PDFs  $g_0(s)$  and  $g_{ij}(s)$  for  $i = 1, \dots, m$  and  $j = 1, \dots, K$

3:   **Initialization:**

4:   Set a starting state  $\bar{\gamma}_1 = (\gamma_1^1, \dots, \gamma_1^m)$  with  $\gamma_1^i \in \Delta^{K-1}, i = 1, \dots, m$

5:   **while** terminal conditions are not met **do**

6:     **for**  $t = 1$  **to**  $M$  **do**

7:       **Simulation:**            $\triangleright$  generate  $(L_1, \dots, L_N)$   $N = N_t$  i.i.d. observations of  $X^{\bar{\gamma}_t}$

8:       **for**  $l = 1$  **to**  $N_t$  **do**

9:           Sample  $(U_1^l, \dots, U_d^l)$  from Algorithm 1 with  $\bar{\gamma}_t = (\gamma_t^1, \dots, \gamma_t^m)$

10:          Set  $L_l = \Psi(F_1^{-1}(U_1^l), \dots, F_d^{-1}(U_d^l))$

11:       **VaR and AV@R Estimation:**

12:       Set  $\hat{v}_p^{N_t} = L_{\lceil N_t p \rceil : N_t}$

13:       Set  $\hat{c}_p^{N_t} = \hat{v}_p^{N_t} + \frac{1}{N_t(1-p)} \sum_{i=1}^{N_t} (L_i - \hat{v}_p^{N_t})^+$

14:       **AV@R Gradient Estimation:**

15:       Set  $f_{X^{\bar{\gamma}_t}}(s) = \alpha_0 g_0(s) + \sum_{i=1}^m \alpha_i \sum_{j=1}^K \gamma_{j,t}^i g_{ij}(s)$   $\triangleright \gamma_{j,t}^i$  is the  $j$ -th component of  $\gamma_t^i$

16:       **for**  $i = 1$  **to**  $m$ , **and**  $j = 1$  **to**  $K$  **do**

17:          Set  $\Delta_{i,j}(\hat{c}_p^{N_t}) = \frac{1}{N_t(1-p)} \sum_{l=1}^{N_t} \frac{\alpha_i g_{ij}(L_l)}{f_{X^{\bar{\gamma}_t}}(L_l)} (L_l - \hat{v}_p^{N_t}) \mathbf{1}_{[L_l \geq \hat{v}_p^{N_t}]}$

18:       **Parameter Update - Multiple Simplexes Projection**

19:       **for**  $i = 1$  **to**  $m$  **do**

20:          Set  $\Delta_t^i = (\Delta_{i,1}, \dots, \Delta_{i,K})$

21:          Update  $\gamma_{t+1}^i = \Pi_{\Delta^{K-1}}(\gamma_t^i + \delta_t \Delta_t^i)$  by Algorithm 3

22:   **Output**  $\hat{c}_p^N$  and  $\gamma_t^1, \dots, \gamma_t^m$

---



---

**Algorithm 3** Euclidean projection of a vector  $y$  onto simplex
 

---

1: **procedure** PROJ $S(y)$   $\triangleright y \in \mathbb{R}^K$

2:   sort  $y$  into  $u$ :  $u_1 \geq u_2 \geq \dots \geq u_K$

3:   find  $\tau = \max\{1 \leq j \leq K : u_j + \frac{1}{j}(1 - \sum_{k=1}^j u_k) > 0\}$

4:   define  $\lambda = \frac{1}{\tau}(1 - \sum_{k=1}^{\tau} u_k)$

5:   **return**  $x$  s.t.  $x_i = \max(y_i + \lambda, 0), i = 1, \dots, K$   $\triangleright x = \Pi_{\Delta^{K-1}}(y) \in \mathbb{R}^K$

---

### 3.1.1 Projected stochastic gradient method

Algorithm 2 seeks to solve the optimization problem (6). At each iteration  $t$  the SA algorithm first generates  $N = N_t$  loss samples  $L_1, \dots, L_{N_t}$  of  $X^{\bar{\gamma}_t}$  according to Algorithm 1. SA then estimates the V@R and AV@R as follows:

$$\hat{v}_p^{N_t} = L_{\lceil N_t p \rceil : N_t}, \quad \hat{c}_p^{N_t} = \hat{v}_p^{N_t} + \frac{1}{N_t(1-p)} \sum_{i=1}^{N_t} (L_i - \hat{v}_p^{N_t})^+.$$

Here,  $\lceil a \rceil$  denotes the smallest integer larger than or equal to  $a$ , and  $L_{s:N}$  is the  $s$ -th order statistic from the  $N$  observations,  $L_{1:N} \leq L_{2:N} \leq \dots \leq L_{N:N}$ .

Second, SA computes the gradients  $\Delta_t^i = (\Delta_{i,1}, \dots, \Delta_{i,K})$  of  $c_p$  at  $\gamma_t^i$  from

$$\begin{aligned} \Delta_{i,j}(\hat{c}_p^N) &= \frac{1}{N(1-p)} \sum_{l=1}^N \frac{\partial \log f_{X^{\bar{\gamma}}}(L_l)}{\partial \gamma_j^i} (L_l - \hat{v}_p^N) \mathbf{1}_{[L_l \geq \hat{v}_p^N]} \\ &= \frac{1}{N(1-p)} \sum_{l=1}^N \frac{\alpha_i g_{ij}(L_l)}{f_{X^{\bar{\gamma}}}(L_l)} (L_l - \hat{v}_p^N) \mathbf{1}_{[L_l \geq \hat{v}_p^N]} \end{aligned} \quad (8)$$

for every  $i = 1, \dots, m$ .

Third, parameter updates are computed for each  $i$ :

$$\gamma_{t+1}^i = \Pi_{\Delta^{K-1}}(\gamma_t^i + \delta_t \Delta_t^i), \quad (9)$$

where  $\Pi_{\Delta^{K-1}}(x) = \arg \min_y \{ \|x - y\| \mid y \in \Delta^{K-1} \}$  is the Euclidean projection of  $x$  onto the simplex, and  $\{\delta_t\}_{t \geq 1}$  is the step size multiplier. This type of algorithm is called *the projected gradient descent algorithm*.

Algorithm 2, a projection onto multiple high dimensional simplexes, applies methods described in Condat (2016). In contrast to these, the simple, classical projection Algorithm 3 that we included for illustration possesses the larger complexity  $O(K^2)$ .

### 3.1.2 Convergence of SA

The convergence of Algorithm 2 is guaranteed if Assumptions 1 & 2 below are satisfied, see Theorem 5.2.1 in Kushner & Yin (2003).

**Assumption 1** (1) *The random variables  $X^{\bar{\gamma}}$  have a continuous distribution with finite support for all  $\bar{\gamma}$ .*

(2) *For all  $\bar{\gamma}$ ,  $i, j$ , the gradients  $\frac{\partial}{\partial \gamma_j^i} v_p(\bar{\gamma})$  and  $\frac{\partial}{\partial \gamma_j^i} c_p(\bar{\gamma})$  are well defined and bounded.*

(3)  *$X^{\bar{\gamma}}$  has a positive and continuously differentiable density  $f_{X^{\bar{\gamma}}}$ , and  $\frac{\partial}{\partial \gamma_j^i} \log f_{X^{\bar{\gamma}}}(s)$  exists and is bounded for all  $s$ ,  $\bar{\gamma}$ ,  $i, j$ .*

**Assumption 2** (1) *The step size sequence  $\{\delta_t\}_{t \geq 1}$  satisfies*

$$\sum_{t=1}^{\infty} \delta_t = \infty, \quad \delta_t \geq 0, \quad \sum_{t=1}^{\infty} \delta_t^2 < \infty.$$

(2)  *$\frac{\partial}{\partial \gamma_j^i} c_p(\bar{\gamma})$  is continuous, and*

$$\sum_{t=1}^{\infty} \delta_t \left| \mathbb{E} [\Delta_{i,j}(\hat{c}_p^{N_t})] - \frac{\partial}{\partial \gamma_j^i} c_p(\bar{\gamma}_t) \right| < \infty$$

*with probability 1 for each  $i$  and  $j$ .*

In specific applications, these sufficient conditions for convergence are not always satisfied. However, the SA algorithm might still produce estimates that approach a solution of the problem. We will impose a switching condition that determines when we move from SA to SAA. SAA is explained in the next section.

**Remark 4** (Concavity of AV@R)

*Algorithm 2 converges to a local maximum. But any local maximum of the problem (6) is even the global maximum, since  $c_p$  is a concave function of  $\bar{\gamma}$ , see Acciaio & Svindland (2013). This property is closely related to the  $m$ -concavity of AV@R, a concavity property on the level of distributions, see Weber (2006) and Bellini & Bigozzi (2015).*

**Remark 5** (Differentiability of AV@R)

*The gradient estimate (8) of AV@R in Algorithm 2 belongs to the Likelihood Ratio (LR) methods due to Tamar, Glassner & Mannor (2015). A LR approach is appropriate, since the distribution of the argument  $X^{\bar{\gamma}}$  of the AV@R depends on  $\bar{\gamma}$ .*

- (a) *The computation (8) needs  $g_{ij}$  as inputs. If their computation is not analytically tractable, an empirical estimator can be chosen. Other options are AEP (Arbenz, Embrechts & Puccetti (2011)) and GAEP (Arbenz, Embrechts & Puccetti (2012)).*
- (b) *An alternative to LR gradient estimation are finite differences, as applied in algorithms of KieferWolfowitz type (Kiefer & Wolfowitz (1952)). Properties of such algorithms are discussed in Brodie, Cicek & Zeevi (2011). Finite differences require less regularity in order to be applicable, but typically exhibit a worse performance.*

### 3.2 Sample average approximation

AV@R belongs to the class of divergence risk measures that coincide, up to a sign change, with optimized certainty equivalents. These admit a representation as the solution of a one-dimensional optimization problem, see Ben-Tal & Teboulle (2007). The minimizer can be characterized by a first order condition. For the specific case of AV@R this representation was previously described in Pflug (2000), Rockafellar & Uryasev (2000), and Rockafellar & Uryasev (2002), and implies the following identity:

$$c_p(\bar{\gamma}) = \min_{u \in \mathbb{R}} \left\{ u + \frac{1}{1-p} \int (L - u)^+ dF_{X^{\bar{\gamma}}}(L) \right\}.$$

The mixture representation (7) of the distribution function of  $X^{\bar{\gamma}}$  provides a reformulation of the original problem (6):

$$\max_{\bar{\gamma} \in (\Delta^{K-1})^m} \min_{u \in \mathbb{R}} \left\{ u + \frac{\alpha_0}{1-p} \mathbb{E}[\Psi^0 - u]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \mathbb{E}[\Psi^{ij} - u]^+ \right\} \quad (10)$$

where  $\Psi^0$  and  $\Psi^{ij}$  are random variables with distributions  $G_0$  and  $G_{ij}$ , respectively.

SAA algorithmically solves the stochastic optimization problem (6) by first approximating the objective function by its sample average estimate and then solving the auxiliary deterministic problem. Eq. (10) suggests the following SAA for (6):

$$\max_{\bar{\gamma} \in (\Delta^{K-1})^m} \min_{u \in \mathbb{R}} \left\{ u + \frac{\alpha_0}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^0 - u]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^{ij} - u]^+ \right\} \quad (11)$$

The SAA procedure is described in Algorithm 4.

---

**Algorithm 4** SAA for the robust AV@R
 

---

1: **procedure** SAA-ROBUSTAV@R  
 2:   **Input** the probability level  $p$  for AV@R,  $\{\Psi_k^0, \Psi_k^{ij}\}_{k=1, \dots, N}$   $N$  realizations  $\Psi^0$  and  $\Psi^{ij}$  for  $i = 1, \dots, m$  and  $j = 1, \dots, K$ ,  $\alpha$ , a grid of  $\bar{\gamma} = (\gamma_1^1, \dots, \gamma_1^m)$   
 3:   **for** every  $\bar{\gamma} = (\gamma_1^1, \dots, \gamma_1^m)$  on the grid **do**  
 4:     **Initialization:** Initialize the lower bound  $u_l$  with  $\bar{p}_N(u_l) < p$ , and the upper bound  $u_u$  with  $\bar{p}_N(u_u) > p$   
 5:     Set  $p_u = \bar{p}_N(u_u)$  and  $p_l = \bar{p}_N(u_l)$   
 6:     **Bisection method:**  
 7:     **while**  $|p_u - p| > \epsilon$  and  $|p_l - p| > \epsilon$  **do**  
 8:        $u_m = (u_u + u_l)/2$  and evaluate  $p_m = \bar{p}_N(u_m)$   
 9:       **if**  $p_m > p$  **then** set  $u_u = u_m$  and  $p_u = p_m$   
 10:       **else** set  $u_l = u_m$ , and  $p_l = p_m$   
 11:       **if**  $|p_u - p| \leq \epsilon$  **then** return  $\check{u}^N(\bar{\gamma}) = u_u$   
 12:       **else** return  $\check{u}^N(\bar{\gamma}) = u_l$   
 13:     **AV@R computation:**  
 14:     Compute  $\check{c}_p^N(\bar{\gamma}) = \check{u}^N + \frac{\alpha_0}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^0 - \check{u}^N]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^{ij} - \check{u}^N]^+$   
 15:     **Output**  $\max \check{c}_p^N(\bar{\gamma})$  on the grid

---

### 3.2.1 Inner minimization

The inner minimization in (10) can numerically be solved on the basis of first order conditions that are specified in the following lemma.

**Lemma 1**

Let  $\zeta(u, \bar{\gamma}) = u + \frac{\alpha_0}{1-p} \mathbf{E}[\Psi^0 - u]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \mathbf{E}[\Psi^{ij} - u]^+$ . Then

$$\begin{aligned} \frac{\partial^+ \zeta}{\partial u}(u, \bar{\gamma}) &= \frac{-p}{1-p} + \frac{\alpha_0}{1-p} \mathbf{P}(\Psi^0 \leq u) + \sum_{i,j} \frac{\alpha_i \gamma_j^i}{1-p} \mathbf{P}(\Psi^{ij} \leq u), \\ \frac{\partial^- \zeta}{\partial u}(u, \bar{\gamma}) &= \frac{-p}{1-p} + \frac{\alpha_0}{1-p} \mathbf{P}(\Psi^0 < u) + \sum_{i,j} \frac{\alpha_i \gamma_j^i}{1-p} \mathbf{P}(\Psi^{ij} < u). \end{aligned}$$

The minima of the function  $u \mapsto \zeta(u, \bar{\gamma})$  are attained and any minimizer  $z$  satisfies

$$\frac{\partial^- \zeta}{\partial u}(u, \bar{\gamma}) \leq 0 \leq \frac{\partial^+ \zeta}{\partial u}(u, \bar{\gamma}). \quad (12)$$

If the distribution functions of  $\Psi^0$  and  $\Psi^{ij}$  are continuous, the first order condition (12) becomes

$$p = \alpha_0 \mathbf{P}(\Psi^0 \leq u) + \sum_{i,j} \alpha_i \gamma_j^i \mathbf{P}(\Psi^{ij} \leq u). \quad (13)$$

*Proof.* See Appendix A2. □

Replacing  $\mathbf{P}(\Psi^0 \leq u)$  and  $\mathbf{P}(\Psi^{ij} \leq u)$  in (12) and (13) by the empirical probabilities, we obtain a SAA approach to solve the root finding problems posed by the first order conditions. The sample version of (13) is

$$\bar{p}_N(u) = \alpha_0 \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{[\Psi_k^0 \leq u]} + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{N} \mathbf{1}_{[\Psi_k^{ij} \leq u]}.$$

Utilizing a simple bisection method, one can determine the root  $\check{u}^N(\bar{\gamma})$  that solves  $\bar{p}_N(u) = p$ .

### 3.2.2 Outer maximization

The sample version of the outer maximization in (10) is

$$\max_{\bar{\gamma}} \left\{ \underbrace{\tilde{u}^N(\bar{\gamma}) + \frac{\alpha_0}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^0 - \tilde{u}^N(\bar{\gamma})]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \frac{1}{N} \sum_{k=1}^N [\Psi_k^{ij} - \tilde{u}^N(\bar{\gamma})]^+}_{=: \tilde{c}_p^N(\bar{\gamma})} \right\}$$

Algorithm 4 evaluates  $\tilde{c}_p^N(\bar{\gamma})$  for all  $\bar{\gamma}$  on a grid, compares the values of the function and thereby determines an approximate solution.

### 3.2.3 Switching condition and dimension reduction

The outer maximization requires the computation at many grid points and is expensive in high dimensions. We propose to identify a suitable lower-dimensional subsimplex in the space of copulas on the basis of SA, before we switch to SAA. This is justified by the fact that the worst-case risk is typically insensitive to contributions of some of the copulas in  $\mathfrak{C}$ . Before we summarize the full procedure, we address the switching condition from SA to SAA.

SA produces a random sequence  $(\bar{\gamma}_t)_{t=1,2,\dots}$ . We choose a certain burn-in period  $t_{\min}$  and a maximal number of SA-iterations  $t_{\max}$  to construct a stopping time  $t^* \in \{t_{\min}, t_{\min}+1, \dots, t_{\max}\}$ . We stop at  $t$  when two consecutive matrices  $\bar{\gamma}_{t-1}$  and  $\bar{\gamma}_t$  are close to each other according to some metrics. In the examples below, we implement the 1-norm  $\|A\| = \sum_i \sum_j |A_{ij}|$  and a threshold level of 0.01. Moreover, we choose  $t_{\min} = 10$  and  $t_{\max} = 50$ .

When switching to SAA, the dimension of the problem is reduced as follows. To simplify the notation, we write

$$\bar{\gamma} = \begin{pmatrix} \gamma_1^1 & \gamma_1^2 & \cdots & \gamma_1^m \\ \gamma_2^1 & \gamma_2^2 & \cdots & \gamma_2^m \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_K^1 & \gamma_K^2 & \cdots & \gamma_K^m \end{pmatrix}$$

instead of  $\bar{\gamma}_{t^*}$  where  $t^*$  is the stopping time described above. Recall that the index  $j = 1, 2, \dots, K$  enumerates the copulas in  $\mathfrak{C}$ , while  $i = 1, 2, \dots, m$  labels the weights  $\alpha_i$  and corresponding distortions  $D_{i1}, D_{i2}, \dots, D_{id}$  in eq. (2) or eq. (4), respectively. We assume that the weights are equal, i.e.,  $\alpha_i = \alpha \forall i = 1, 2, \dots, m$ ; this assumption ensures that the probability fraction of the total dependence that is governed by each column of  $\bar{\gamma}$  is equal for all columns.

We first select the number of copulas  $K^* < K$  we wish to select from  $\mathfrak{C}$  for the application of SAA. We distinguish the cases  $K^* \leq m$  and  $K^* > m$ . In the first case, we identify the largest entry<sup>3</sup> from  $\bar{\gamma}$  and select the copula corresponding to it. We remove the corresponding row and the corresponding column from  $\bar{\gamma}$ , identify the largest entry from the remaining matrix, and remove again the corresponding row and column. We proceed iteratively until  $K^*$  copulas are selected. In the second case, i.e.,  $K^* > m$ , all rows and columns are removed from  $\bar{\gamma}$ , after  $m$  copulas were selected. In this case, we proceed with selecting copulas  $m+1, m+2, \dots$  as follows. We remove all rows corresponding to the  $m$  copulas that were already selected, and then proceed in the same manner as described above to select the remaining copulas.

#### Remark 6

*For each  $i = 1, 2, \dots, m$ , the mixture copula corresponding to  $\gamma^i$  governs a probability fraction  $\alpha$  of the overall dependence structure in a region determined by the distortions  $D_{i1}, D_{i2}, \dots, D_{id}$ . The algorithm consecutively selected for different  $i$  the most important element from the copulas in  $\mathfrak{C}$  that were not previously selected. This guarantees that the contributions of the vectors of*

<sup>3</sup>If there is a tie, we select the one with the larger gradient.

distortion functions corresponding to different values of  $i$  are taken into consideration when the  $K^*$  copulas are chosen.

### 3.3 Full procedure

We finally present a brief summary of our proposed algorithm.

#### Step 1 – Sampling.

1. Generate  $N = N_t$  independent copies of the random variables  $\Psi^0, \Psi^{ij}$  as described in Section 2.3.
2. Use the samples to estimate the densities  $g_0(s), g_{ij}(s)$ , and  $f_{X^\tau}(s)$ . The values are stored for the inter- and extrapolation.
3. If necessary, generate new samples according to an importance sampling density  $h$ .

#### Step 2 – SA Algorithm.

4. Apply PSG-ROBUSTAV@R Algorithm 2. Importance sampling techniques can be adopted as illustrated in Section 3.4 below – if applicable.
5. If the switching condition described in Section 3.2.3 is met, terminate the algorithm and determine a selection of the most important copulas in order to reduce the dimensionality of the problem.

#### Step 3 – SAA Algorithm.

6. Construct a suitable grid on the lower-dimensional simplex. Adaptively refine the grid in a smaller domain on the basis of the results of the application of the algorithm specified in the next step, and apply the algorithm again on the new grid.
7. Apply SAA-ROBUSTAV@R Algorithm 4 to find the worst case over grid points. The worst case is the estimated solution to the original problem (6) on the lower-dimensional mixture space of copulas chosen in Task 5.

### 3.4 A motivating example

Before we discuss applications to finance markets and cyber risk, we illustrate our procedure in the context of a simple example motivated by Li et al. (2014).

**Example 2** ( $m = 2, K = 5, d = 2$ )

Consider aggregate losses  $X = X_1 + X_2$  with individual losses  $X_1, X_2 \in L^1$ . The distributions of the individual positions are inverse Gaussian with density

$$x \mapsto \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda}{2\mu^2 x}(x - \mu)^2\right).$$

The dependence of the positions is uncertain, and we would like to evaluate the worst-case AV@R at level  $p \in (0, 1)$ . Letting  $\alpha_0 = 1 - 2\alpha$  and  $\alpha_1 = \alpha_2 = \alpha$  with  $\alpha = 0.1$ , we assume that  $D_i = D_{i1} = \dots = D_{id}$  for all  $i = 0, 1, 2$  and choose the distortion functions

$$D_1(x) = \frac{x - \alpha x^2}{\alpha + (1 - 2\alpha)x}, \quad D_2(x) = \frac{\alpha x^2}{\alpha + (1 - 2\alpha)(1 - x)}, \quad D_0(x) = \frac{x - \alpha D_1(x) - \alpha D_2(x)}{1 - 2\alpha}. \quad (14)$$

The copula capturing dependence in the typical area is modeled by a  $d$ -dimensional Gaussian copula

$$C_0 = C_{\Sigma}^{Ga} = \Phi_{\mathbf{0}, \Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d));$$

here,  $\Phi$  and  $\Phi_{\mathbf{0}, \Sigma}$  signify the standard univariate normal distribution function and the multivariate Gaussian distribution function with covariance matrix  $\Sigma$ , respectively.

The family of copulas  $\mathfrak{C} = \{C_1, C_2, C_3, C_4, C_5\}$  is specified as follows:

$C_1$ : a  $t$ -copula  $C_{\nu, P}^t(u_1, \dots, u_d) = \mathbf{t}_{\nu, P}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d))$  where  $t_{\nu}$  is a standard univariate  $t$  distribution with  $\nu$  degree of freedom and  $\mathbf{t}_{\nu, P}$  is the joint distribution with a correlation matrix  $P$ ;

$C_2$ : a Clayton copula  $C^{Cl}(u_1, \dots, u_d) = \left(\sum_{i=1}^d u_i^{-\theta} - d + 1\right)^{-1/\theta}$ ,  $0 < \theta < \infty$ ;

$C_3$ : a Gumbel copula  $C^{Gu}(u_1, \dots, u_d) = \exp\left\{-\left[\sum_{i=1}^d (-\log u_i)^{\theta}\right]^{1/\theta}\right\}$ ,  $1 \leq \theta < \infty$ ;

$C_4$ : a Frank's copula  $C^{Fr}(u_1, \dots, u_d) = \log_{\theta}\left\{1 + \frac{\prod_{i=1}^d (\theta^{u_i} - 1)}{(\theta - 1)^{d-2}}\right\}$ ,  $\theta \geq 0$ ;

$C_5$ : the independence copula  $\Pi(u_1, \dots, u_d) = \prod_{i=1}^d u_i$ .

## SA Algorithm

Step 2 in the full procedure summarized in Section 3.3 is the SA Algorithm 2. Its step size is given by  $t^{-a}$  for  $0.5 < a \leq 1$ . Figure 1 illustrates the dynamics of the corresponding weights in the random sequence  $(\bar{\gamma}_t)_t$  for the five copulas in  $\mathfrak{C}$  and the distortions  $D_1$  and  $D_2$  for a specific numerical example. We vary the step size and compare  $a = 0.6, 0.7, 0.8, 0.9$  for the first 200 iterations. The approximation becomes faster for smaller  $a$ .

The downside risk measures by AV@R is mainly governed by the upper tail of the losses whose dependence structure is encoded by the distortion function  $D_2$ . This is captured by the second column in Figure 1 which shows that the weights of copulas  $C_2$  (Clayton copula),  $C_4$  (Frank's copula), and  $C_5$  (independence copula) decrease quickly to zero. The maximal AV@R is mainly determined by  $\gamma_1^2$  (the weight of  $t$ -copula  $C_1$  for the upper tail,  $D_2$ ) and  $\gamma_3^2$  (the weight of Gumbel copula  $C_3$  for the upper tail,  $D_2$ ). These observations suggest that dimension reduction as described in Section 3.2.3 can successfully be implemented for this example.

The initial AV@R at  $p = 0.99$  for uniform<sup>4</sup>  $\bar{\gamma}_1$  is reported as 13.8657 for  $a = 0.6$ , while AV@R is increased to 14.6832 just after five iterations. In fact, this number is hardly distinguishable from the estimated optimal value found in SAA later on. We observe that AV@R values become insensitive to changes in  $\bar{\gamma}_t$  after just a few iterations. This observation provides further motivation for the suggested approach to reduce the dimension of the problem (see Section 3.2.3).

## Importance sampling

We explore the potential to reduce the variance of the estimators by an application of importance sampling. Recall the notation introduced in Section 2.3.2. If  $h$  is a density that dominates  $f_{X\bar{\gamma}}$ , we may sample from  $h$  and modify Algorithm 2 to obtain importance sampling estimators replacing (i) VaR  $\hat{v}_p^N$ , (ii) AV@R  $\hat{c}_p^N$ , and (iii) the AV@R gradient  $\Delta_{i,j}(\hat{c}_p^N)$ .

Letting  $\mathcal{L} = \frac{f_{X\bar{\gamma}}}{h}$  be the likelihood ratio, we estimate the corresponding IS empirical distribution  $\tilde{F}_{X\bar{\gamma}}^{IS}(s)$  by

$$\tilde{F}_{X\bar{\gamma}}^{IS}(s) = \frac{1}{N} \sum_{l=1}^N \mathcal{L}(y_l) \mathbf{1}_{|y_l| \leq s}, \quad s \in \mathbb{R},$$

<sup>4</sup>All entries of the matrix are equal.

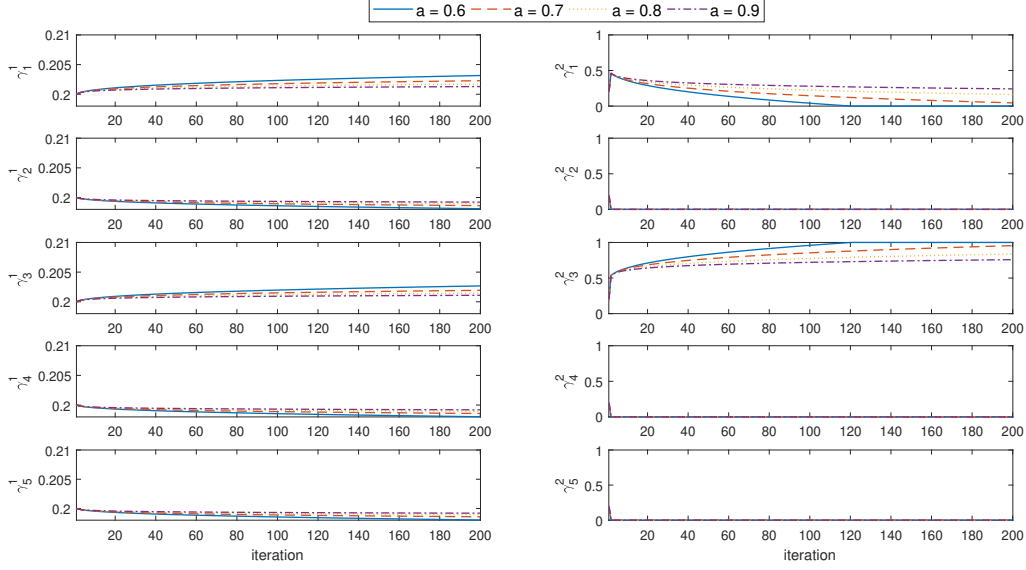


Figure 1: SA-results for varying  $a \in (0.5, 1]$  with step sizes  $\{\delta_t\} = \{t^{-a}\}_{t \geq 1}$  in Example 2. The off-diagonal elements of  $\Sigma$  equal 0.7, the diagonal elements 1. We set  $\nu = 1$ ,  $\theta = 0.7565, 1.7095, 1.2$  for  $C_2, C_3, C_4$ , respectively. The IG parameters are  $\mu_1 = \mu_2 = 1, \lambda_1 = 0.5, \lambda_2 = 1.2$ . The sample size is fixed as  $N_t = 10^6$  for every iteration  $t$  and a kernel density at 1000 equally spaced points is used based on  $5 \times 10^7$  sample data.

with  $y_l$  drawn iid from  $h$ . The corresponding IS estimators are

$$\tilde{v}_p^N = \inf\{s : \tilde{F}_{X^{\tilde{\gamma}}}^{IS}(s) \geq p\}; \quad \tilde{c}_p^N = \tilde{v}_p^N + \frac{1}{N(1-p)} \sum_{l=1}^N (y_l - \tilde{v}_p^N)^+ \mathcal{L}(y_l);$$

$$\Delta_{i,j}(\tilde{c}_p^N) = \frac{1}{N(1-p)} \sum_{l=1}^N \frac{\alpha_i g_{ij}(y_l)}{f_{X^{\tilde{\gamma}}}(y_l)} (y_l - \tilde{v}_p^N) \mathcal{L}(y_l) \mathbf{1}_{[y_l \geq \tilde{v}_p^N]}.$$

Motivated by eq. (7), we propose to define the IS density  $h$  as a mixture that relies on measure changes of the distribution functions  $G_0, G_{ij}$  with densities  $g_0, g_{ij}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, K$ :

$$h(x) = \alpha_0 h_0(x) + \sum_{i=1}^m \alpha_i \sum_{j=1}^K \gamma_j^i h_{ij}(x).$$

For simplicity, we modify only two ingredients:

We replace the central copula  $C_0$  by an importance sampling copula  $\tilde{C}_0$  and the marginal distributions  $F_i$  by importance sampling distributions  $\tilde{F}_i$ ; all other ingredients of the family of joint distributions of  $X_1, X_2, \dots, X_d$  in Example 2, in particular the collection  $\mathfrak{C}$ , are not changed. We thus obtain the following identities:

$$h_0(x) = \frac{\partial}{\partial x} \int \mathbf{1}_{\tilde{A}(x)} d\tilde{C}_0(D_{01}(u_1), \dots, D_{0d}(u_d));$$

$$h_{ij}(x) = \frac{\partial}{\partial x} \int \mathbf{1}_{\tilde{A}(x)} dC_j(D_{i1}(u_1), \dots, D_{id}(u_d)) \quad \forall i, j;$$

$$\tilde{A}(x) = \left\{ (u_1, \dots, u_d) : \Psi \left( \tilde{F}_1^{-1}(u_1), \dots, \tilde{F}_d^{-1}(u_d) \right) \leq x \right\}.$$

Many other strategies to design IS distributions are, of course, possible. However, good IS methodologies for copulas are challenging. At the same time, the total computational effort



must be estimated in order to evaluate the overall efficiency of competing algorithms. These issues constitute an interesting topic for future research.

To illustrate the potential of IS, we consider Example 2. As suggested by Huang, Subramanian & Xu (2010), we shift the mean vector of the Gaussian copula  $C_0$  to obtain  $\tilde{C}_0$ . On the marginal distributions, we utilize for each  $i = 1, 2, \dots, m$  an Esscher measure change with parameter  $w_i$  that transforms an inverse Gaussian distribution  $IG(\mu_i, \lambda_i)$  to a shifted IG distribution  $\tilde{F}_i \sim IG\left(\frac{\mu_i \sqrt{\lambda_i}}{\sqrt{\lambda_i - 2\mu_i^2 w_i}}, \lambda_i\right)$  with  $w_i \leq \frac{\lambda_i}{2\mu_i^2}$ .

Numerical results display significant variance reduction. For example, in a typical case study with  $10^6$  samples the variance of the crude MC estimator of robust AV@R is 0.00247 while the importance sampling variances are reported as 0.00044 and 0.00042 for the historical likelihood estimator and the kernel estimator, respectively. We observe average variance reduction ratios around 5 to 7 across samples with the following new parameters: for exponential tilting  $w_1 = 0.1$  (new  $\mu_1^{IG} = 1.2910$ ),  $w_2 = 0.3$  (new  $\mu_2^{IG} = 1.4142$ ) and for the shifted drift for Gaussian distribution  $\mu_1^G = 0.5$ , and  $\mu_2^G = 1$ .

### Switching to SAA

We apply the methodology described in Section 3.2.3. Setting  $t_{\min} = 10$  and  $t_{\max} = 50$ , we run SA with uniform initial values, i.e., all entries of  $\bar{\gamma}_1$  are  $1/5$ , and with a sample size  $N_t = 10^5$  for step size  $a = 0.6$ . Recall Algorithm 2 for a description of the parameters. The initial choice of  $\bar{\gamma}$  corresponds to an AV@R at level 0.99 of 13.8046. This result differs slightly from the initial value reported in Figure 1 due to sampling error.

The stopping time equals  $t^* = 17$  with corresponding

$$\bar{\gamma}^\top = \begin{pmatrix} 0.2008 & 0.1994 & 0.2007 & 0.1994 & 0.1994 \\ 0.3309 & 0 & 0.6690 & 0 & 0 \end{pmatrix}$$

and AV@R at level 0.99 of 14.8094 with an empirical standard deviation of 0.1326 computed from the last ten iterations. The increments of the sequence  $(\bar{\gamma}_t)_{t=1,2,\dots}$  are already small at the stopping time  $t^*$ :

$$(\bar{\gamma}_{17} - \bar{\gamma}_{16})^\top = \begin{pmatrix} 0.00002 & -0.00001 & 0.00001 & -0.00001 & -0.00001 \\ -0.004 & 0 & 0.004 & 0 & 0 \end{pmatrix}.$$

For comparison, at iteration 100 we obtain a corresponding

$$\bar{\gamma}^\top = \begin{pmatrix} 0.2017 & 0.1989 & 0.2015 & 0.1988 & 0.1988 \\ 0.1484 & 0 & 0.8515 & 0 & 0 \end{pmatrix}$$

and AV@R at level 0.99 of 14.7715 with an empirical standard deviation of 0.1553 computed from the last ten iterations. These observations indicate that SA quickly approximates the worst-case AV@R. However, the precision improves only very slowly afterwards. The convergence to the optimal value of  $\bar{\gamma}$  is slow for some components.

In order to reduce the dimension of the problem according to Section 3.2.3, we set  $K^* = 2$  and select for the application of SAA the copulas  $C_1$  (t-copula) and  $C_3$  (Gumbel copula) on the basis of the estimate  $\bar{\gamma}_{17}$ . Thus, SAA needs to be applied to a two-dimensional grid for

$$\bar{\gamma}^\top = \begin{pmatrix} \gamma_1^1 & 0 & \gamma_3^1 & 0 & 0 \\ \gamma_1^2 & 0 & \gamma_3^2 & 0 & 0 \end{pmatrix}, \quad \gamma_1^1 + \gamma_3^1 = 1, \gamma_1^2 + \gamma_3^2 = 1, \gamma_1^1, \gamma_3^1, \gamma_1^2, \gamma_3^2 \geq 0.$$

On the basis of SAA with  $5 \cdot 10^7$  samples one observes that the worst-case risk is insensitive to dependence in the lower tail. The worst-case risk is attained for a  $\gamma_3^2 = 1$  (upper tail) with an AV@R at level 0.99 of 14.71. This is illustrated in Figure 2. The worst-case risk in the

considered model is lower than the sum of the marginal AV@Rs which is equal to 15.49; this value corresponds to the comonotonicity of all components. This confirms that the underlying assumption (i.e., that dependence in different regions can be modeled separately and that an expert’s opinion limits the choices of copulas) reduces model uncertainty.

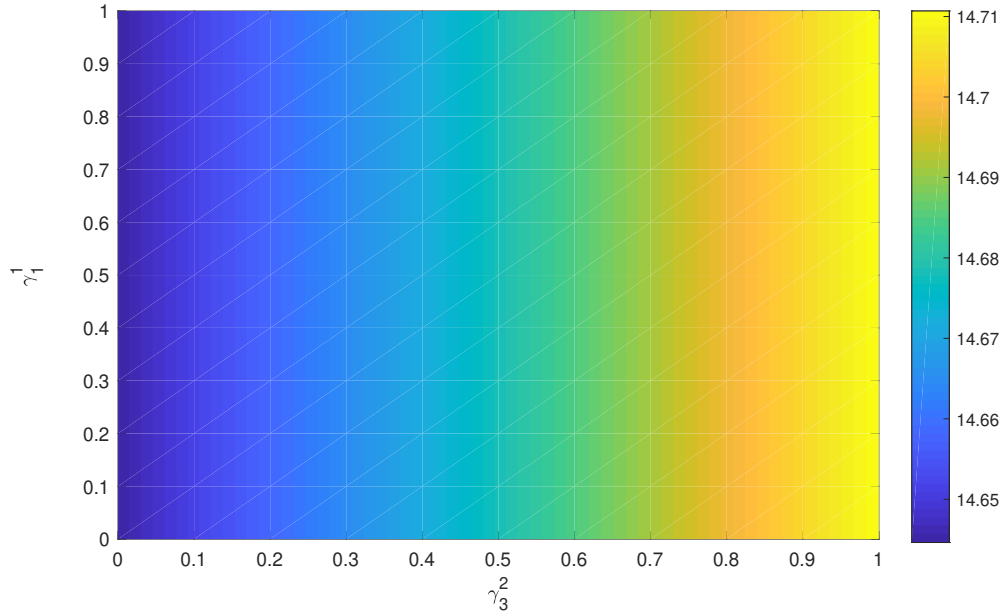


Figure 2: Color map of AV@R for parameters  $\gamma_1^1$  and  $\gamma_3^2$ .

In summary, in this case study SA is capable of quickly computing a reasonable estimate of the worst-case risk. Suitable worst-case copulas, encoded by the matrix  $\bar{\gamma}$ , in a lower-dimensional mixture space can be determined by a combination of SA, the copula selection method described in Section 3.2.3, and SAA.

## 4 Applications

Our method is flexible and can be used in multiple application domains. For the purpose of illustrating its applicability, we consider two case studies. The first example in Section 4.1 is based on a substantial amount of financial market data and allows a good calibration of copulas. Model risk can thereby be reduced.<sup>5</sup> For cyber risk, the second example discussed in Section 4.2, only few observations are available which also increases the model risk.

### 4.1 Financial markets

We apply our methodology to a data set spanning the time interval 2005/01/01 to 2019/12/31 that contains the daily closing values of the following stock indices:

i	Index
1	S&P 500
2	NASDAQ Composite
3	Dow Jones Industrial Average
4	DAX Performance Index
5	EURONEXT 100
6	KOSPI Composite Index
7	Nikkei 225

<sup>5</sup>In order to illustrate this claim, we include an additional case study in Appendix A.2.

This data period also includes extreme events during the 2007 – 2008 global financial crisis. The 7-dimensional time series is labeled by trading days  $t = 1, 2, \dots, 3358$  and quoted in US\$:

$$\text{PRICE}_t = (\text{PRICE}_{1,t} \quad \text{PRICE}_{2,t} \quad \text{PRICE}_{3,t} \quad \text{PRICE}_{4,t} \quad \text{PRICE}_{5,t} \quad \text{PRICE}_{6,t} \quad \text{PRICE}_{7,t})$$

with  $\text{PRICE}_{i,t}$  being the time  $t$  US\$-price of index  $i$ ,  $i = 1, 2, \dots, 7$ . We consider a 7-dimensional random vector<sup>6</sup>  $\mathbf{X} = (X_1, X_2, \dots, X_7)$  that models the negative returns of the terminal US\$-value of the indices over a 10-day horizon. The corresponding time series is given by

$$\mathbf{x}_t = \left( \{-1\} \cdot \frac{\text{PRICE}_{i,t+10} - \text{PRICE}_{i,t}}{\text{PRICE}_{i,t}} \right)_{i=1,2,\dots,7}, \quad t = 1, 2, \dots, \underbrace{3348}_{=: D}$$

We investigate the robust AV@R at level 0.95 over a 10-day horizon of a portfolio that invests an equal dollar amount into each index. To be more specific, we consider the robust AV@R of the losses  $X = \sum_{i=1}^7 X_i$ .

#### 4.1.1 Marginal distributions

We apply a semi-parametric approach to the seven marginal distributions. For the central part of the distributions we linearly interpolate the empirical distribution. Less data are available in the tail, and we fit Generalized Pareto Distributions (GPD) to the data which allow a convenient extrapolation of samples.

The CDF of a GPD with two parameters  $\xi$  and  $\vartheta$  is given by

$$G_{\xi,\vartheta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\vartheta}\right)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\vartheta), & \text{if } \xi = 0. \end{cases}$$

The GPD is supported on  $x \geq 0$ , if  $\xi \geq 0$ , and on  $0 \leq x \leq -\vartheta/\xi$ , if  $\xi < 0$ .

To be specific, for any  $i = 1, 2, \dots, 7$ , let  $(x_{i,(t)})_{t=1,\dots,D}$  be the ordered sample of the data  $(x_{i,t})_{t=1,\dots,D}$  with  $x_{i,(1)} \leq \dots \leq x_{i,(D)}$ . As GPD approximates a tail distribution for the excesses above some high threshold, we let  $x_{i,l}$  and  $x_{i,u}$  be suitably chosen thresholds of lower and upper parts. We apply a graphical diagnostic for the threshold choice; it is based on a mean excess plot that should be linear in the threshold for a GPD. For alternative, more sophisticated methods we refer to Scarrott & MacDonald (2012). The two parameters  $(\xi_{i,l}, \vartheta_{i,l})$  and  $(\xi_{i,u}, \vartheta_{i,u})$  are then determined by maximum likelihood estimation based on the lower and upper excess data  $(x_{i,l} - x_{i,(1)}, \dots, x_{i,l} - x_{i,(t_{i,l}-1)})$  and  $(x_{i,(t_{i,u}+1)} - x_{i,u}, \dots, x_{i,(D)} - x_{i,u})$ , respectively. The estimated thresholds (i.e., the upper and lower boundaries:  $x_{i,u} = x_{i,(t_{i,u})}$ ,  $x_{i,l} = x_{i,(t_{i,l})}$ ) and the corresponding parameters are reported in Table 1.

The linearly interpolated empirical distribution function truncated in  $[x_{i,l}, x_{i,u}]$  for index  $i$  is

$$H_i(x) = \begin{cases} 0 & x < x_{i,l} \\ \frac{k-1}{t_{i,u}-t_{i,l}} + \frac{x-x_{i,(k)}}{(t_{i,u}-t_{i,l})(x_{i,(k+1)}-x_{i,(k)})} & x_{i,(k)} \leq x < x_{i,(k+1)} \\ 1 & x \geq x_{i,u}, \end{cases}$$

where  $t_{i,l}, t_{i,u}$  are the indices in the ordered data corresponding to the lower and upper thresholds.

The distribution of  $X_i$  is finally modeled by

$$F_i(x) = \begin{cases} p_l(1 - G_{\xi_{i,l}, \vartheta_{i,l}}(x_{i,l} - x)) & x \leq x_{i,l} \\ p_l + (1 - p_l - p_u)H_i(x) & x_{i,l} < x \leq x_{i,u} \\ (1 - p_u) + p_u G_{\xi_{i,u}, \vartheta_{i,u}}(x - x_{i,u}) & x > x_{i,u}, \end{cases} \quad (15)$$

<sup>6</sup>For simplicity, we do not considered any dependence of the returns on time or on underlying factors. In practice, conditional distributions are typically important to reflect the market conditions.

Index $i$	Lower tail			Upper tail		
	boundary $x_{i,l}$	shape $\xi_{i,l}$	scale $\vartheta_{i,l}$	boundary $x_{i,u}$	shape $\xi_{i,u}$	scale $\vartheta_{i,u}$
1	-0.0466	0.3172	0.0131	0.0445	0.2749	0.0219
2	-0.0594	0.1826	0.0173	0.0562	0.2801	0.0212
3	-0.0355	0.1791	0.0124	0.0446	0.2355	0.0209
4	-0.0578	-0.1227	0.0269	0.0495	0.1042	0.0263
5	-0.0640	0.0453	0.0176	0.0656	-0.0307	0.0266
6	-0.0658	0.4326	0.0211	0.0600	0.2211	0.0335
7	-0.0568	0.3802	0.0139	0.0413	0.1226	0.0236

Table 1: The thresholds (boundaries) and the estimated shape and scale parameters of GPDs in the lower and upper tail parts.

Lower Tail	index	1	2	3	4	5	6	7
Anderson-Darling	statistic	0.3069	0.3889	0.4054	0.2127	0.4712	0.2882	0.6304
	p value	0.6675	0.4972	0.4642	0.9255	0.3742	0.7022	0.1497
Cramér-von-Mises	statistic	0.0386	0.0369	0.0580	0.0227	0.0562	0.0256	0.0564
	p value	0.6984	0.7462	0.4371	0.9628	0.4876	0.9024	0.4177
Upper tail	index	1	2	3	4	5	6	7
Anderson-Darling	statistic	0.5173	0.4178	0.4654	0.5473	0.3104	0.4984	0.2146
	p value	0.2706	0.4240	0.3491	0.2637	0.7126	0.3030	0.9053
Cramér-von-Mises	statistic	0.0634	0.0432	0.0747	0.0463	0.0362	0.0694	0.0186
	p value	0.3582	0.6264	0.2660	0.6104	0.7917	0.3106	0.9827

Table 2: The results of Anderson-Darling and Cramér-von-Mises tests for the lower and upper GPD approximations.

where  $p_l = P(x \leq x_{i,l})$  and  $p_u = P(x \leq x_{i,u})$ .

The results of Anderson-Darling and Cramér-von-Mises tests for the GPD lower and upper tails are reported in Table 2. For the goodness of test of GPD, we follow the approach in Choulakian & Stephens (2001). The results do not provide evidence against the estimated marginal distributions.

#### 4.1.2 Dependence

Since AV@R focuses on the upper tails, we consider the following distortion functions with parameter  $\alpha_1 = \alpha_2 = 0.04$  and  $\alpha_0 = 1 - \alpha_1 - \alpha_2$ :

$$\begin{aligned}
D_0(x) &= \begin{cases} \frac{x}{\alpha_0} & \text{if } x \leq \alpha_0 \\ 1 & \text{if } x > \alpha_0 \end{cases} & D_1(x) &= \begin{cases} 0 & \text{if } x \leq \alpha_0 \\ \frac{x-\alpha_0}{\alpha_1} & \text{if } \alpha_0 < x \leq \alpha_0 + \alpha_1 \\ 1 & \text{if } x > \alpha_0 + \alpha_1 \end{cases} \\
D_2(x) &= \begin{cases} 0 & \text{if } x \leq \alpha_0 + \alpha_1 \\ \frac{x-\alpha_0-\alpha_1}{\alpha_2} & \text{if } \alpha_0 + \alpha_1 < x \leq 1. \end{cases} \tag{16}
\end{aligned}$$

We split the data on the basis of the aggregate loss function  $\{\sum_{i=1}^7 x_{i,t}\}_{t=1,\dots,D}$  into three parts: extreme upper tail (-4%), upper tail (4%-8%), and the remaining center and lower tail (8%-100%). A scatter plot of  $(X_1, X_2), \dots, (X_1, X_7)$  in Figure 3 illustrates the procedure of partitioning the data.

The dependence in the central part is modeled by a Gaussian copula corresponding to the estimated linear correlations. The estimates are based on the 92% data (8%-100%) and can be found in Section A.3. For the tail parts,  $K = 16$  candidate copulas<sup>7</sup> are considered:

<sup>7</sup>The method is very flexible and could equally be applied to a larger set of copulas. The specific copulas are

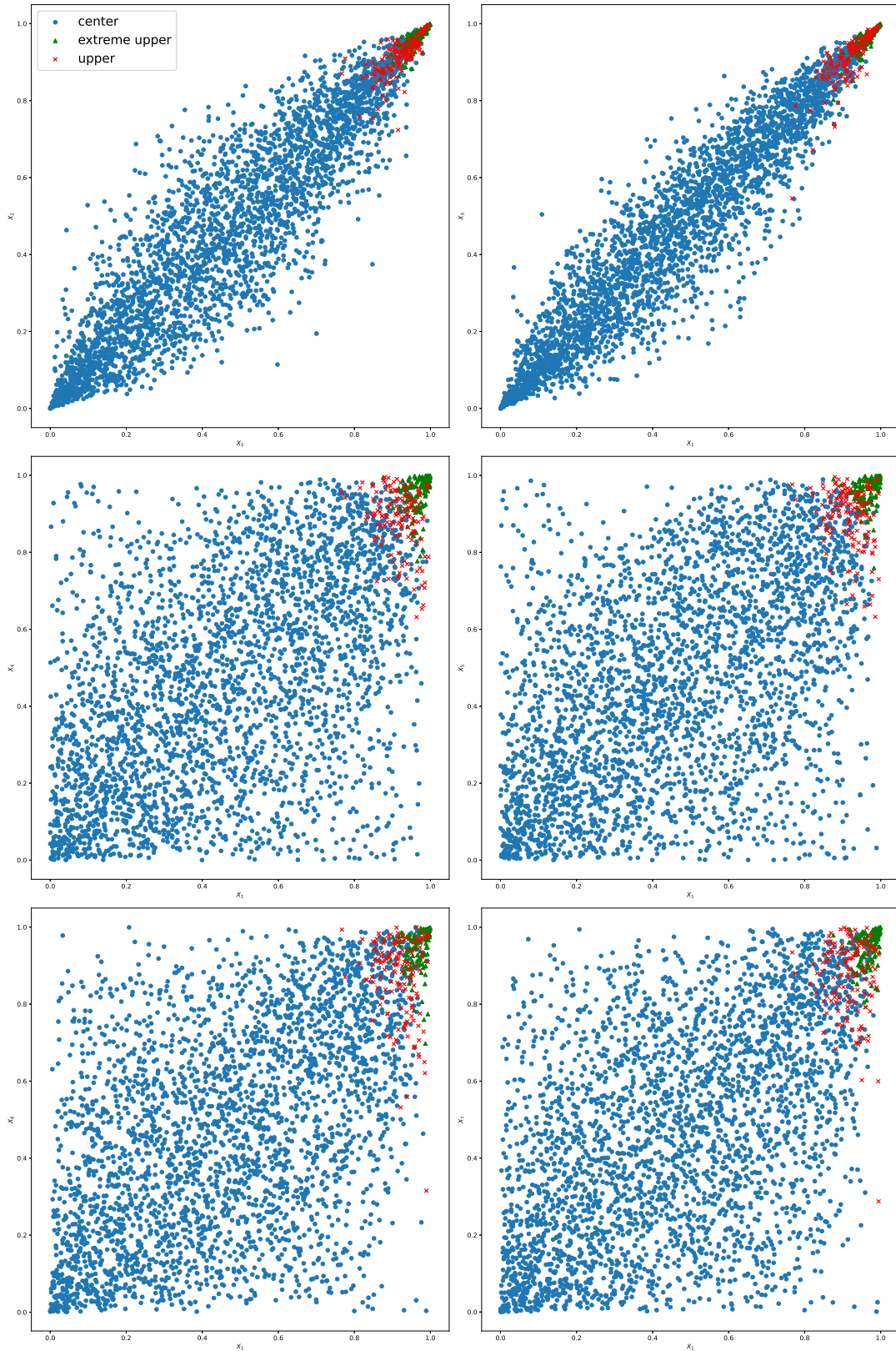


Figure 3: Scatter plot of  $X_i$  against  $X_1$  for  $i = 2, \dots, 7$ . The green triangles depict the data points in the extreme upper tail (-4%), the red x the data points in the upper tail (4%-8%), and the blue circles the remaining data points (8%-100%).

---

a potential choice due to an expert's opinion. This corresponds to model uncertainty that a priori is limited in this way.

- Copulas  $C_1, C_2$  are Gaussian, matching the linear correlation or Kendall's tau estimated from the upper (4%-8%) data, respectively.
- The copulas  $C_3, C_4, \dots, C_8$  are t-copulas whose parameters are calibrated on the basis of the upper (4%-8%) data:
  - the multivariate meta t-copulas  $(C_3, C_4, C_5) = (C_{\nu^1, P^1}^t, C_{\nu^2, P^2}^t, C_{\nu^3, P^3}^t)$  with parameters  $(\nu^l, P^l)$ ,  $l = 1, 2, 3$ ; the superscript  $l$  indicates the estimation method explained below;
  - the grouped t-copula  $(C_6, C_7, C_8) = (C_{\nu^1, P^1}^{Gt}, C_{\nu^2, P^2}^{Gt}, C_{\nu^3, P^3}^{Gt})$  that allow different subsets of the random variates to have different degrees of freedom parameters; we divide the indices into the three subgroups US, Europe, and Asia; for each  $l = 1, 2, 3$ , the vector  $\nu^l = (\nu_1^l, \nu_2^l, \nu_3^l)$  specifies the degrees of freedom for these three subgroups, and the superscript  $l$  refers to the estimation method.

The first method ( $l = 1$ ) is ML estimation. The second method ( $l = 2$ ) exploits the approximated log-likelihood for the degrees of freedom parameter which increases the speed of the estimation. The third method ( $l = 3$ ) estimates the correlation matrix  $P^3$  by Kendall's tau for each pair and then estimates the scalar degree of freedom by ML given the fixed  $P^3$ . This method is useful when the dimension of the data is large, because the numerical optimization quickly becomes infeasible. The estimated correlation matrices as well as the degrees of freedom are not identical and sometimes even very different.

- The copulas  $C_9, C_{10}, \dots, C_{16}$  are constructed analogously to  $C_1, C_2, \dots, C_8$ , but based on the extreme upper (-4%) data.

The calibration results can be found in Section A.3.

#### 4.1.3 Case studies

As a benchmark, we compute the AV@R at level 0.95 when dependence is modeled by a single Gaussian copula estimated from the entire data set. The correlation matrix is given in Section A.3, and the AV@R equals 0.514928 when the number of samples is  $10^7$ .

We compare the benchmark to the algorithm based on DM copulas described in Section 3. With an equal initial weight of  $1/16$ , a constant sample size  $N_t = 10^6$  and step size  $a = 0.7$ , the initial AV@R at level 0.95 corresponds to 0.652009 and is substantially higher than the benchmark. The DM method with copulas fitted to tail data provides a much better methodology in assessing downside risk than single Gaussian copulas. In fact, even if a DM method combines only Gaussian copulas for central and tail areas, the estimation results are often reasonable.<sup>8</sup> For the considered data, results are quite insensitive to the considered copulas, as long as they are fitted to different parts of the distribution and a DM copula is used.

When running SA, the stopping time  $t^* = t_{\min}$  equals 10 with an AV@R at level 0.95 of 0.655033 and an empirical standard deviation of 0.0019 computed from the last ten iterations. The corresponding weights  $\bar{\gamma}_{10}^\top$  are

$$\begin{pmatrix} 0.0497 & 0.0494 & 0.0676 & 0.0693 & 0.0503 & 0.0675 & 0.0693 & 0.0503 & 0.0630 & 0.0624 & 0.0696 & 0.0699 & 0.0616 & 0.0696 & 0.0694 & 0.0613 \\ 0.0595 & 0.0598 & 0.0643 & 0.0657 & 0.0618 & 0.0650 & 0.0624 & 0.0611 & 0.0624 & 0.0616 & 0.0649 & 0.0652 & 0.0628 & 0.0638 & 0.0573 & 0.0623 \end{pmatrix}$$

with an increment  $\bar{\gamma}_{10} - \bar{\gamma}_9$  with components of very small modulus (roughly less than  $1/1000$ ).

Now we switch to SAA. Setting  $K^* = 3$ , our procedure selects on the basis of the estimated  $\bar{\gamma}_{10}$  for the application of SAA the copulas  $C_4$  (t-copula estimated from the upper (4%-8%) data using approximate ML),  $C_{12}$  (t-copula estimated from the extreme upper (-4%) data using approximate ML), and  $C_{14}$  (grouped t-copula estimated from the extreme upper (-4%) data

<sup>8</sup>See Section A.2 for further evidence.

using ML). SAA with sample size  $10^7$  and grid size 0.1 applied to the corresponding three-dimensional grid picks only one copula,  $C_{12}$  (t-copula estimated from the extreme upper (-4%) data using approximate ML), associated with an AV@R of 0.655452. The numerical analysis confirms that the AV@R values are insensitive to  $\bar{\gamma}$  near the solution. In the current example, the worst case AV@R is very close to the sum of the marginal AV@Rs, i.e., to comonotonic dependence of all components, that equals 0.656120.

In summary, when computing AV@R at level 0.95, in the current example a reasonable amount of tail data is available to estimate the dependence of the factors in different parts of the distribution. In contrast to a single Gaussian copula, the DM method provides solutions in the considered family that are not very sensitive to the choice of the estimated component copulas. But instead of making ad hoc assumptions that select a specific components copula a priori, our algorithm demonstrates explicitly the strength of the DM method, identifies and substantiates the insensitivity to components a posteriori and finally reduces the dimensionality of the problem in the worst-case analysis.

#### 4.1.4 Robustness

Another question one may ask is how the initial choice of  $\alpha_0, \alpha_1, \alpha_2$  influences the final result of the algorithm. In fact, since the distortion functions in eq. (16) and thus the regions of different copulas depend on these parameters, we cannot expect that the copulas are invariant if  $\alpha_0, \alpha_1, \alpha_2$  change.

We follow the same methodology as described above but vary  $\alpha_1$ . Recall that  $\alpha_0 = 1 - \alpha_1 - \alpha_2$  and that  $\alpha_2$  is equal to  $\alpha_1$ . For a given  $\alpha_1$ , we split the data into three parts depending on the value of  $\alpha_0, \alpha_1$  and  $\alpha_2$ ; the data are segmented into the extreme upper  $-\alpha_1 \cdot 100\%$  part, the upper  $\alpha_1 \cdot 100\% - 2 \cdot \alpha_2 \cdot 100\%$ , and the remaining part. We apply this procedure for  $\alpha_1$ : 0.03, 0.04, 0.05, 0.06.

Table 3 displays the worst case DM AV@R values together with the copulas chosen by our algorithm for the application of SAA when the dimension of the simplex is reduced. Both the worst case DM AV@R and the corresponding copulas for both  $D_1$  and  $D_2$ , in both cases  $C_{12}$ , are robust with respect to the choice of  $\alpha_1$ . However, as indicated in the third row of Table 3, when we reduced the dimension before applying SAA and select  $K^*$  candidate copulas from the SA results, different candidates are picked.

$\alpha_1$	0.03	0.04	0.05	0.06
Worst case DM AV@R	0.655472	0.655452	0.655967	0.655872
Dimension reduction: selected copulas	C8, C12, C14	C4, C12, C14	C4, C8, C12	C4, C8, C12

Table 3: The AV@R values of the worst case DM copula (corresponding to  $C_{12}$  for both  $D_1$  and  $D_2$ ) and the selected candidate copulas for the SAA procedure with  $K^* = 3$ .

## 4.2 Cyber risk

In an application to cyber risk, we study cyber incidents in USA from Privacy Rights Clearinghouse (<https://privacyrights.org/>) collected from January 2005 until October 2019. For a time window ending in 2016 the data set was also analyzed by Eling & Jung (2018). We consider loss records in periods of two months and rearrange the data accordingly. This reduces the number of zero entries and admits a tractable analysis that does not separate zero entries from strictly positive losses.<sup>9</sup>

<sup>9</sup>In contrast to our simplified approach, Eling & Jung (2018) build their analysis on a methodology described in Erhardt & Czado (2012) that expresses the joint probability function by copulas with discrete and continuous margins. Our algorithmic approach can also be applied to their methodology. The statistical estimation is, however, more difficult in this case.

The data records contain information on the period, the number of events in each period and the corresponding losses. We consider five types of breaches:<sup>10</sup>

i	Data Breaches Type (Number of zero data): description
1	DISC (0): Unintended Disclosure Not Involving Hacking, Intentional Breach or Physical Loss
2	HACK (0): Hacked by an Outside Party or Infected by Malware
3	INSD (17): Insider - employee, contractor or customer
4	PHYS (5): Physical - paper documents that are lost, discarded or stolen
5	PORT (16): Portable Device - lost, discarded or stolen laptop, smartphone, memory stick, etc.

The data set contains 89 two months periods. The number of dates with observations of zero losses or no incidents is provided in parenthesis. In order to simplify the analysis, we replace all zero entries by a uniform  $(0, 1)$  random variable; since the severity for non-zero losses is typically on the order of  $10^3$  or more, this approach does not substantially modify the data, but admits a simplified model with strictly positive marginal densities.

#### 4.2.1 Losses due to data breaches

We assume that the two months breach records can be modeled by a 5-dimensional random vector  $\mathbf{L} = (L_1, L_2, \dots, L_5)$ . We employ a loss distribution approach to the marginal distributions, i.e.,

$$L_i = \sum_{j=1}^{N^i} R_j^i$$

where  $R_j^i, j = 1, \dots, N^i$  are iid random variables representing the severity of individual loss records and  $N^i$  signifies the random number of losses. The dependence among  $\mathbf{L}$  is captured by a copula which will be modelled as a DM copula. The details of selection and calibration will be given below. In general, one is interest in measuring the risk of some functional of  $\mathbf{L}$ . As an illustrative example, we focus on the AV@R at level 0.95 of the sum of its components,  $X = \sum_{i=1}^5 L_i$ .

#### 4.2.2 Marginal distributions

Motivated by Eling & Jung (2018), we model the frequency and the severity of loss records separately, choosing a lognormal distribution for the severity and a negative binomial distribution for the number of losses in each period. We estimate the parameters of the distributions and summarize the results in Table 4; for the negative binomial we applied MLE, for the lognormal unbiased estimates of mean and variance of the log-data.

Type	Negative binomial		Lognormal	
	$r$	$p$	$\mu$	$\sigma$
1	2.8684	0.1209	9.8543	2.4364
2	1.6333	0.0543	11.7851	2.5086
3	0.9250	0.1196	6.9622	4.4965
4	1.3117	0.0632	7.9432	2.9350
5	0.9685	0.0685	7.9445	4.7539

Table 4: Estimation results for the loss frequency and severity of  $\mathbf{L}$ .

For the implementation of our algorithm, we finally generate and store  $10^7$  samples of each distribution.

<sup>10</sup>The description was obtained from the website <https://privacyrights.org/>.



### 4.2.3 Dependence

Since the AV@R focuses on the upper tails, we continue to use the distortion functions in (16), again choosing  $\alpha_1 = \alpha_2 = 0.04$  and  $\alpha_0 = 1 - \alpha_1 - \alpha_2$ . If AV@R at level 0.95 is computed by a corresponding DM copula, the dependence on the central and lower part (captured by  $D_0$ ) is very low; this is confirmed in numerical experiments. For this reason, we focus on a particularly simple approach and use a Gaussian copula for this part with linear correlation estimated from the data. The correlation matrix  $\Sigma_1$  is given in Section A.4 in the appendix.

For the upper tails (captured by  $D_1$  and  $D_2$ ), we consider  $K = 8$  candidate copulas, namely two Gaussian copulas, two t-copulas, two Gumbel copulas, and two vine copulas; for the latter we refer to Dißmann, Brechmann, Czado & Kurowicka (2013) for further information. More specifically, the copulas are estimated as follows; the corresponding parameters for Gaussian, t and vine copulas are given in Section A.4:

- $C_1$ : Gaussian copula with the estimated linear correlation  $\Sigma_1$ ;
- $C_2$ : Gaussian copula that matches the estimated Kendall's tau with corresponding correlation matrix  $\Sigma_2$ ;
- $C_3$ : t-copula with parameters  $\nu_1$  and  $P_1$  estimated by MLE;
- $C_4$ : t-copula with parameters  $\nu_2$  and  $P_2$  with  $P_2$  matching Kendall's tau and  $\nu_2$  estimated by MLE;
- $C_5$ : Gumbel copula estimated by MLE with parameter  $\theta = 1.875123$ ;
- $C_6$ : Gumbel copula estimated on the basis of a minimal Cramér-von Mises distance according to Hofert, Mächler & McNeil (2013) with parameter  $\theta = 1.000061$ ;
- $C_7$ : Regular vine copula estimated according to AIC;
- $C_8$ : Regular vine copula estimated according to BIC.

### 4.2.4 Case studies

As a benchmark, we compute the AV@R at level 0.95 when dependence is modeled by the single Gaussian copula  $C_1$  estimated from the entire data set. The unit for the reported AV@R values is always one million. The estimated AV@R at level 0.95 equals 45.6533 in this case on the basis of  $10^7$  samples. If we use copulas  $C_2, C_3, \dots, C_8$  points estimates range from about 45.2 to 53.1 with significant sampling error. As in Section 4.1, we compare this benchmark to the result of the algorithm with DM copulas that was described in Section 3. The DM approach provides a more sophisticated analysis of the worst case.

With an equal initial weight of  $1/8$ , a constant sample size  $N_t = 10^6$  and step size  $a = 0.7$ , the initial AV@R at level 0.95 corresponds to 49.4159. Stopping SA according to the our stopping rule at  $t^* = 10$ , we obtain an estimated AV@R at level 0.95 of 53.4793 with an empirical standard deviation of 3.8764 computed from the last ten iterations. The corresponding  $\bar{\gamma}^\top$  equals

$$\begin{pmatrix} 0.1187 & 0.1191 & 0.1232 & 0.1193 & 0.1307 & 0.1307 & 0.1291 & 0.1292 \\ 0 & 0 & 0 & 0 & 0.5751 & 0.4249 & 0 & 0 \end{pmatrix}$$

with increments  $(\bar{\gamma}_{10} - \bar{\gamma}_9)^\top$  of an order of  $1/500$  or less.

Setting  $K^* = 3$  and switching to SAA, our algorithm selects the copulas  $C_5$  (Gumbel copula with  $\theta = 1.875123$ ),  $C_6$  (Gumbel copula with  $\theta = 1.000061$ ), and  $C_8$  (Regular vine copula according to BIC, Table 8) on the basis of the estimate  $\bar{\gamma}_{10}$ . Thus, SAA needs to be applied to a three-dimensional grid on

$$\gamma_5^i + \gamma_6^i + \gamma_8^i = 1, \quad i = 1, 2, \quad \gamma_j^i \geq 0, \quad i = 1, 2, j = 5, 6, 8.$$

With a sample size of  $10^7$  sample size and 0.1 grid size, SAA selects the Gumbel copula  $C_5$  for  $D_1$  and the Gumbel copula  $C_6$  for  $D_2$  with a worst-case AV@R at level 95% of 53.6990. For the distortion  $D_1$ , the copula  $C_6$  leads to almost the same result, i.e., for  $D_1$  the sensitivity of the AV@R with respect to  $C_5$  and  $C_6$  is almost zero. This worst case AV@R may be compared to the comonotonic case, i.e., the sum of the marginal AV@Rs, which equals 55.133051. This shows that within our setting (that limits the admissible dependence structure on the basis of an expert’s opinion) model uncertainty is already reduced.

In summary, when computing AV@R at level 0.95, DM methods provide an excellent method for identifying the relevant low-dimensional dependence structures, when many data are available as illustrated in Section 4.1. In the current example on cyber risk, data are scarce and tail copulas are chosen ad hoc on the basis of an expert’s opinion. In this case, our algorithm easily identified the worst-case dependence and reduces the dimensionality at the same time. If only few data are available in the tail, the choice of tail copulas is restricted by only few constraints and the sensitivities of the AV@R within this class are more significant. In all cases, the worst-case AV@R on the basis of the DM copula provides a substantially better understanding of downside risk than single copulas fitted to the whole data.

## 5 Conclusion

Uncertainty requires suitable techniques for risk assessment. In this paper, we combined stochastic approximation and stochastic average approximation to develop an efficient algorithm to compute the worst case average value at risk in the face of tail uncertainty. Dependence was modelled by the distorted mix method that flexibly assigns different copulas to different regions of multivariate distributions. The method is computationally efficient and allows at the same time to identify copulas in a lower-dimensional mixture space that capture the worst case with high precision. We illustrated the application of our approach in the context of financial markets and cyber risk. Distorted mix copulas can flexibly adjust the dependence structure in different regions of a multivariate distribution. Our research indicated that they provide a powerful and flexible tool for capturing dependence in both the central area and tails of distributions.

## References

- Acciaio, B. & G. Svindland (2013), ‘Are law-invariant risk functions concave on distributions?’, *Dependence Modeling* **1**, 54–64.
- Arbenz, P., P. Embrechts & G. Puccetti (2011), ‘The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables’, *Bernoulli* **17**(2), 562–591.
- Arbenz, P., P. Embrechts & G. Puccetti (2012), ‘The GAEP algorithm for the fast computation of the distribution of a function of dependent random variables’, *Stochastics* **84**(5-6), 569–597.
- Bardou, O.A., N. Frikha & G. Pagès (2009), ‘Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling’, *Monte Carlo Methods and Applications* **15**(3), 173–210.
- Bardou, O.A., N. Frikha & G. Pagès (2016), ‘CVaR hedging using quantization-based stochastic approximation algorithm’, *Mathematical Finance* **26**(1), 184–229.
- Bartl, D., S. Drapeau & L. Tangpi (2020), ‘Computational aspects of robust optimized certainty equivalents and option pricing’, *Mathematical Finance* **30**(1), 287–309.

- Bellini, F. & V. Bignozzi (2015), ‘On elicitable risk measures’, *Quantitative Finance* **15**(5), 725–733.
- Ben-Tal, A. & M. Teboulle (2007), ‘An old-new concept of convex risk measures: The optimized certainty equivalent’, *Mathematical Finance* **17**(3), 449–476.
- Bernard, C. & S. Vanduffel (2015), ‘A new approach to assessing model risk in high dimensions’, *Journal of Banking & Finance* **58**, 166–178.
- Bernard, C., X. Jiang & R. Wang (2014), ‘Risk aggregation with dependence uncertainty’, *Insurance: Mathematics and Economics* **54**, 93–108.
- Bhatnagar, S., H.L. Prasad & L.A. Prashanth (2013), *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods*, Springer.
- Blanchet, J. & K. Murthy (2019), ‘Quantifying distributional model risk via optimal transport’, *Mathematics of Operations Research* **44**(2), 565–600.
- Breuer, T. & I. Csiszár (2016), ‘Measuring distribution model risk’, *Mathematical Finance* **26**(2), 395–411.
- Broadie, M., D. Cicek & A. Zeevi (2011), ‘General bounds and finite-time improvement for the Kiefer-Wolfowitz stochastic approximation algorithm’, *Operations Research* **59**(5), 1211–1224.
- Choulakian, V. & M. A. Stephens (2001), ‘Goodness-of-fit tests for the generalized pareto distribution’, *Technometrics* **43**(4), 478–484.
- Condat, L. (2016), ‘Fast projection onto the simplex and the  $l_1$ -ball’, *Mathematical Programming* **158**(1-2), 575–585.
- Cont, R., R. Deguest & G. Scandolo (2010), ‘Robustness and sensitivity analysis of risk measurement procedures’, *Quantitative Finance* **10**(6), 593–606.
- Dißmann, J., E.C. Brechmann, C. Czado & D. Kurowicka (2013), ‘Selecting and estimating regular vine copulae and application to financial returns’, *Computational Statistics & Data Analysis* **59**, 52 – 69.
- Dunkel, J. & S. Weber (2007), Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models, WSC ’07, IEEE Press, pp. 958–966.
- Dunkel, J. & S. Weber (2010), ‘Stochastic root finding and efficient estimation of convex risk measures’, *Operation Research* **58**(5), 1505–1521.
- Eling, M. & K. Jung (2018), ‘Copula approaches for modeling cross-sectional dependence of data breach losses’, *Insurance: Mathematics and Economics* **82**, 167–180.
- Embrechts, P., G. Puccetti & L. Rüschendorf (2013), ‘Model uncertainty and VaR aggregation’, *Journal of Banking & Finance* **37**(8), 2750–2764.
- Embrechts, P., H. Liu & R. Wang (2018), ‘Quantile-based risk sharing’, *Operations Research* **66**(4), 936–949.
- Erhardt, V. & C. Czado (2012), ‘Modeling dependent yearly claim totals including zero claims in private health insurance’, *Scandinavian Actuarial Journal* (2), 106–129.
- Fu, M.C. (2006), Chapter 19 gradient estimation, in S. G.Henderson & B. L.Nelson, eds, ‘Simulation’, Vol. 13 of *Handbooks in Operations Research and Management Science*, Elsevier, pp. 575 – 616.

- Föllmer, H. & A. Schied (2004), *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin.
- Ghosh, S. & H. Lam (2019), ‘Robust analysis in stochastic simulation: Computation and performance guarantees’, *Operations Research* **67**(1), 232–249.
- Glasserman, P. & X. Xu (2014), ‘Robust risk measurement and model risk’, *Quantitative Finance* **14**(1), 29–58.
- Hamm, A., T. Knispel & S. Weber (2020), ‘Optimal risk sharing in insurance networks’, *European Actuarial Journal* **10**(1), 203–234.
- Hampel, F. R. (1971), ‘A general qualitative definition of robustness’, *The Annals of Mathematical Statistics* **42**(6), 1887–1896.
- Hofert, M., M. Mächler & A.J. McNeil (2013), ‘Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications’, *Journal de la Société Française de Statistique* **154**(1), 25–63.
- Hu, L. (2006), ‘Dependence patterns across financial markets: a mixed copula approach’, *Applied Financial Economics* **16**(10), 717–729.
- Hu, Z. & L.J. Hong (2013), ‘Kullback-Leibler divergence constrained distributionally robust optimization’, *Available at Optimization Online* .
- Huang, P., D. Subramanian & J. Xu (2010), ‘An importance sampling method for portfolio CVaR estimation with Gaussian copula models’, *Proceedings of the 2010 Winter Simulation Conference (WSC)* pp. 2790–2800.
- Kakouris, I. & B. Rustem (2014), ‘Robust portfolio optimization with copulas’, *European Journal of Operational Research* **235**(1), 2837.
- Kiefer, J. & J. Wolfowitz (1952), ‘Stochastic estimation of the maximum of a regression function’, *The Annals of Mathematical Statistics* **23**(3), 462–466.
- Kim, S., R. Pasupathy & S.G. Henderson (2015), A guide to sample average approximation, in F.M.C., ed., ‘Handbook of Simulation Optimization. International Series in Operations Research & Management Science’, Vol. 216, Springer, New York.
- Kou, S., X. Peng & C. C. Heyde (2013), ‘External risk measures and Basel accords’, *Mathematics of Operations Research* **38**(3), 393–417.
- Krätschmer, V., A. Schied & H. Zähle (2014), ‘Comparative and qualitative robustness for law-invariant risk measures’, *Finance and Stochastics* **18**, 271295.
- Kushner, H.J. & G.G. Yin (2003), *Stochastic approximation algorithms and applications*, Springer-Verlag.
- Li, L., H. Shao, R. Wang & J. Yang (2018), ‘Worst-case range value-at-risk with partial information’, *SIAM Journal on Financial Mathematics* **9**(1), 190–218.
- Li, L., K.C. Yuen & J. Yang (2014), ‘Distorted mix method for constructing copulas with tail dependence’, *Insurance: Mathematics and Economics* **57**, 77–89.
- McNeil, A.J., R. Frey & P. Embrechts (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, NJ, USA.

- Meng, F.W., J. Sun & M. Goh (2010), ‘Stochastic optimization problems with CVaR risk measure and their sample average approximation’, *Journal of Optimization Theory and Applications* **146**(2), 399–418.
- Pflug, G.C. (2000), Some remarks on the value-at-risk and the conditional value-at-risk, in S.Uryasev, ed., ‘Probabilistic Constrained Optimization. Nonconvex Optimization and Its Applications’, Vol. 49, Springer, Boston, MA, pp. 272–281.
- Puccetti, G., L. Rüschendorf, D. Small & S. Vanduffel (2017), ‘Reduction of value-at-risk bounds via independence and variance information’, *Scandinavian Actuarial Journal* **2017**(3), 245–266.
- Rockafellar, R.T. & S. Uryasev (2000), ‘Optimization of conditional value-at-risk’, *Journal of Risk* **2**(3), 21–41.
- Rockafellar, R.T. & S. Uryasev (2002), ‘Conditional value-at-risk for general loss distributions’, *Journal of Banking & Finance* **26**, 1443–1471.
- Rüschendorf, L. (2017), Risk bounds and partial dependence information, in D. Fergert, W. González Manteiga, T. Schmidt, J.L. Wang, ed., ‘From Statistics to Mathematical Finance’, Springer, Cham, pp. 345–366.
- Scarrott, C. & A. MacDonald (2012), ‘A review of extreme value threshold estimation and uncertainty quantification’, *Revstat Statistical Journal* **10**(1), 33–60.
- Shapiro, A. (2003), ‘Monte Carlo sampling methods’, *Handbooks in Operations Research and Management Science* **10**, 353–425.
- Sun, H., H. Xu & Y. Wang (2014), ‘Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions’, *Journal of Optimization Theory and Applications* **161**, 257–284.
- Tamar, A., Y. Glassner & S. Mannor (2015), ‘Optimizing the CVaR via sampling’, *Proceedings of the 29th AAAI Conference on Artificial Intelligence* pp. 2993–2999.
- Weber, S. (2006), ‘Distribution-invariant risk measures, information, and dynamic consistency’, *Mathematical Finance* **16**(2), 419–441.
- Weber, S. (2018), ‘Solvency II, or how to sweep the downside risk under the carpet’, *Insurance: Mathematics and Economics* **82**, 191 – 200.
- Zhu, S. & M. Fukushima (2009), ‘Worst-case conditional value-at-risk with application to robust portfolio management’, *Operations Research* **57**(5), 1155–1168.

## A Appendix

### A.1 Proof of Lemma 1

Define  $\zeta(u, \bar{\gamma}) = u + \frac{\alpha_0}{1-p} \mathbb{E}[\Psi^0 - u]^+ + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \mathbb{E}[\Psi^{ij} - u]^+$ , where  $\Psi^0$  and  $\Psi^{ij}$  are random variables having the distribution  $G_0$  and  $G_{ij}$  in (7), respectively. The finiteness of the function  $\zeta$  is guaranteed by the existence of the AV@R, or equivalently by  $\mathbb{E}|\Psi^0| < \infty$  and  $\mathbb{E}|\Psi^{ij}| < \infty$  for each  $i$  and  $j$ . Moreover, a convex function  $\zeta(\cdot, \bar{\gamma})$  has finite right and left derivatives for any  $\bar{\gamma}$ . Observe that

$$\frac{\zeta(u', \bar{\gamma}) - \zeta(u, \bar{\gamma})}{u' - u} = 1 + \frac{\alpha_0}{1-p} \frac{\mathbb{E}[\Psi^0 - u']^+ - \mathbb{E}[\Psi^0 - u]^+}{u' - u} + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \frac{\mathbb{E}[\Psi^{ij} - u']^+ - \mathbb{E}[\Psi^{ij} - u]^+}{u' - u}.$$

When  $u' > u$ ,

$$\frac{\mathbb{E}[\Psi^0 - u']^+ - \mathbb{E}[\Psi^0 - u]^+}{u' - u} = \begin{cases} -1 & \text{if } \Psi^0 \geq u' \\ 0 & \text{if } \Psi^0 \leq u \\ \mathbb{E} \left[ \frac{-\Psi^0 + u}{u' - u} \right] \in (-1, 0) & \text{if } u < \Psi^0 < u'. \end{cases}$$

Then there exist  $\rho(u, u') \in [0, 1]$  for which

$$\frac{\mathbb{E}[\Psi^0 - u']^+ - \mathbb{E}[\Psi^0 - u]^+}{u' - u} = -(1 - \mathbb{P}(\Psi^0 \leq u')) - \rho(u, u')(\mathbb{P}(\Psi^0 \leq u') - \mathbb{P}(\Psi^0 \leq u)).$$

By letting  $u' \downarrow u$ , we have  $\mathbb{P}(\Psi^0 \leq u')$  converges to  $\mathbb{P}(\Psi^0 \leq u)$  which makes

$$\lim_{u' \downarrow u} \frac{\mathbb{E}[\Psi^0 - u']^+ - \mathbb{E}[\Psi^0 - u]^+}{u' - u} = \mathbb{P}(\Psi^0 \leq u) - 1.$$

Similarly, we can compute

$$\begin{aligned} \lim_{u' \downarrow u} \frac{\zeta(u', \bar{\gamma}) - \zeta(u, \bar{\gamma})}{u' - u} &= 1 + \frac{\alpha_0}{1-p} (\mathbb{P}(\Psi^0 \leq u) - 1) + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} (\mathbb{P}(\Psi^{ij} \leq u) - 1) \\ &= 1 - \frac{1}{1-p} + \frac{\alpha_0}{1-p} \mathbb{P}(\Psi^0 \leq u) + \sum_{i=1}^m \sum_{j=1}^K \frac{\alpha_i \gamma_j^i}{1-p} \mathbb{P}(\Psi^{ij} \leq u) \end{aligned}$$

which is  $\frac{\partial^+ \zeta}{\partial u}(u, \bar{\gamma})$ . Analogously, we can compute  $\frac{\partial^- \zeta}{\partial u}(u, \bar{\gamma})$ . The remaining results are now straightforward.

### A.2 Calibrations with large amounts of data

We provide an illustrative example supporting the claim that the DM method provides a good statistical framework for estimating the risk measure AV@R if a large amount of data is available. This claim refers, of course, to the DM method of Li et al. (2014) itself. Computing a worst-case is thus not an issue in this case study.

We consider a setting that modifies Example 2 ( $d = 2$ ) as follows. Data are generated by a collection of models with  $X = X_1 + X_2$  where the dependence between  $X_1$  and  $X_2$  is correctly described by one of the following copulas

- a Gaussian copula with  $\rho = 0.7$ ;
- a t-copula with  $\rho = 0.7, \nu = 1$ ;

- a Gumbel copula with  $\theta = 1.7095$ ;
- a Cuadras-Augé copula with  $\theta = 0.8$ .

The Cuadras-Augé copula

$$C^{CA}(u_1, u_2) = \min(u_1, u_2)^\theta (u_1 u_2)^{1-\theta}, \quad \theta \in (0, 1],$$

is upper tail dependent and an extreme value copula. The marginal distributions of  $X_1$  and  $X_2$  are inverse Gaussian with parameters  $(\mu, \lambda)$  equal to  $(1, 0.5)$  and  $(1, 1.2)$ , respectively. For each model we compute the AV@R from SAA using the  $5 \cdot 10^7$  number of samples as an approximation of the true AV@R; the results are displayed on the second row of Table 6 and denoted by ‘True AV@R’.

The numerical experiment is then conducted as follows. We generate data with sample size  $10^3$  and  $10^4$  from the given model (the given true copula). These are then used to estimate parameters of the following copulas and to finally compute AV@R measurements from them using a sample size  $10^6$ :

- a Gaussian DM copula with  $\alpha_1 = 0.1$ , whose component copulas for the three parts,  $D_0, D_1, D_2$  are all Gaussian copulas;
- a Gaussian DM copula for the optimal  $\alpha_1$  (as explained below);
- a Gaussian copula;
- a t-copula;
- a Gumbel copula.

Letting  $\alpha_0 = 1 - \alpha_1$ , and  $\alpha_1 = \alpha_2$ , the distortion functions are:

$$D_0(x) = \begin{cases} \frac{x}{\alpha_0} & \text{if } x \leq \alpha_0 \\ 1 & \text{if } x > \alpha_0 \end{cases} \quad D_1(x) = \begin{cases} 0 & \text{if } x \leq \alpha_0 \\ \frac{x-\alpha_0}{\alpha_1} & \text{if } \alpha_0 < x \leq \alpha_0 + \alpha_1 \\ 1 & \text{if } x > \alpha_0 + \alpha_1 \end{cases}$$

$$D_2(x) = \begin{cases} 0 & \text{if } x \leq \alpha_0 + \alpha_1 \\ \frac{x-\alpha_0-\alpha_1}{\alpha_2} & \text{if } \alpha_0 + \alpha_1 < x \leq 1. \end{cases}$$

The results of the parameter estimation are displayed in Table 5.

Data size	Copulas	Parameters	Gaussian	t	Gumbel	Cuadras-Augé
$10^3$	Gaussian DM	$\alpha_1$	0.1	0.1	0.1	0.1
		$\hat{\rho}$ for $D_0$	0.5407	0.5833	0.3273	0.6718
		$\hat{\rho}$ for $D_1$	-0.6668	-0.5114	-0.7146	-0.5774
		$\hat{\rho}$ for $D_2$	0.016	0.1752	0.2868	0.0187
	Gaussian DM	$\alpha_1$	0.2	0.12	0.15	0.1
		$\hat{\rho}$ for $D_0$	0.4514	0.5849	0.2602	0.6718
		$\hat{\rho}$ for $D_1$	-0.6254	-0.5371	-0.6672	-0.5774
		$\hat{\rho}$ for $D_2$	0.1467	0.1889	0.2683	0.0187
	Gaussian	$\hat{\rho}$	0.7032	0.7002	0.5973	0.761
		$\hat{\nu}$	3338586	1.0062	5.3304	1
	t	$\hat{\rho}$	0.6991	0.7102	0.577	1
		$\hat{\theta}$	1.8341	2.1882	1.6542	3.1677
	$10^4$	Gaussian DM	$\alpha_1$	0.1	0.1	0.1
$\hat{\rho}$ for $D_0$			0.5652	0.5936	0.3456	0.1191
$\hat{\rho}$ for $D_1$			-0.7711	-0.5744	-0.6837	-0.562
$\hat{\rho}$ for $D_2$			-0.1119	-0.1026	0.1121	0.1368
Gaussian DM		$\alpha_1$	0.15	0.12	0.13	0.1
		$\hat{\rho}$ for $D_0$	0.5185	0.6076	0.3006	0.1191
		$\hat{\rho}$ for $D_1$	-0.6983	-0.605	-0.6758	-0.562
		$\hat{\rho}$ for $D_2$	-0.0157	-0.1003	0.1191	0.1368
Gaussian		$\hat{\rho}$	0.7028	0.6281	0.5906	0.742
		$\hat{\nu}$	4669186	1.0107	7.7096	1
t		$\hat{\rho}$	0.7064	0.6973	0.5938	1
		$\hat{\theta}$	1.8476	2.0485	1.6837	2.9518
Gumbel		$\hat{\rho}$	0.7028	0.6281	0.5906	0.742
	$\hat{\theta}$	1.8476	2.0485	1.6837	2.9518	

Table 5: Estimated parameters.



True copulas		Gaussian	t	Gumbel	Cuadras-Augé	
Data size	True AV@R	8.8405	9.0755	9.0728	9.2205	
$10^3$	Gaussian DM AV@R ( $\alpha_1 = 0.1$ )	9.2404	9.2879	9.3345	9.2093	
	increment to true AV@R	-0.3999	-0.2125	-0.2617	0.0112	
	Gaussian DM AV@R (optimal $\alpha_1$ )	8.7948	9.1645	9.0836	9.2093	
	increment to true AV@R	(0.2)	(0.12)	(0.15)	(0.1)	
	increment to true AV@R	0.0457	-0.0890	-0.0107	0.0112	
	Gaussian AV@R	8.8089	8.5252	8.5674	8.9012	
	increment to true AV@R	0.0316	0.5503	0.5054	0.3193	
	t AV@R	8.8132	9.0212	8.6768	9.6490	
	increment to true AV@R	0.0273	0.0543	0.3960	-0.4285	
	Gumbel AV@R	9.1872	9.3263	9.0231	9.5304	
	increment to true AV@R	-0.3467	-0.2508	0.0498	-0.3098	
	$10^4$	Gaussian DM AV@R ( $\alpha_1 = 0.1$ )	9.1480	9.1498	9.2478	9.2560
		increment to true AV@R	-0.3075	-0.0744	-0.1750	-0.0467
		Gaussian DM AV@R (optimal $\alpha_1$ )	8.8558	9.0162	9.0541	9.2560
increment to true AV@R		(0.15)	(0.12)	(0.13)	(0.1)	
increment to true AV@R		-0.0153	0.0593	0.0187	-0.0467	
Gaussian AV@R		8.8574	8.6748	8.5548	8.9677	
increment to true AV@R		-0.0169	0.4007	0.5181	0.2529	
t AV@R		8.8647	9.0564	8.6662	9.6537	
increment to true AV@R		-0.0242	0.0190	0.4067	-0.4331	
Gumbel AV@R		9.2021	9.2689	9.0558	9.5103	
increment to true AV@R		-0.3616	-0.1935	0.0171	-0.2897	

Table 6: AV@R values in estimated models and increments to true AV@R.

The AV@Rs calculated in various estimated models and the resulting increments to the true AV@R are displayed in Table 6. The Gaussian model with optimal  $\alpha_1$  is estimated using the empirical maximal likelihood. The main observation is that, despite being based on Gaussian copulas only, the DM method performs quite well in estimating the true AV@R. In particular, the Gaussian DM AV@R with optimal  $\alpha_1$  outperforms all other models. This is particularly striking in the case of the Cuadras-Augé copula, an extreme value copula. All copulas including the Gumbel copula perform worse in this case.

### A.3 Data in Section 4.1

#### Dependence in the central part

Dependence in the central part is modeled as the Gaussian copula whose correlation matrix consists of the estimated linear correlations. The estimated correlation matrix based on the 92% data is

$$\Sigma = \begin{bmatrix} 1 & 0.9170 & 0.9494 & 0.4656 & 0.4706 & 0.5211 & 0.5012 \\ 0.9170 & 1 & 0.8185 & 0.4599 & 0.4495 & 0.5072 & 0.4724 \\ 0.9494 & 0.8185 & 1 & 0.4440 & 0.4389 & 0.4799 & 0.4698 \\ 0.4656 & 0.4599 & 0.4440 & 1 & 0.9152 & 0.2288 & 0.3191 \\ 0.4706 & 0.4495 & 0.4389 & 0.9152 & 1 & 0.2295 & 0.3060 \\ 0.5211 & 0.5072 & 0.4799 & 0.2288 & 0.2295 & 1 & 0.5717 \\ 0.5012 & 0.4724 & 0.4698 & 0.3191 & 0.3060 & 0.5717 & 1 \end{bmatrix}.$$

## Dependence in the upper tail part

We consider  $K = 16$  candidate copulas in the tail parts. The calibration results are summarized in the following.

- $C_1$ : Gaussian copula matching the estimated linear correlation in the upper (4%-8%) data

$$\Sigma_1 = \begin{bmatrix} 1 & 0.8316 & 0.8727 & -0.2373 & -0.1814 & -0.1005 & -0.1121 \\ 0.8316 & 1 & 0.6434 & -0.1834 & -0.1717 & -0.0088 & -0.0635 \\ 0.8727 & 0.6434 & 1 & -0.2906 & -0.2663 & -0.2072 & -0.1561 \\ -0.2373 & -0.1834 & -0.2906 & 1 & 0.7351 & -0.2962 & -0.2112 \\ -0.1814 & -0.1717 & -0.2663 & 0.7351 & 1 & -0.2406 & -0.2076 \\ -0.1005 & -0.0088 & -0.2072 & -0.2962 & -0.2406 & 1 & 0.0762 \\ -0.1121 & -0.0635 & -0.1561 & -0.2112 & -0.2076 & 0.0762 & 1 \end{bmatrix}.$$

- $C_2$ : Gaussian copula matching the estimated Kendall's tau in the upper (4%-8%) data

$$\Sigma_2 = \begin{bmatrix} 1 & 0.8011 & 0.8571 & -0.2492 & -0.2328 & -0.1505 & -0.0449 \\ 0.8011 & 1 & 0.5947 & -0.1970 & -0.2012 & -0.0481 & -0.0090 \\ 0.8571 & 0.5947 & 1 & -0.3024 & -0.3445 & -0.1766 & -0.0952 \\ -0.2492 & -0.1970 & -0.3024 & 1 & 0.7583 & -0.3262 & -0.2632 \\ -0.2328 & -0.2012 & -0.3445 & 0.7583 & 1 & -0.2438 & -0.2332 \\ -0.1505 & -0.0481 & -0.1766 & -0.3262 & -0.2438 & 1 & 0.0580 \\ -0.0449 & -0.0090 & -0.0952 & -0.2632 & -0.2332 & 0.0580 & 1 \end{bmatrix}.$$

- $C_3$ : t-copula estimated from the upper (4%-8%) data using ML

$$\nu^1 = 10.8802, \quad P^1 = \begin{bmatrix} 1 & 0.9763 & 0.9808 & 0.7991 & 0.8093 & 0.7712 & 0.7786 \\ 0.9763 & 1 & 0.9461 & 0.7950 & 0.7996 & 0.7855 & 0.7806 \\ 0.9808 & 0.9461 & 1 & 0.7725 & 0.7811 & 0.7382 & 0.7615 \\ 0.7991 & 0.7950 & 0.7725 & 1 & 0.9529 & 0.6949 & 0.7192 \\ 0.8093 & 0.7996 & 0.7811 & 0.9529 & 1 & 0.7050 & 0.7172 \\ 0.7712 & 0.7855 & 0.7382 & 0.6949 & 0.7050 & 1 & 0.7332 \\ 0.7786 & 0.7806 & 0.7615 & 0.7192 & 0.7172 & 0.7332 & 1 \end{bmatrix}.$$

- $C_4$ : t-copula estimated from the upper (4%-8%) data using approximate ML

$$\nu^2 = 4.8409, \quad P^2 = \begin{bmatrix} 1 & 0.9884 & 0.9910 & 0.8914 & 0.8922 & 0.8662 & 0.8741 \\ 0.9884 & 1 & 0.9745 & 0.8876 & 0.8870 & 0.8765 & 0.8761 \\ 0.9910 & 0.9745 & 1 & 0.8813 & 0.8800 & 0.8507 & 0.8679 \\ 0.8914 & 0.8876 & 0.8813 & 1 & 0.9762 & 0.8283 & 0.8439 \\ 0.8922 & 0.8870 & 0.8800 & 0.9762 & 1 & 0.8318 & 0.8405 \\ 0.8662 & 0.8765 & 0.8507 & 0.8283 & 0.8318 & 1 & 0.8459 \\ 0.8741 & 0.8761 & 0.8679 & 0.8439 & 0.8405 & 0.8459 & 1 \end{bmatrix}.$$

- $C_5$ : t-copula estimated from the upper (4%-8%) data using Kendall's tau and ML

$$\nu^3 = 1.1237, \quad P^3 = \begin{bmatrix} 1 & 0.8011 & 0.8571 & -0.2492 & -0.2328 & -0.1505 & -0.0449 \\ 0.8011 & 1 & 0.5947 & -0.1970 & -0.2012 & -0.0481 & -0.0090 \\ 0.8571 & 0.5947 & 1 & -0.3024 & -0.3445 & -0.1766 & -0.0952 \\ -0.2492 & -0.1970 & -0.3024 & 1 & 0.7583 & -0.3262 & -0.2632 \\ -0.2328 & -0.2012 & -0.3445 & 0.7583 & 1 & -0.2438 & -0.2332 \\ -0.1505 & -0.0481 & -0.1766 & -0.3262 & -0.2438 & 1 & 0.0580 \\ -0.0449 & -0.0090 & -0.0952 & -0.2632 & -0.2332 & 0.0580 & 1 \end{bmatrix}.$$

- $C_6$ : Grouped t-copula estimated from the upper (4%-8%) data using ML

$$\boldsymbol{\nu}^1 = (4.8284, 4.4677, 12.8199),$$

$$\mathbf{P}^1 = \begin{bmatrix} 1 & 0.9734 & 0.9780 \\ 0.9734 & 1 & 0.9388 \\ 0.9780 & 0.9388 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9472 \\ 0.9472 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.7655 \\ 0.7655 & 1 \end{bmatrix}.$$

- $C_7$ : Grouped t-copula estimated from the upper (4%-8%) data using approximate ML

$$\boldsymbol{\nu}^2 = (1.2569, 1.7534, 13040952.7492),$$

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 0.9713 & 0.9793 \\ 0.9713 & 1 & 0.9375 \\ 0.9793 & 0.9375 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9682 \\ 0.9682 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8517 \\ 0.8517 & 1 \end{bmatrix}.$$

- $C_8$ : Grouped t-copula estimated from the upper (4%-8%) data using Kendall's tau and ML

$$\boldsymbol{\nu}^3 = (0.7287, 0.9180, 1.9547),$$

$$\mathbf{P}^3 = \begin{bmatrix} 1 & 0.8011 & 0.8571 \\ 0.8011 & 1 & 0.5947 \\ 0.8571 & 0.5947 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.7583 \\ 0.7583 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.0580 \\ 0.0580 & 1 \end{bmatrix}.$$

- $C_9$ : Gaussian copula matching the estimated linear correlation in the extreme upper (-4%) data

$$\Sigma_3 = \begin{bmatrix} 1 & 0.9197 & 0.9592 & 0.5265 & 0.5336 & 0.5598 & 0.5913 \\ 0.9197 & 1 & 0.8860 & 0.4640 & 0.5141 & 0.5420 & 0.5060 \\ 0.9592 & 0.8860 & 1 & 0.4624 & 0.4786 & 0.4820 & 0.5998 \\ 0.5265 & 0.4640 & 0.4624 & 1 & 0.8503 & 0.5107 & 0.4501 \\ 0.5336 & 0.5141 & 0.4786 & 0.8503 & 1 & 0.4144 & 0.3761 \\ 0.5598 & 0.5420 & 0.4820 & 0.5107 & 0.4144 & 1 & 0.5388 \\ 0.5913 & 0.5060 & 0.5998 & 0.4501 & 0.3761 & 0.5388 & 1 \end{bmatrix}.$$

- $C_{10}$ : Gaussian copula matching the estimated the Kendall's tau in the extreme upper (-4%) data

$$\Sigma_4 = \begin{bmatrix} 1 & 0.9088 & 0.9498 & 0.5068 & 0.4949 & 0.5044 & 0.5333 \\ 0.9088 & 1 & 0.8722 & 0.4565 & 0.4878 & 0.5028 & 0.4495 \\ 0.9498 & 0.8722 & 1 & 0.4615 & 0.4213 & 0.3868 & 0.5318 \\ 0.5068 & 0.4565 & 0.4615 & 1 & 0.8604 & 0.5050 & 0.4718 \\ 0.4949 & 0.4878 & 0.4213 & 0.8604 & 1 & 0.4223 & 0.3349 \\ 0.5044 & 0.5028 & 0.3868 & 0.5050 & 0.4223 & 1 & 0.4835 \\ 0.5333 & 0.4495 & 0.5318 & 0.4718 & 0.3349 & 0.4835 & 1 \end{bmatrix}.$$

- $C_{11}$ : t-copula estimated from the extreme upper (-4%) data using ML

$$\nu_e^1 = 25.9712, \quad P_e^1 = \begin{bmatrix} 1 & 0.9824 & 0.9909 & 0.8853 & 0.8961 & 0.8843 & 0.8906 \\ 0.9824 & 1 & 0.9738 & 0.8691 & 0.8926 & 0.8790 & 0.8685 \\ 0.9909 & 0.9738 & 1 & 0.8688 & 0.8825 & 0.8620 & 0.8919 \\ 0.8853 & 0.8691 & 0.8688 & 1 & 0.9553 & 0.8532 & 0.8378 \\ 0.8961 & 0.8926 & 0.8825 & 0.9553 & 1 & 0.8324 & 0.8264 \\ 0.8843 & 0.8790 & 0.8620 & 0.8532 & 0.8324 & 1 & 0.8620 \\ 0.8906 & 0.8685 & 0.8919 & 0.8378 & 0.8264 & 0.8620 & 1 \end{bmatrix}.$$

- $C_{12}$ : t-copula estimated from the extreme upper (-4%) data using approximate ML

$$\nu_e^2 = 2.8781, \quad P_e^2 = \begin{bmatrix} 1 & 0.9864 & 0.9950 & 0.9062 & 0.9175 & 0.9321 & 0.9239 \\ 0.9864 & 1 & 0.9804 & 0.8876 & 0.9103 & 0.9243 & 0.9044 \\ 0.9950 & 0.9804 & 1 & 0.8965 & 0.9092 & 0.9202 & 0.9280 \\ 0.9062 & 0.8876 & 0.8965 & 1 & 0.9692 & 0.8957 & 0.8760 \\ 0.9175 & 0.9103 & 0.9092 & 0.9692 & 1 & 0.8838 & 0.8749 \\ 0.9321 & 0.9243 & 0.9202 & 0.8957 & 0.8838 & 1 & 0.9134 \\ 0.9239 & 0.9044 & 0.9280 & 0.8760 & 0.8749 & 0.9134 & 1 \end{bmatrix}.$$

- $C_{13}$ : t-copula estimated from the extreme upper (-4%) data using Kendall's tau and ML

$$\nu_e^3 = 2.2042, \quad P_e^3 = \begin{bmatrix} 1 & 0.9088 & 0.9498 & 0.5068 & 0.4949 & 0.5044 & 0.5333 \\ 0.9088 & 1 & 0.8722 & 0.4565 & 0.4878 & 0.5028 & 0.4495 \\ 0.9498 & 0.8722 & 1 & 0.4615 & 0.4213 & 0.3868 & 0.5318 \\ 0.5068 & 0.4565 & 0.4615 & 1 & 0.8604 & 0.5050 & 0.4718 \\ 0.4949 & 0.4878 & 0.4213 & 0.8604 & 1 & 0.4223 & 0.3349 \\ 0.5044 & 0.5028 & 0.3868 & 0.5050 & 0.4223 & 1 & 0.4835 \\ 0.5333 & 0.4495 & 0.5318 & 0.4718 & 0.3349 & 0.4835 & 1 \end{bmatrix}.$$

- $C_{14}$ : Grouped t-copula estimated from the extreme upper (-4%) data using ML

$$\nu_e^1 = (6.7057, 109.1713, 709778.4720),$$

$$P_e^1 = \begin{bmatrix} 1 & 0.9781 & 0.9890 \\ 0.9781 & 1 & 0.9668 \\ 0.9890 & 0.9668 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9580 \\ 0.9580 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8848 \\ 0.8848 & 1 \end{bmatrix}.$$

- $C_{15}$ : : Grouped t-copula estimated from the extreme upper (-4%) data using approximate ML

$$\nu_e^2 = (1.0001, 1.1370, 1.7893),$$

$$P_e^2 = \begin{bmatrix} 1 & 0.9510 & 0.9792 \\ 0.9510 & 1 & 0.9234 \\ 0.9792 & 0.9234 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9301 \\ 0.9301 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9092 \\ 0.9092 & 1 \end{bmatrix}.$$

- $C_{16}$ : : Grouped t-copula estimated from the extreme upper (-4%) data using Kendall's tau and ML

$$\nu_e^3 = (1.2977, 1.2977, 1.5911),$$

$$P_e^3 = \begin{bmatrix} 1 & 0.9088 & 0.9498 \\ 0.9088 & 1 & 0.8722 \\ 0.9498 & 0.8722 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8604 \\ 0.8604 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.4835 \\ 0.4835 & 1 \end{bmatrix}.$$

**Single Gaussian copula estimated from the entire data set – correlation matrix**

$$\begin{bmatrix} 1 & 0.9388 & 0.9625 & 0.5961 & 0.6031 & 0.6302 & 0.6172 \\ 0.9388 & 1 & 0.8662 & 0.5906 & 0.5865 & 0.6197 & 0.5946 \\ 0.9625 & 0.8662 & 1 & 0.5775 & 0.5771 & 0.5965 & 0.5922 \\ 0.5961 & 0.5906 & 0.5775 & 1.0000 & 0.9333 & 0.4003 & 0.4699 \\ 0.6031 & 0.5865 & 0.5771 & 0.9333 & 1 & 0.4035 & 0.4623 \\ 0.6302 & 0.6197 & 0.5965 & 0.4003 & 0.4035 & 1.0000 & 0.6579 \\ 0.6172 & 0.5946 & 0.5922 & 0.4699 & 0.4623 & 0.6579 & 1 \end{bmatrix}.$$

## A.4 Data in Section 4.2

### Dependence in the central part

Dependence in the central part is modeled as the Gaussian copula whose correlation matrix consists of the estimated linear correlations as

$$\Sigma_1 = \begin{bmatrix} 1 & -0.0086 & -0.0224 & 0.0260 & -0.3324 \\ -0.0086 & 1 & 0.1179 & -0.0210 & -0.2222 \\ -0.0224 & 0.1179 & 1 & -0.1795 & 0.2620 \\ 0.0260 & -0.0210 & -0.1795 & 1 & -0.1342 \\ -0.3324 & -0.2222 & 0.2620 & -0.1342 & 1 \end{bmatrix}.$$

### Dependence in the tail parts

- $\Sigma_2 = \begin{bmatrix} 1 & 0.0605 & 0.1459 & 0.0437 & -0.1554 \\ 0.0605 & 1 & -0.0233 & 0.0313 & -0.2593 \\ 0.1459 & -0.0233 & 1 & -0.1787 & 0.3949 \\ 0.0260 & 0.0313 & -0.1787 & 1 & -0.1677 \\ -0.1554 & -0.2593 & 0.3949 & -0.1677 & 1 \end{bmatrix}$

- $\nu_1 = 27.5747$ ,  $P_1 = \begin{bmatrix} 1 & 0.6895 & 0.2085 & 0.5510 & 0.3492 \\ 0.6895 & 1 & 0.2257 & 0.6665 & 0.4838 \\ 0.2085 & 0.2257 & 1 & -0.3175 & -0.4300 \\ 0.5510 & 0.6665 & -0.3175 & 1 & 0.8533 \\ 0.3492 & 0.4838 & -0.4300 & 0.8533 & 1 \end{bmatrix}$

- $\nu_2 = 3.6372$ ,  $P_2 = \Sigma_2$

- $C_7$ : The regular vine copula is estimated according to AIC. For more information, we refer to Dißmann et al. (2013). The estimation was conducted by the vine copula package in R. <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>. The selected trees, pair copulas and the estimated parameters are given in Table 7.
- $C_8$ : The regular vine copula is estimated according to BIC. For more information, we refer to Dißmann et al. (2013). The estimation was conducted by the vine copula package in R. <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>. The selected trees, pair copulas and the estimated parameters are provided in Table 8.

Tree	pair	copula	parameters	
1	3,4	Frank	3.63	
	5,2	Frank	6.12	2
	5,1	Frank	4.39	0.06
	5,3	Tawn type 2 180 degrees	3.57	0.46
2	5,4 ; 3	Tawn type 2	4.12	0.37
	1,2 ; 5	Tawn type 2 180 degrees	1.78	0.41
	3,1 ; 5	Survival BB8	6	0.17
3	1,4 ; 5,3	Tawn type 1	3.61	0.39
	3,2 ; 1,5	Joe	1.11	
4	2,4 ; 1,5,3	Tawn type 2 180 degrees	1.6	0.31

Table 7: The structure, pair copulas, and parameters of the regular vine copula  $C_7$  estimated according to AIC

Tree	pair	copula	parameters	
1	3,4	Frank	3.63	
	5,2	Frank	6.12	2
	5,1	Frank	4.39	0.06
	5,3	Tawn type 2 180 degrees	3.57	0.46
2	5,4 ; 3	Tawn type 2	4.12	0.37
	1,2 ; 5	Survival Joe	1.55	
	3,1 ; 5	Independence		
3	1,4 ; 5,3	Tawn type 1	3.49	0.39
	3,2 ; 1,5	Independence		
4	2,4 ; 1,5,3	Clayton	0.52	

Table 8: The structure, pair copulas, and parameters for the regular vine copula  $C_8$  according to BIC