Optimal Expected-Shortfall Portfolio Selection With Copula Induced Dependence

Irène Gijbels^{1,*} and Klaus Herrmann^{1,2}

¹KU Leuven, Department of Mathematics and Leuven Statistics Research Centre (LStat), Belgium

²Concordia University, Department of Mathematics and Statistics, Montréal, Canada.

Abstract: We provide a computational framework for the selection of weights $(\omega_1, \dots, \omega_d)$ that minimize the Expected-Shortfall of the aggregated risk $Z = \sum_{i=1}^d \omega_i X_i$. Contrary to classic and recent results we neither restrict the marginal distributions nor the dependence structure of $(X_1, \dots, X_d)$ to any specific type. While the margins can be set to any absolutely continuous random variable with finite expectation, the dependence structure can be modelled by any absolutely continuous copula function. A real world application to portfolio selection illustrates the usability of the new framework.

Keywords: Aggregation of Risk, Capital Allocation, Copula, Dependence, Portfolio Selection, Smolyak Integration, Sums of Random Variables.

2010 Mathematics Subject Classification: 62E15, 62E17, 62Hxx

1 Introduction

In this paper, we revisit the portfolio selection problem, where an investor needs to decide how to invest a given amount into a portfolio of stocks out of all possible asset combinations. The seminal solution proposed by Markowitz (1952) was extended in various ways, two of which are relevant starting points for our investigation. First, the portfolio choice needs to be based on an appropriate optimization criterion. While Markowitz (1952) relied on the portfolio variance, better suited alternatives such as the Value-at-Risk and the Expected-Shortfall are now commonly accepted, see e.g. McNeil et al. (2015). The Expected-Shortfall enjoys favourable

*Corresponding Author. KU Leuven, Statistics Section, Celestijnenlaan 200b - bus 2400, 3001 Heverlee, Belgium. E-mail: irene.gijbels@wis.kuleuven.be.

This research is supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy), and the project GOA/12/014 of the KU Leuven Research Fund. Part of the research was carried out when the second author visited RiskLab at ETH Zürich, supported by a research stay grant from the Flemish Science Foundation (FWO). The second author thanks Professor Embrechts for hosting him at RiskLab. He also acknowledges indirect support by NSERC grants RGPIN-5010-2015 and RGPIN-2015-05447. The authors thank an anonymous reviewer for valuable comments on a previous version.

theoretical properties as a coherent risk measure in the sense of Artzner et al. (1999). Second, the joint normality assumption of Markowitz (1952) is incompatible with empirical findings on joint asset returns. In recent years, copulas have been a popular means through which more realistic models for joint asset return distributions can be constructed; see e.g. Nelsen (2006).

There has been notable success in providing specific models that allow to compute optimal portfolios addressing both of the above points, see e.g. Bernardi (2013) and Cossette et al. (2013) for recent contributions, and the latter also for an overview of the topic. However, a typical drawback of the current approaches is that the marginal distributions cannot be chosen independently from the copula of the resulting joint asset return distribution, which severely limits the flexibility of these models.

The aim of this paper is to introduce a framework that does not place any restrictions on the margins or the copula of the joint asset returns, while still allowing to carry out an Expected-Shortfall based portfolio optimization. We contribute to the existing literature in three ways.

First, we connect the Expected-Shortfall optimization based approach of Rockafellar and Uryasev (2000) with the computational framework of Gijbels and Herrmann (2014). This leads to a portfolio optimization framework that is Expected-Shortfall based, as well as fully flexible in terms of the underlying distribution of the joint asset returns. Second, we add to the applicability of the approach of Gijbels and Herrmann (2014) by providing necessary auxiliary functions for the most popular copula families, as well as further generalizing their formulas when necessary. Third, our results show how sensitive the optimization results are to model misspecification in the given context, even when closely comparable distributions are considered. This adds further evidence against the application of simplifying modelling assumptions, such as joint normality, in a risk management context.

The remainder of this paper is organized as follows. In Section 2, we introduce the portfolio selection problem and relate our approach to the existing literature. Section 3 reviews necessary properties of sums of dependent random variables when dependence is specified by an absolutely continuous copula. The numerical treatment of our approach is discussed in Section 4, where Section 4.1 provides the necessary auxiliary functions for popular copula families and Section 4.2 deals with the actual numerical computation. We illustrate our approach with examples in Section 5.1, while Section 5.2 provides a real data application where we compute Expected-

Shortfall optimal weights for a portfolio consisting of Apple, Microsoft, IBM, Bank of America Corporation, and Pacific Gas and Electric Company shares. Final conclusions are drawn in Section 6.

2 Risk Measures and Portfolio Selection

2.1 Notation

To formalize the portfolio choice problem, we introduce the aggregated risk as a weighted sum of random variables

$$Z = Z(\boldsymbol{\omega}) := \sum_{i=1}^d \omega_i X_i,$$

where the random vector $\mathbf{X} = (X_1, \dots, X_d)^T$ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Throughout the text, all vectors are understood as column vectors, and T denotes the transpose of a vector or matrix. The joint density and distribution function of $\mathbf{X}$ are denoted by h and H respectively. We assume that all random variables X_i are absolutely continuous with respect to the Lebesgue measure, denoted by $\boldsymbol{\lambda}$, with density, distribution and quantile function denoted by f_i , F_i and F_i^{-1} . For Z we likewise use f_Z , F_Z and F_Z^{-1} . We call $\boldsymbol{\omega} = (\omega_i)_{i=1}^d \in \mathbb{R}^d$ the portfolio weights, or weights for short, and in unambiguous cases the dependence of Z on $\boldsymbol{\omega}$ will be omitted to streamline the presentation.

2.2 Setting of the Optimization Problem

We consider an agent who seeks to invest a fixed amount in a portfolio composed of different assets with joint gross returns $\mathbf{X}$. The weights $\boldsymbol{\omega} = (\omega_i)_{i=1}^d$ represent the decision variables that determine the gross return of the portfolio $Z(\boldsymbol{\omega})$, and their optimal value will be obtained under the classical portfolio selection constraints

$$(C1): \omega_i \geq 0 \text{ for all } i \in \{1, \dots, d\},$$

$$(C2): \sum_{i=1}^d \omega_i = 1,$$

$$(C3): \mathbb{E}[Z(\boldsymbol{\omega})] = \sum_{i=1}^d \omega_i \mathbb{E}[X_i] \geq r \text{ for a given } r \in \mathbb{R}.$$

Given the constraints on the decision variables $(\omega_i)_{i=1}^d$ we denote by $\mathcal{X}_r \subseteq \mathbb{R}^d$ the set of all weights satisfying (C1)–(C3). In the case where only (C1) and (C2) are imposed, it should be

understood that the index r in $\mathcal{X}_r$ is superfluous and hence is dropped. Constraint (C1) prohibits short selling of assets, but we discuss in Section 3 how negative weights can be incorporated. Constraint (C2) formalizes that the amounts invested in the different assets add up to the a priori fixed investment amount. Taking constraints (C1) and (C2) together, we directly see that $\mathcal{X}_r$ is contained in the standard $d - 1$ simplex. Finally, constraint (C3) guarantees a minimum expected return on the investment. The effect of (C3) is further visualized in the examples in Section 5.

2.3 Portfolio Risk Measures and Optimization

Denoting by ρ a risk measure applicable to $Z(\omega)$ for all $\omega \in \mathcal{X}_r$, we call weights ω^* optimal, if $\rho(Z(\omega^*)) \geq \rho(Z(\omega))$ for all $\omega \in \mathcal{X}_r$. We call the random variable $Z^* = Z(\omega^*)$ the optimal portfolio with respect to ρ . Clearly, the choice of ρ plays a pivotal role in the portfolio selection, with alternative choices leading to different advantages and disadvantages. In the classical approach of Markowitz (1952), the variance of Z is used to construct the corresponding risk measure, leading to closed form solutions in the case of joint normality. However, as shown in Embrechts et al. (2002), the mean-variance reasoning fails when $\mathbf{X}$ does not follow an elliptical joint distribution. We therefore focus on the Expected-Shortfall¹ (ES) as an alternative. For $p \in (0, 1)$, the Value-at-Risk (VaR) at level p of Z is given by the corresponding quantile function

$$\text{VaR}_p(Z(\omega)) := F_{Z(\omega)}^{-1}(p) := \inf\{z \in \mathbb{R} : F_{Z(\omega)}(z) \geq p\}.$$

The Expected-Shortfall at level p is then defined as the conditional expectation of Z , given that Z falls below the Value-at-Risk at level p , i.e.

$$\text{ES}_p(Z(\omega)) := \mathbb{E}[Z(\omega) | Z(\omega) \leq \text{VaR}_p(Z(\omega))].$$

The calculation of ES and ES-based capital allocation has attracted attention in recent years, see for example Cossette et al. (2013). However, most of the literature so far focuses on specific margins and dependence structures, such as mixtures of multivariate skew normal distributions in Bernardi (2013), elliptical distributions in Landsman and Valdez (2003) and Dhaene et

¹Under the given continuity assumptions on Z the Expected-Shortfall is equal to the closely related Tail-Value-at-Risk, see Acerbi et al. (2001) and Acerbi and Tasche (2002). It is therefore a coherent risk measure in the sense of Artzner et al. (1999).

al. (2008), or multivariate distributions with mixed Erlang marginals and a Farlie-Gumbel-Morgenstern copula in Cossette et al. (2013). While this comes with the benefit of providing (semi-) closed form expressions for the quantities of interest, it is then not possible to disentangle inference on the margins from that of the dependence structure. As a notable exception, Bargès et al. (2009) provide an algorithm to compute the distribution function and Expected-Shortfall of Z for a general copula and continuous margins that are almost surely positive. As their approach is based on approximating the continuous margins by discrete random variables, a direct formulation for the density is not available. While simulations show the convergence of the approximation in test cases, analytic convergence rates are not quantified.

In this paper, we are concerned with providing a non-random computational framework to find the optimal portfolio with respect to ES_p , without imposing restrictions on the copula or the margins. In the context of portfolio selection based on financial returns, it is of crucial importance not to impose bounds on the random variables in question. The positivity constraint of Bargès et al. (2009) is thus undesirable and needs to be relaxed. To this end, we utilize the approach of Rockafellar and Uryasev (2000) and combine it with the results of Gijbels and Herrmann (2014) in Section 3.

Following Rockafellar and Uryasev (2000), we can compute ES in terms of a convex minimization problem on the loss variable² $Y := Y(\boldsymbol{\omega}) := -Z(\boldsymbol{\omega})$. To compute the portfolio VaR and ES, they define the objective function

$$\mathcal{F}_\beta: \mathcal{X}_r \times \mathbb{R} \rightarrow \mathbb{R}, (\boldsymbol{\omega}, \alpha) \mapsto \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = \alpha + (1 - \beta)^{-1} \int_{\{\mathbf{x} \in \mathbb{R}^d: -\boldsymbol{\omega}^\top \mathbf{x} \geq \alpha\}} (-\boldsymbol{\omega}^\top \mathbf{x} - \alpha) h(\mathbf{x}) d\mathbf{x}.$$

Then, the following theorem can be established.

Theorem 2.1 (Rockafellar and Uryasev (2000), Theorem 1). As a function of α , $\mathcal{F}_\beta(\boldsymbol{\omega}, \alpha)$ is convex and continuously differentiable with $\frac{\partial}{\partial \alpha} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = \frac{F_{Y(\boldsymbol{\omega})}(\alpha) - \beta}{1 - \beta}$. Furthermore:

- (i) For any $\boldsymbol{\omega} \in \mathcal{X}_r$ we can compute $\mathbb{E}[Y(\boldsymbol{\omega}) | Y(\boldsymbol{\omega}) \geq F_{Y(\boldsymbol{\omega})}^{-1}(\beta)] = \min_{\alpha \in \mathbb{R}} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha)$.
- (ii) The set consisting of values of α for which the minimum in (i) is attained, namely $A_\beta(\boldsymbol{\omega}) = \operatorname{argmin}_{\alpha \in \mathbb{R}} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha)$, is a non-empty, closed, bounded interval (perhaps reducing to a single point), and $F_Y^{-1}(\beta)$ is given by $F_{Y(\boldsymbol{\omega})}^{-1}(\beta)$, the left endpoint of $A_\beta(\boldsymbol{\omega})$.

²In Rockafellar and Uryasev (2000) more general loss functions than linear aggregations are considered. We present all following theorems adjusted to our particular setting.

In particular, one always has $F_{Y(\omega)}^{-1}(\beta) \in A_\beta(\omega)$ and $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)] = F_\beta(\omega, F_{Y(\omega)}^{-1}(\beta))$. $\blacksquare$

The importance of Theorem 2.1 is that continuously differentiable functions are particularly good targets for numerical optimization, while the convexity guarantees that any found optimum is equal to the global one. Furthermore, it turns out that the computation of $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)]$ does not depend on a prior computation of $F_{Y(\omega)}^{-1}(\beta)$. Instead, $F_{Y(\omega)}^{-1}(\beta)$ is obtained as a byproduct of the optimization procedure. In terms of portfolio selection, we will utilize the following theorem also due to Rockafellar and Uryasev (2000).

Theorem 2.2 (Rockafellar and Uryasev (2000), Theorem 2). We have:

- (i) Minimizing $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)]$ over all ω in $\mathcal{X}_r$ is equivalent to minimizing $\mathcal{F}_\beta(\omega, \alpha)$ over all $(\omega, \alpha) \in \mathcal{X}_r \times \mathbb{R}$, in the sense that $\min_{\omega \in \mathcal{X}_r} \mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)] = \min_{(\omega, \alpha) \in \mathcal{X}_r \times \mathbb{R}} \mathcal{F}_\beta(\omega, \alpha)$, where moreover a pair (ω^*, α^*) achieves the second minimum if and only if ω^* achieves the first minimum and $\alpha^* \in A_\beta(\omega^*)$.
- (ii) If the interval $A_\beta(\omega^*)$ reduces to a single point (as is typical), the minimization of $\mathcal{F}_\beta(\omega, \alpha)$ over $(\omega, \alpha) \in \mathcal{X}_r \times \mathbb{R}$ produces a pair (ω^*, α^*) , not necessarily unique, such that ω^* minimizes $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)]$ and α^* gives the corresponding $F_{Y(\omega^*)}^{-1}(\beta)$.
- (iii) $\mathcal{F}_\beta(\omega, \alpha)$ is convex with respect to (ω, α) , and $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)]$ is convex with respect to ω . If the constraints are such that $\mathcal{X}_r$ is a convex set, the joint minimization is an instance of convex programming. $\blacksquare$

Theorem 2.2 therefore allows to work with the convex and directly accessible function $\mathcal{F}_\beta$, instead of the more complicated $\mathbb{E}[Y(\omega)|Y(\omega) \geq F_{Y(\omega)}^{-1}(\beta)]$, which is defined over the inverse $F_{Y(\omega)}^{-1}$. It remains to find an approximation to the integral term in the formulation of $\mathcal{F}_\beta$. While Rockafellar and Uryasev (2000) advocate an approximation based on random sampling from the joint distribution H of $\mathbf{X}$, we will connect their theoretical framework with the non-random computational methods of Gijbels and Herrmann (2014).

To this end, we reformulate our problem entirely in terms of Z by relying on the identities $F_Y(y) = 1 - F_Z(-y)$, $F_Y^{-1}(\beta) = -F_Z^{-1}(1 - \beta)$ and $\mathbb{E}[Y|Y \geq y] = -\mathbb{E}[Z|Z \leq -y]$, that hold for

continuous F_Y , F_Z and $\beta \in (0, 1)$. In terms of the ES, we then also have

$$\mathbb{E}[Y|Y \geq F_Y^{-1}(\beta)] = -\mathbb{E}[Z|Z \leq F_Z^{-1}(1 - \beta)].$$

For optimal weights $\boldsymbol{\omega}^*$, we then immediately get

$$\mathbb{E}[Y^*|Y^* \geq F_{Y^*}^{-1}(\beta)] \leq \mathbb{E}[Y(\boldsymbol{\omega})|Y(\boldsymbol{\omega}) \geq F_{Y(\boldsymbol{\omega})}^{-1}(\beta)]$$

for all $\boldsymbol{\omega} \in \mathcal{X}_r$, if and only if $\mathbb{E}[Z^*|Z^* \leq F_{Z^*}^{-1}(1 - \beta)] \geq \mathbb{E}[Z(\boldsymbol{\omega})|Z(\boldsymbol{\omega}) \leq F_{Z(\boldsymbol{\omega})}^{-1}(1 - \beta)]$ for all $\boldsymbol{\omega} \in \mathcal{X}_r$. Optimal weights found via Theorem 2.2 will therefore also serve as optimal weights in ES_p -sense. Finally we can restate the function $\mathcal{F}_\beta$ in terms of Z for our later purpose as

$$\mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = \alpha - \frac{\alpha F_Z(-\alpha) + F_Z(-\alpha)\mathbb{E}[Z|Z \leq -\alpha]}{1 - \beta}, \quad (1)$$

where the corresponding derivative with respect to α is given by $\frac{\partial}{\partial \alpha} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = 1 - \frac{F_Z(-\alpha)}{1 - \beta}$. For a given (typically small) $p \in (0, 1)$, we consequently compute

$$\phi_{1-p}^* = \min_{\boldsymbol{\omega} \in \mathbb{R}^d} \min_{\alpha \in \mathbb{R}} \mathcal{F}_{1-p}(\boldsymbol{\omega}, \alpha), \text{ subject to: (C1), (C2) and (C3)}. \quad (2)$$

Given the optimal decision variables $(\boldsymbol{\omega}^*, \alpha^*)$, we have $\text{ES}_p(Z^*) = -\phi_{1-p}^*$ with a corresponding Value-at-Risk of $\text{VaR}_p(Z^*) = -\alpha^*$. For practical purposes, constraint (C2) can be directly incorporated into the objective function by setting $\omega_d = 1 - \sum_{i=1}^{d-1} \omega_i$ thus reducing the outer minimization by one dimension to $\mathbb{R}^{d-1}$.

3 Properties of Sums of Random Variables

In the previous section, we gave a representation of the portfolio optimization problem in terms of F_Z and $\mathbb{E}[Z|Z \leq z]$. To successfully apply this approach, we now discuss a computational framework that allows to compute F_Z and $\mathbb{E}[Z|Z \leq z]$ under few assumptions on the margins and the dependence structure of $\mathbf{X}$. In order to capture the dependence between the components of $\mathbf{X}$, we resort to the copula framework, see Nelsen (2006).

Definition 3.1. A function $C: [0, 1]^d \rightarrow [0, 1]$ is called d -copula (or copula for short), if C is the distribution function of a d -dimensional random vector $\mathbf{U} = (U_1, \dots, U_d)$ with standard uniform

marginals, i.e. $\mathbb{P}[U_k \leq u_k] = u_k$ for all $k \in \{1, \dots, d\}$ and $u_k \in [0, 1]$. ■

The importance of copulas stems from Sklar's Theorem, which states that any multivariate distribution function can be decomposed in the following way (where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$).

Theorem 3.1 (Sklar (1959)). Let H be a d -dimensional distribution function with margins $F_1, \dots, F_d$. Then, there exists a d -copula C such that for all $\mathbf{x} \in \overline{\mathbb{R}}^d$:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (3)$$

If $F_1, \dots, F_d$ are all continuous, then C is unique. Otherwise, C is uniquely determined on $\text{Ran } F_1 \times \dots \times \text{Ran } F_d$, where $\text{Ran } F_j = F_j(\overline{\mathbb{R}})$ for $j = 1, \dots, d$. Conversely, if C is a d -copula and $F_1, \dots, F_d$ are distribution functions, then the function H defined by (3) is a d -dimensional distribution function with margins $F_1, \dots, F_d$. ■

In what follows, we will always assume that the copulas we consider are everywhere differentiable. For a d -dimensional random vector $\mathbf{X}$, we denote the corresponding unique copula by C and its density by c . With μ_C we denote the measure on $[0, 1]^d$ implied by C , that is, $\mu_C(B) := \mathbb{P}[\mathbf{U} \in B]$ for any Borel measurable $B \subseteq [0, 1]^d$ and $\mathbf{U} \sim C$. Combining (3) with our assumptions for the margins also yields the joint density of $\mathbf{X}$ as $h(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i)$. Under the given conditions, the distribution function of Z is available in terms of an integral of the copula density over a subset of $[0, 1]^d$.

Proposition 3.1 (Embrechts et al. (2005), Gijbels and Herrmann (2014)). Denote by A_z the convex set $A_z = \left\{ \mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^d \omega_i x_i \leq z \right\}$, then

$$F_Z(z) = \mu_C(B_z) = \int_{B_z} c d\boldsymbol{\lambda}, \quad (4)$$

where $B_z = \varphi(A_z) \subseteq [0, 1]^d$ and $\varphi: \mathbb{R}^d \rightarrow [0, 1]^d, (x_1, \dots, x_d) \mapsto (F_1(x_1), \dots, F_d(x_d))$. ■

It is important to notice, that in (4), the influence of the dependence structure, given by the integrand, is separated from that of the marginal distribution functions, which is reflected in the integration area. A detailed description of the integration area B_z is given in Gijbels and

Herrmann (2014). Most importantly, for $j \in \{1, \dots, d\}$, we can define a function

$$\tau_z^j: (0, 1)^{d-1} \rightarrow (0, 1), \mathbf{u} \mapsto \tau_z^j(\mathbf{u}) = F_j \left(\frac{z}{\omega_j} - \sum_{i=1}^{j-1} \frac{\omega_i}{\omega_j} F_i^{-1}(u_i) - \sum_{i=j+1}^d \frac{\omega_i}{\omega_j} F_i^{-1}(u_{i-1}) \right), \quad (5)$$

where we set the last sum to zero if $j = d$, such that $B_z = B_z^j$ is given by

$$B_z^j = \{ \mathbf{u} \in (0, 1)^d : 0 < u_j \leq \tau_z^j(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_d), 0 < u_k < 1 \text{ for } k \neq j \}. \quad (6)$$

Using integration, it is now possible to give the following alternative representation for F_Z and $\mathbb{E}[Z|Z \leq z]$, the quantities of interest in the portfolio optimization problem (2). We specifically keep track of the indices that are singled out, leading to slightly more general formulas than the ones presented in Gijbels and Herrmann (2014).

Theorem 3.2 (Gijbels and Herrmann (2014)). Define Z as above. For a given copula C with density c , denote, for a fixed index $j \in \{1, \dots, d\}$, by C^{-j} the function obtained by omitting the j -th partial derivative of the copula C , i.e. $C^{-j}: (0, 1)^d \rightarrow \mathbb{R}, \mathbf{u} \mapsto \partial_1 \dots \partial_{j-1} \partial_{j+1} \dots \partial_d C(\mathbf{u})$. Then the following statements hold.

- (i) $F_Z(z) = \int_{[0,1]^{d-1}} C^{-j}(u_1, \dots, u_{j-1}, \tau_z^j(u_1, \dots, u_{d-1}), u_j, \dots, u_{d-1}) d\boldsymbol{\lambda}$.
- (ii) Given a function $g \in \mathcal{L}^1$ such that $\left| \frac{d}{dz} C^{-j}(u_1, \dots, u_{j-1}, \tau_z^j(u_1, \dots, u_{d-1}), u_j, \dots, u_{d-1}) \right| \leq g(u_1, \dots, u_{d-1})$ for almost all $\mathbf{u} \in (0, 1)^{d-1}$, the density f_Z of Z is given as

$$f_Z(z) = \int_{[0,1]^{d-1}} c(u_1, \dots, u_{j-1}, \tau_z^j(u_1, \dots, u_{d-1}), u_j, \dots, u_{d-1}) \\ \times f_j \left(\frac{z}{\omega_j} - \sum_{i=1}^{j-1} \frac{\omega_i}{\omega_j} F_i^{-1}(u_i) - \sum_{i=j+1}^d \frac{\omega_i}{\omega_j} F_i^{-1}(u_{i-1}) \right) \frac{1}{\omega_j} d\boldsymbol{\lambda}.$$

- (iii) If $\mathbb{E}[|X_i|] < \infty$ for all $1 \leq i \leq d$ and $z \in \mathbb{R}$ is such that $F_Z(z) > 0$, then

- (a) the conditional expectation $\mathbb{E}[Z|Z \leq z]$ is given by

$$\mathbb{E}[Z|Z \leq z] = \frac{1}{F_Z(z)} \sum_{i=1}^d \omega_i \int_{B_z^{j_i}} F_i^{-1}(u_i) c(u_1, \dots, u_d) du_d \dots du_1,$$

where $j_i \in \{1, \dots, d\}$ and $B_z^{j_i}$ as in (6),

(b) and for $i, j_i \in \{1, \dots, d\}$ such that $i \neq j_i$, we have

$$\begin{aligned} & \int_{B_z^{j_i}} F_i^{-1}(u_i) c(u_1, \dots, u_d) du_d \dots du_1 \\ &= \int_{[0,1]^{d-1}} F_i^{-1}(y_{j_i}^i) C^{-j_i}(y_1, \dots, y_{j_i-1}, \tau_z^{j_i}(\mathbf{y}), y_{j_i}, \dots, y_{d-1}) dy_{d-1} \dots y_1, \end{aligned}$$

where $\mathbf{y} = (u_1, \dots, u_{j_i-1}, u_{j_i+1}, \dots, u_d)$ and $y_{j_i}^i = y_i$, if $i < j_i$ and $y_{j_i}^i = y_{i-1}$ if $i > j_i$. ■

Of practical importance in Theorem 3.2 is that the dimension of the integration domain is transformed from $B_z^j \subseteq [0, 1]^d$ to $[0, 1]^{d-1}$. This not only reduces the dimension of the integration problem, but also provides an easier starting point for numerical integration algorithms as discussed in Section 4.2. Notably, the quantile function $F_Z^{-1}(p)$ for $p \in (0, 1)$ can be computed by numerical root searching algorithms.

While Proposition 3.1 and Theorem 3.2 are obtained under the positivity constraint (C1) for the portfolio weights, it is possible to obtain similar results without imposing (C1). For a fixed choice of j , it is only the sign of ω_j that is of importance. If $\omega_j > 0$, then Proposition 3.1 and Theorem 3.2 can be applied without modifications, even if $\omega_i < 0$ for some $i \neq j$. If $\omega_j < 0$, the integration in Proposition 3.1 and Theorem 3.2 (iii)(a) then needs to be carried out over the alternative integration domain

$$\tilde{B}_z^j = \{\mathbf{u} \in (0, 1)^d : \tau_z^j(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_d) \leq u_j < 1, 0 < u_k < 1 \text{ for } k \neq j\}.$$

This, however, generally excludes the usage of Theorem 3.2 (i), (ii) and (iii)(b) as an evaluation of the function C^{-j} for a boundary point, i.e., $C^{-j}(\dots, 1, \dots)$ is for most copula families not possible. Such an evaluation would however be needed when adapting the proof outlined in Gijbels and Herrmann (2014) to the new scenario. Due to (C2) there is however always at least one index j such that $w_j > 0$. A suitable choice of j therefore generally permits an application of Theorem 3.2 in this context.

Applying Theorem 3.2 requires knowledge of the function C^{-j} for the copula family in question, as well as an efficient and stable method to numerically evaluate the multivariate integral. Both problems are discussed in the following section.

4 Numerical Computation

During the optimization procedure, it is possible that one or more weights will be set to zero. Potentially this may also be true for the optimal weights ω^* . To successfully apply the methodology presented in the previous section, it is necessary that all $\sum_{k=1}^d \binom{d}{k} = 2^d - 1$ lower dimensional distributions associated with $\mathbf{X}$ are known. For example, in the case $d = 3$, we need the distributions of (X_1) , (X_2) , (X_3) , (X_1, X_2) , (X_1, X_3) , (X_2, X_3) and (X_1, X_2, X_3) . If dealing with all lower dimensional joint distributions becomes too cumbersome, either when d is too large or when it is analytically too difficult to trace all cases, a minimum investment constraint $\omega_i \geq c_i > 0$ for all $i \in \{1, \dots, d\}$ can be imposed. This alleviates the issue as the guaranteed positivity of all weights implies that only the joint distribution of $(X_1, \dots, X_d)$ has to be known compared to knowing all $2^d - 1$ possible lower dimensional distributions when zero weights are allowed.

For the examples and real data example considered in Section 5, we do not impose any minimum investment constraint. This is possible since the number of dimensions is limited to $d \leq 5$, keeping the programming effort for all cases still manageable. Furthermore, for some popular copula families, especially those discussed in the following Examples 4.1–4.3 and later utilized in the applications in Section 5, the necessary lower dimensional copulas can be obtained rather easily. For the Gaussian copula, all lower dimensional copulas are Gaussian, with the respective covariance matrix obtained by deleting those rows and columns which are associated with components of $\mathbf{X}$ that receive weight zero. Archimedean copulas have the property, that the lower dimensional distributions are Archimedean of the exact same type. The only adjustment done is in the number of input arguments. In a similar fashion, lower dimensional copulas of the Farlie-Gumbel-Morgenstern copula are again Farlie-Gumbel-Morgenstern copulas. For details and references see Nelsen (2006), p. 108. Only for the vine copulas of Example 4.4 is it not as straightforward to obtain the unconditional lower dimensional copulas as for the previous examples.

4.1 Formulas for C^{-j}

In this subsection, we discuss the necessary formulas for C^{-j} for some important copula classes. We also point out potential problems that complicate the computation of C^{-j} . The Gaussian copula is our first example.

Example 4.1 (Gaussian Copula). We denote by $\Phi_{\mathbf{R}}$ the multivariate normal distribution with standard normal margins and a given correlation matrix $\mathbf{R} \in \mathbb{R}^{d \times d}$ in d dimensions. Then, the Gaussian copula $C_{\mathbf{R}}$ is given by $C_{\mathbf{R}}(u_1, \dots, u_d) := \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$, where Φ^{-1} is the inverse of the univariate standard normal distribution function Φ . To define C^{-j} we further write $y_i = \Phi^{-1}(u_i)$ and $\mathbf{y}_{-j} = (y_1, \dots, \cancel{y_j}, \dots, y_d)^T$. Given the correlation matrix $\mathbf{R}$, we denote by $\mathbf{S}_{-j}$ the $(d-1) \times (d-1)$ matrix obtained by deleting the j -th row and j -th column of the inverse matrix $\mathbf{R}^{-1}$, and set $r_j = \mathbf{R}_{jj}^{-1}$, the j -th diagonal element of $\mathbf{R}^{-1}$. Finally, we define $\mathbf{s}_{-j}$ as a row-vector given by $\mathbf{s}_{-j} = (\mathbf{R}_{j1}^{-1}, \dots, \cancel{\mathbf{R}_{jj}^{-1}}, \dots, \mathbf{R}_{jd}^{-1})$. Then, we have

$$C^{-j}(u_1, \dots, u_d) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}_{-j}^T(\mathbf{S}_{-j} - \mathbf{I})\mathbf{y}_{-j} + \frac{(\mathbf{s}_{-j}\mathbf{y}_{-j})^2}{2r_j}\right)}{\sqrt{r_j}\sqrt{\det \mathbf{R}}} \Phi\left(\sqrt{r_j}\left(y_j + \frac{\mathbf{s}_{-j}\mathbf{y}_{-j}}{r_j}\right)\right),$$

where $\mathbf{I}$ denotes the $(d-1) \times (d-1)$ identity matrix.

To see this, we proceed as follows. First note that $r_j > 0$, which follows by symmetric positive definiteness of $\mathbf{R}^{-1}$. For the joint normal distribution with zero mean and correlation matrix $\mathbf{R}$, the density is given by

$$\varphi_{\mathbf{R}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}\right)}{(2\pi)^{d/2}\sqrt{\det \mathbf{R}}}, \text{ for } \mathbf{x} \in \mathbb{R}^d.$$

We now integrate out over the variable x_j . For $z \in \mathbb{R}$, denote by $\mathbf{x}_z$ the vector $\mathbf{x}_z = (x_1, \dots, x_{j-1}, z, x_{j+1}, \dots, x_d)^T$. Then, we can split $\mathbf{x}_z^T\mathbf{R}^{-1}\mathbf{x}_z = \mathbf{x}_{-j}^T\mathbf{S}_{-j}\mathbf{x}_{-j} + 2\mathbf{s}_{-j}\mathbf{x}_{-j}z + r_jz^2$, where $\mathbf{x}_{-j}$ is a column-vector given by $\mathbf{x}_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d)^T$. This yields

$$\begin{aligned} \Phi_{\mathbf{R},j}(\mathbf{x}) &= \int_{-\infty}^{x_j} \varphi_{\mathbf{R}}(\mathbf{x}_z) dz \\ &= \frac{1}{(2\pi)^{d/2}\sqrt{\det \mathbf{R}}} \int_{-\infty}^{x_j} \exp\left(-\frac{1}{2}(\mathbf{x}_{-j}^T\mathbf{S}_{-j}\mathbf{x}_{-j} + 2\mathbf{s}_{-j}\mathbf{x}_{-j}z + r_jz^2)\right) dz \\ &= \frac{\exp\left(-\frac{1}{2}\mathbf{x}_{-j}^T\mathbf{S}_{-j}\mathbf{x}_{-j}\right)}{(2\pi)^{d/2}\sqrt{\det \mathbf{R}}} \int_{-\infty}^{x_j} \exp\left(-\mathbf{s}_{-j}\mathbf{x}_{-j}z - \frac{1}{2}r_jz^2\right) dz \\ &= \frac{\exp\left(-\frac{1}{2}\mathbf{x}_{-j}^T\mathbf{S}_{-j}\mathbf{x}_{-j}\right)}{(2\pi)^{d/2}\sqrt{\det \mathbf{R}}} \exp\left(\frac{1}{2}\frac{(\mathbf{s}_{-j}\mathbf{x}_{-j})^2}{r_j}\right) \sqrt{\frac{2\pi}{r_j}} \Phi\left(\sqrt{r_j}\left(x_j + \frac{\mathbf{s}_{-j}\mathbf{x}_{-j}}{r_j}\right)\right), \end{aligned}$$

where the last line follows by completing the square in the exponential integral. For the Gaussian

copula, we then have by the chain rule

$$\begin{aligned}
C_{\mathbf{R}}^{-j}(\mathbf{u}) &= \Phi_{\mathbf{R},j}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)) \prod_{\substack{i=1 \\ i \neq j}}^d \sqrt{2\pi} \exp\left(\frac{1}{2}\Phi^{-1}(u_i)^2\right) \\
&= \frac{\exp\left(-\frac{1}{2}\mathbf{y}_{-j}^T \mathbf{S}_{-j} \mathbf{y}_{-j}\right)}{(2\pi)^{d/2} \sqrt{\det \mathbf{R}}} \exp\left(\frac{1}{2} \frac{(\mathbf{s}_{-j} \mathbf{y}_{-j})^2}{r_j}\right) \sqrt{\frac{2\pi}{r_j}} \Phi\left(\sqrt{r_j} \left(\Phi^{-1}(u_j) + \frac{\mathbf{s}_{-j} \mathbf{y}_{-j}}{r_j}\right)\right) \\
&\quad \times \prod_{\substack{i=1 \\ i \neq j}}^d \sqrt{2\pi} \exp\left(\frac{1}{2}\Phi^{-1}(u_i)^2\right) \\
&= \frac{\exp\left(-\frac{1}{2}\mathbf{y}_{-j}^T \mathbf{S}_{-j} \mathbf{y}_{-j}\right)}{\sqrt{r_j} \sqrt{\det \mathbf{R}}} \exp\left(\frac{1}{2} \frac{(\mathbf{s}_{-j} \mathbf{y}_{-j})^2}{r_j}\right) \Phi\left(\sqrt{r_j} \left(\Phi^{-1}(u_j) + \frac{\mathbf{s}_{-j} \mathbf{y}_{-j}}{r_j}\right)\right) \\
&\quad \times \exp\left(\frac{1}{2} \sum_{\substack{i=1 \\ i \neq j}}^d \Phi^{-1}(u_i)^2\right) \\
&= \frac{\exp\left(-\frac{1}{2}\mathbf{y}_{-j}^T (\mathbf{S}_{-j} - \mathbf{I}) \mathbf{y}_{-j} + \frac{(\mathbf{s}_{-j} \mathbf{y}_{-j})^2}{2r_j}\right)}{\sqrt{r_j} \sqrt{\det \mathbf{R}}} \Phi\left(\sqrt{r_j} \left(\Phi^{-1}(u_j) + \frac{\mathbf{s}_{-j} \mathbf{y}_{-j}}{r_j}\right)\right). \quad \blacksquare
\end{aligned}$$

Aside from the Gaussian copula, the (nested) Archimedean copulas are a popular copula family that admits a closed-form representation for C^{-j} in terms of the corresponding generator.

Example 4.2 ((Nested) Archimedean Copulas). A multivariate copula is called Archimedean if it admits a representation $C(u_1, \dots, u_d) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d))$, where $\psi: [0, \infty] \rightarrow [0, 1]$ is a continuous and decreasing function with $\psi(0) = 1$ and $\lim_{t \rightarrow \infty} \psi(t) = 0$, which is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$. The function ψ is called the generator of C and naturally possesses an inverse ψ^{-1} . The function C^{-j} is then given in terms of ψ and ψ^{-1} as

$$C^{-j}(u_1, \dots, u_d) = \psi^{(d-1)}\left(\sum_{i=1}^d \psi^{-1}(u_i)\right) \prod_{\substack{i=1 \\ i \neq j}}^d (\psi^{-1})'(u_i),$$

where $\psi^{(k)}$ denotes the k -th derivative of ψ . For popular Archimedean copula families, explicit formulas for the necessary derivatives of the generator are available in Hofert et al. (2012). Although formulas become more cumbersome, the nested Archimedean copula framework (see Hofert (2010) for an introduction) still allows for a representation of C^{-j} in terms of the involved generators and their inverses. Expressions of densities of nested Archimedean copulas can be found in Hofert and Pham (2013). ■

As a prominent example of a non-elliptical, non-Archimedean copula family, we discuss the Farlie-Gumbel-Morgenstern copula next.

Example 4.3 (Farlie-Gumbel-Morgenstern (FGM) Copula). In the case of the multivariate FGM copula, see Nelsen (2006) p. 108, the copula density is given as

$$c(u_1, \dots, u_d) = 1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1, j_2, \dots, j_k} (1 - 2u_{j_1})(1 - 2u_{j_2}) \cdots (1 - 2u_{j_k}). \quad (7)$$

To guarantee a non-negative density, the $(2^d - d - 1)$ parameters need to fulfil the 2^d constraints

$$1 + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \varepsilon_{j_1} \varepsilon_{j_2} \cdots \varepsilon_{j_k} \theta_{j_1, j_2, \dots, j_k} \geq 0, \text{ for all } \varepsilon_{j_1}, \varepsilon_{j_2}, \dots, \varepsilon_{j_d} \in \{-1, +1\}.$$

To find C^{-j} , we integrate the density with respect to u_j . Integrating the summands in (7) yields

$$I(j_1, j_2, \dots, j_k) = \begin{cases} u_j \prod_{\ell=1}^k (1 - 2u_{j_\ell}), & \text{if } \nexists \ell^* \in \{1, \dots, k\} : j_{\ell^*} = j, \\ u_j (1 - u_j) \prod_{\substack{\ell=1 \\ \ell \neq \ell^*}}^k (1 - 2u_{j_\ell}), & \text{if } \exists \ell^* \in \{1, \dots, k\} : j_{\ell^*} = j. \end{cases}$$

The formula for C^{-j} is then given by

$$C^{-j}(u_1, \dots, u_d) = u_j + \sum_{k=2}^d \sum_{1 \leq j_1 < \dots < j_k \leq d} \theta_{j_1, j_2, \dots, j_k} I(j_1, j_2, \dots, j_k). \quad \blacksquare$$

While the previous examples allow for a straightforward application of the methodology presented in Theorem 3.2, not all popular copula families are equally convenient to deal with. For the class of vine copulas, in a setting with d as low as three, it might be necessary to revert to utilizing Proposition 3.1 and Theorem 3.2 (iii)(a), integrating over B_z^j instead of using C^{-j} .

Example 4.4 (Vine Copula Constructions). Vine copulas, see Czado (2010) for an introduction, utilize the idea of building multivariate distributions from bivariate building blocks by recursive conditioning of the densities. For our purpose a three dimensional example suffices, but vine copulas are available in arbitrary finite dimensions. Taking a three-dimensional random vector $\mathbf{U} = (U_1, U_2, U_3)$ with uniform margins, we can decompose the density c into, for example,

$$c(u_1, u_2, u_3) = c_{1,3|2} \left(C_{1|2}(u_1|u_2), C_{3|2}(u_3|u_2) | u_2 \right) c_{2,3}(u_2, u_3) c_{1,2}(u_1, u_2),$$

where $c_{1,2}$ and $c_{2,3}$ denote the copulas of (U_1, U_2) and (U_2, U_3) , and $c_{1,3|2}$ denotes the copula of (U_1, U_3) conditional on $U_2 = u_2$. Concerning the marginal conditional distributions $C_{1|2}$ and $C_{3|2}$, we can use $C_{1|2}(u_1|u_2) := \mathbb{P}(U_1 \leq u_1|U_2 = u_2) = \frac{\partial}{\partial u_2} C_{1,2}(u_1, u_2)$, leading finally to

$$c(u_1, u_2, u_3) = c_{1,3|2} \left(\frac{\partial}{\partial u_2} C_{1,2}(u_1, u_2), \frac{\partial}{\partial u_2} C_{2,3}(u_2, u_3)|u_2 \right) c_{2,3}(u_2, u_3) c_{1,2}(u_1, u_2).$$

Once the copulas $C_{1,2}$, $C_{2,3}$ and $C_{1,3|2}$ are fixed, we can compute C^{-j} by integration. With our conditioning order we get

$$C^{-3}(u_1, u_2, u_3) = c_{1,2}(u_1, u_2) \int_0^{u_3} c_{1,3|2} \left(\frac{\partial}{\partial u_2} C_{1,2}(u_1, u_2), \frac{\partial}{\partial u_2} C_{2,3}(u_2, x)|u_2 \right) c_{2,3}(u_2, x) dx,$$

showing that the computation of C^{-j} may already be infeasible in three dimensions for some copula choices. Even if we employ the commonly-used simplifying assumption, i.e. $c_{1,3|2}(\cdot, \cdot|u_2) = c_{1,3|2}(\cdot, \cdot)$, the integration problem does not in general possess a closed form solution. ■

Having dealt with the integrands in Theorem 3.2, we turn to the numerical evaluation of the integrals in the following subsection.

4.2 Numerical Integration

Dealing with the minimization problem in (2) is only feasible if the corresponding integrals in Theorem 3.2 can be computed efficiently. While univariate numerical integration is a standard topic in numerical analysis, see for example Stoer and Bulirsch (1993), the multivariate counterpart is more diverse in nature. Shifting away from the suggested Monte-Carlo method of Rockafellar and Uryasev (2000) and the Quasi-Monte-Carlo method utilized in Gijbels and Herrmann (2014), we opt here for numerical integration using sparse grids, also called Smolyak quadrature referring to their first introduction by Smolyak (1963). Sparse grids, see Gerstner and Griebel (1998) for an overview, utilize univariate quadrature formulas as building blocks for multivariate numerical integration, which we discuss without loss of generality for functions defined on $[-1, 1]^d$. Typical for numerical integration, sparse grid approaches provide node points $(\mathbf{x}_i)_{i=1}^N$, with $\mathbf{x}_i \in [-1, 1]^d$, and corresponding weights $(\alpha_i)_{i=1}^N$, with $\alpha_i \in \mathbb{R}$, such that the integral of a function $f: [-1, 1]^d \rightarrow \mathbb{R}$ can be approximated by a weighted sum of function evaluations $\int_{[-1, 1]^d} f(\mathbf{x}) d\mathbf{x} \approx \sum_{i=1}^N \alpha_i f(\mathbf{x}_i)$. We denote by $e(N)$ the approximation error

$e(N) = \left| \int_{[-1,1]^d} f(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^N \alpha_i f(\mathbf{x}_i) \right|$. We use the big- $\mathcal{O}$ notation

$f(x) = \mathcal{O}(g(x))$ if and only if $\exists M > 0 : |f(x)/g(x)| \leq M$ for x sufficiently large

to quantify convergence rates of $e(N)$ to zero.

To construct the multivariate Smolyak rule we first consider a univariate quadrature rule Q acting on a function $g: [-1, 1] \rightarrow \mathbb{R}$ via $Q[g] = \sum_{j=1}^{n(Q)} \alpha_j(Q) g(x_j(Q))$, where we collect the nodes and weights in the sets $\Gamma(Q) = \{x_j(Q), 1 \leq j \leq n(Q)\}$ and $W(Q) = \{\alpha_j(Q), 1 \leq j \leq n(Q)\}$ for $n(Q) \in \mathbb{N}_1$, where $\mathbb{N}_k = \{k, k+1, \dots\}$ for an integer k . We will only consider quadrature rules that belong to families that can be indexed over the natural numbers. For example, in the case of the closed trapezoidal rule T_n over $[-1, 1]$ with n node points, $n \geq 2$, we have $n(T_n) = n$, and $\Gamma(T_n) = \left\{ -1 + \frac{2(j-1)}{n-1}, 1 \leq j \leq n \right\}$, $W(T_n) = \left\{ \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{2}{n-1}, \frac{1}{n-1} \right\}$ and $\#W(T_n) = \#\Gamma(T_n)$.

A sequence of quadrature rules $(Q_\ell)_{\ell \geq 1}$ is called nested if $\Gamma(Q_\ell) \subsetneq \Gamma(Q_{\ell+1})$ for all $\ell \geq 1$, and for Q_1 it holds that $n(Q_1) = 1$, $\Gamma(Q_1) = \{0\}$ and $W(Q_1) = \{2\}$. For an indexed family of quadrature rules we can typically pick a subsequence $(Q_\ell)_{\ell \geq 1}$ that will be nested. The index ℓ is called the level, and we always refer to a nested sequence whenever the index ℓ is used. In the case of the trapezoidal rule, we can for example set $Q_\ell = T_{n_\ell}$ with $n_\ell = 2^{\ell-1} + 1$ for $\ell \geq 2$ and Q_1 as above to obtain a nested sequence.

For two quadrature rules, we define their difference as $\Delta(Q_2, Q_1) = Q_2 - Q_1$, meaning $\Delta(Q_2, Q_1)[g] = Q_2[g] - Q_1[g]$ for a suitable function g . For elements of a sequence, we use the shorthand notation $\Delta_n = \Delta(Q_n, Q_{n-1})$. Clearly, the difference of quadrature rules yields again a quadrature rule. When applying the differencing to a nested sequence, we directly find $n(\Delta_\ell) = n(Q_\ell)$ and $\Gamma(\Delta_\ell) = \Gamma(Q_\ell)$. In this case, the differencing is therefore only a re-weighting of the node points in Q_ℓ . Making the step to multivariate integration rules requires us to define the tensor product $Q_1 \otimes \dots \otimes Q_d$ of quadrature rules, which is given as

$$Q_1 \otimes \dots \otimes Q_d[f] = \sum_{i_1=1}^{n(Q_1)} \dots \sum_{i_d=1}^{n(Q_d)} \alpha_{i_1}(Q_1) \dots \alpha_{i_d}(Q_d) f(x_{i_1}(Q_1), \dots, x_{i_d}(Q_d)),$$

for a function $f: [-1, 1]^d \rightarrow \mathbb{R}$. For a nested sequence of quadrature rules $(Q_\ell)_{\ell \geq 1}$ the d -

dimensional Smolyak quadrature rule Q_ℓ^d with level ℓ is now given as

$$Q_\ell^d[f] = \sum_{|\mathbf{k}|_1 \leq \ell+d-1} \Delta_{k_1} \otimes \cdots \otimes \Delta_{k_d}[f], \quad (8)$$

where $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}_1^d$ denotes a multi-index, and $|\mathbf{k}|_1 := \sum_{i=1}^d k_i$. Although (8) can be implemented directly, algorithms to compute $\Gamma(Q_\ell^d)$ and $W(Q_\ell^d)$ in an efficient and numerically stable way are available, see Gerstner and Griebel (1998). The combination of univariate quadrature rules according to (8) has some favourable properties that depend on the univariate building blocks $(Q_\ell)_{\ell \geq 1}$. When Q_ℓ satisfies $n(Q_\ell) = \mathcal{O}(2^\ell)$ for ℓ sufficiently large, we have $n(Q_\ell^d) = \mathcal{O}(2^\ell \ell^{d-1})$. Comparing this to the product rule node growth of $\mathcal{O}(2^{\ell d})$ justifies the sparse grid terminology.

To discuss convergence rates of the approximation error $e(N)$, we consider two function spaces. First, $\mathcal{C}^l$ denotes the space of univariate functions with finite derivatives up to order l , given as

$$\mathcal{C}^l := \left\{ g: [-1, 1] \rightarrow \mathbb{R} : \sup_{x \in (-1, 1)} \left| \frac{d^s g(x)}{dx^s} \right| < \infty, s \leq l \right\}.$$

Second, $\mathcal{W}_d^k$ denotes the space of d -variate functions with existing and bounded mixed partial derivatives

$$\mathcal{W}_d^k := \left\{ f: [-1, 1]^d \rightarrow \mathbb{R} : \left\| \frac{\partial^{|\mathbf{s}|_1} f}{\partial x_1^{s_1} \cdots \partial x_d^{s_d}} \right\|_\infty < \infty, \|\mathbf{s}\|_\infty \leq k \right\},$$

where $\|\mathbf{s}\|_\infty := \max_{1 \leq i \leq d} s_i$ for a multi-index $\mathbf{s} \in \mathbb{N}_0^d$. For a sequence of univariate quadrature rules $(Q_\ell)_{\ell \geq 1}$ with approximation error

$$\left| \int_{[-1, 1]} g(x) dx - Q_\ell[g] \right| = e(n(Q_\ell)) = \mathcal{O}((n(Q_\ell))^{-k}) \quad (9)$$

for $g \in \mathcal{C}^k$, we have an approximation error of the multivariate integral as

$$\left| \int_{[-1, 1]^d} f(\mathbf{x}) d\mathbf{x} - Q_\ell^d[f] \right| = e(n(Q_\ell^d)) = \mathcal{O}(2^{-\ell k} \ell^{(d-1)(k+1)})$$

for $f \in \mathcal{W}_d^k$; see e.g. Gerstner and Griebel (1998). The requirement given in (9) is not restrictive,

	Product Rule	Smolyak Quadrature with $N = n(Q_\ell^d)$	Monte Carlo	Quasi Monte Carlo
$e(N)$	$\mathcal{O}\left(\frac{1}{N^{k/d}}\right)$	$\mathcal{O}\left(\frac{(\log N)^{(d-1)(k+1)}}{N^k}\right)$	$\mathcal{O}\left(\frac{1}{N^{1/2}}\right)$	$\mathcal{O}\left(\frac{(\log N)^d}{N}\right)$
Nr. node points	$\mathcal{O}(2^{\ell d})$	$\mathcal{O}(2^\ell \ell^{d-1})$	-	-

Table 1: Asymptotic convergence rates of $e(N)$ for different numerical integration schemes for a function $f \in \mathcal{W}_d^k$. Results can amongst others be found in Davis and Rabinowitz (1984) (Product Rule, Monte Carlo), Gerstner and Griebel (1998) or Kaarnioja (2013) (Smolyak Quadrature) and Niederreiter (1992) (Quasi Monte Carlo). For the product rule and Smolyak quadrature the second line gives the asymptotic growth of the number of node points when both methods are based on a univariate integration method with $\mathcal{O}(2^\ell)$ node points.

as e.g. the Gauss-Legendre, Gauss-Patterson and Clenshaw-Curtis rules fulfil it.

The convergence rate of the approximation error can also be expressed in terms of the number of node points associated with Q_ℓ^d to allow a comparison to other multivariate integration schemes, see Table 1. Kaarnioja (2013) compares the Smolyak quadrature convergence rate in terms of node points to the product rule, Monte Carlo and Quasi Monte Carlo integration schemes. Table 1 shows that the regularity k of the univariate function g in (9) determines the asymptotic convergence speed relative to the competitors. For sufficiently regular functions the Smolyak quadrature is therefore, at least in moderate dimensions, a first choice. In the case of our examples and real data application in Sections 5.1 and 5.2, the corresponding marginal distributions and copulas are highly regular and we therefore utilize the Smolyak quadrature for our numerical computations.

When applying numerical integration methods to compute the Expected-Shortfall via Theorem 3.2, using (iii)(b) is clearly preferable to a direct computation over $B_z^{j_i}$ in (iii)(a) as it reduces the dimensionality of the problem and thus leads to faster computation times. In Table 7 in Section 5.1, we compare execution times of the two approaches. However, purely numerically the approach of using Theorem 3.2 (iii)(b) can lead to difficulties for certain constellations of weights. The numerical problem arises when the input argument to F_j in (5) is so big or small that τ_z^j gets numerically evaluated to either zero or one. This in turn poses problems when computing $C^{-j_i}(y_1, \dots, y_{j_i-1}, \tau_z^{j_i}(\mathbf{y}), y_{j_i}, \dots, y_{d-1})$ in the integral in Theorem 3.2 (i) and (iii)(b), as the evaluation of C^{-j_i} on the boundary might not be defined for every copula in d dimensions.

To better understand the problem, we denote by $\omega_{(k)}$ the k -th largest value in $\{\omega_1, \dots, \omega_d\}$, e.g. $\omega_{(1)} = \max_{1 \leq k \leq d} \omega_k$ and $\omega_{(d)} = \min_{1 \leq k \leq d} \omega_k$. Notice that the indices of the associated ω_i are not necessarily unique due to possible ties in the weights. If we choose for example j_i

such that $\omega_{j_i} = \omega_{(d)}$, all terms in $\frac{z}{\omega_{j_i}} - \sum_{k=1}^{j_i-1} \frac{\omega_k}{\omega_{j_i}} F_k^{-1}(u_k) - \sum_{k=j_i+1}^d \frac{\omega_k}{\omega_{j_i}} F_k^{-1}(u_{k-1})$ get amplified, leading potentially to the numerical breakdown outlined above. On the other hand, choosing j_i such that $\omega_{j_i} = \omega_{(1)}$ will shrink all ω_k/ω_{j_i} terms and leads to an improved stability. Fixing therefore an index $j^* \in \{1, \dots, d\}$ such that $\omega_{j^*} = \omega_{(1)}$ allows to apply this shrinking strategy for the $d - 1$ cases when $i \neq j^*$, see Theorem 3.2 (iii)(b). In the case $i = j^*$ it is advised to pick j_i such that $\omega_{j_i} = \omega_{(2)}$ to mitigate the blow up effect. Depending on the ratio $\omega_{(1)}/\omega_{(2)}$, numerical problems might still occur. For this reason we make use of Theorem 3.2 (iii)(b) if $\omega_{(1)}/\omega_{(2)} \leq K$, where a value of $K = 1.5$ is applied in our examples. The value of K has to be chosen with the target distributions, their numerical implementation and chosen floating point representations in mind. Considering for example the Laplace, logistic, normal and Student-t distributions it is clear that, from a mathematical point of view, their distribution functions are never zero, regardless of the argument. Due to the limitations of floating point arithmetic, this is however not true for numerical implementations. While we exemplify this issue using the Boost C++ library, it needs to be understood that this is an unavoidable issue in any programming language, and not a shortcoming of this specific implementation. We denote by $>^{\text{float}}, =^{\text{float}}$ and $>^{\text{double}}, =^{\text{double}}$ the comparison operators for numbers in float and double precision. Considering the implementation of the standard normal distribution in float precision we have $\Phi(-14.14)^{\text{float}} > 0$, yet $\Phi(-14.15)^{\text{float}} = 0$. In double precision we have $\Phi(-38.47)^{\text{double}} > 0$, yet $\Phi(-38.48)^{\text{double}} = 0$. For other distributions the numerical difficulties arise around different points: The standard Laplace distribution returns zero at -102.88 (float) and -744.04 (double). The standard logistic distribution returns zero at -88.01 (float) and -709.01 (double). The heavy tailed Student-t distribution with 4 degrees-of-freedom can be evaluated for float and double values smaller than -10000 without returning zero. This confirms the intuition that for heavy tailed distributions, this problem is less of an issue, due to the slow decay of the distribution function towards zero. Using double precision in our numerical experiments, a value of $K = 1.5$ was sufficient to prevent numerical problems. In other settings, the value can be adjusted as needed.

When the ratio $\omega_{(1)}/\omega_{(2)}$ overshoots K , we apply Theorem 3.2 (iii)(a) directly, trading a longer execution time for numerical stability. It is important to notice that K is not a tuning parameter of the underlying statistical model, and that it does not depend on the considered data. In Table 7, we report the percentage of overshoots over this a priori chosen threshold

compared to the total number of function evaluations for one of the examples in Section 5.

5 Applications

5.1 Illustrative Examples

To illustrate the impact of dependence structures and margins on the portfolio selection problem, we compute $\mathbb{E}[Y|Y \geq F_Y^{-1}(\beta)]$ for different scenarios using the computational set up outlined in Sections 3 and 4. We therefore consider eight random vectors $\{\mathbf{X}_i\}_{i=1}^8$ in three dimensions that differ in terms of their copula and margins. To allow for a fair comparison, we however select the margins such that they have component-wise the same mean and variance over all $\mathbf{X}_i$.

We further make the pairwise dependence structures comparable in terms of Kendall's tau. For two random variables X_k and X_l with copula C_{X_k, X_l} , Kendall's tau measure of association can be expressed as $\tau_K(X_k, X_l) = 4 \int_{[0,1]^2} C_{X_k, X_l}(u, v) dC_{X_k, X_l}(u, v) - 1$, see Nelsen (2006) for details. We therefore choose the copula parameter(s) such that all pairwise dependences are the same in this sense over all $\mathbf{X}_i$.

To break ties with symmetrical marginal distributions, we utilize the skew or type I generalized logistic distribution; see Johnson et al. (1995). Additional properties, maximum likelihood estimation and limiting behaviour are discussed in Shao (2002). The skew logistic distribution ($\mathcal{SL}$) is constructed from the logistic distribution by distorting its distribution function with a power function. Denote by $\mathcal{L}_{m,s}$ the logistic distribution with location parameter $m \in \mathbb{R}$ and scale parameter $s > 0$. For $\xi > 0$, we denote by V a random variable that follows a skew logistic distribution $\mathcal{SL}(z; m, s, \xi) = (\mathcal{L}_{m,s}(z))^\xi = (1 + \exp(-\frac{z-m}{s}))^{-\xi}$. The corresponding quantile function is given as $\mathcal{SL}^{-1}(p) = \mathcal{L}_{m,s}^{-1}(p^{1/\xi}) = m + s \ln(p^{1/\xi}/(1 - p^{1/\xi}))$. We then have the first three centred moments, see Windberger and Zeileis (2014), as $\mathbb{E}[V] = m + s(\psi_{\text{DG}}(\xi) - \psi_{\text{DG}}(1))$, $\mathbb{E}[(V - \mathbb{E}[V])^2] = s^2(\psi'_{\text{DG}}(\xi) + \psi'_{\text{DG}}(1))$ and $\mathbb{S}[V] = \frac{\mathbb{E}[(V - \mathbb{E}[V])^3]}{\mathbb{E}[(V - \mathbb{E}[V])^2]^{3/2}} = \frac{\psi''_{\text{DG}}(\xi) - \psi''_{\text{DG}}(1)}{(\psi'_{\text{DG}}(\xi) + \psi'_{\text{DG}}(1))^{3/2}}$, where $\psi_{\text{DG}}(x) = \frac{d}{dx} \log(\Gamma(x))$ denotes the digamma function. To be prepared for the case $\omega_i = 1$, $i = 1, 2, 3$, we additionally compute $\mathbb{E}[V|V \leq z]$ as

$$\mathbb{E}[V|V \leq z] = m + s \left(\log \left(\frac{\mathcal{L}_{m,s}(z)}{1 - \mathcal{L}_{m,s}(z)} \right) - \frac{1}{\xi} - \mathcal{L}_{m,s}(z) \mathcal{L}(\mathcal{L}_{m,s}(z), 1, \xi + 1) \right), \quad (10)$$

where $\mathcal{L}(z, s, \xi) = \sum_{n=0}^{\infty} \frac{z^n}{(n+\xi)^s}$ is the Lerch transcendent. The result is derived in Appendix A, while an efficient computation of $\mathcal{L}$ is discussed in Jentschura et al. (1999) and Aksenov et al.

(2003). Depending on available software, the following link between the Gauss hypergeometric series ${}_2F_1$ and $\mathcal{L}$, given as ${}_2F_1(1, \xi; 1 + \xi; z) = {}_2F_1(\xi, 1; 1 + \xi; z) = \xi \mathcal{L}(z, 1, \xi)$ can be exploited.

To showcase the effect of semi-heavy and heavy tails we also use the Laplace and Student-t distributions as marginal distributions. The necessary conditional expectations comparable to (10) are derived in Appendices B and C for location scale models build on both distributions. See also Table 10.

In our study, we start from the familiar jointly Gaussian setting and then replace the copula and / or margins to assess the effect of the individual building blocks. The details for each case can be found in Table 2. With the parameters fixed as in Table 2, we have $\tau_K(X_{i,k}, X_{i,l}) \approx 0.297$ for $1 \leq i \leq 8$ and $1 \leq k < l \leq 3$. To obtain this result, we use that Kendall's tau can be computed in closed form for the respective copulas, see Table 3 for details.

Figures 1 and 2 show the resulting contour plots of $\mathbb{E}[Y|Y \geq F_Y^{-1}(0.95)]$ with $Y = -\boldsymbol{\omega}^T \mathbf{X}_i$ for the different settings $i = 1, \dots, 8$ in Table 2 over the (ω_1, ω_2) space. Here, we use the fact that ω_3 is given implicitly. For the Smolyak quadrature we use nested Gaussian quadrature and level $\ell = 7$ for the examples. On top of the contour lines the minimum return constraint (C3) has been added for a minimum return of $r = 0.48$, showing how the constraint impacts the feasible (ω_1, ω_2) set. In the figures, black dots represent the weights of the global 5% minimum Expected-Shortfall portfolio (MESP), while circles show optimal weights for different active minimum return constraints. When numerically carrying out the optimization of $\mathcal{F}_\beta$ with respect to $\boldsymbol{\omega}$, one can opt for gradient-based or derivative-free methods. While we utilize a derivative-free method in our computations, we derive all necessary formulas to compute $\partial/\partial\omega_k \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha)$ in Appendix D, enabling the potential usage of gradient-based methods as well.

In Figure 1, we explore the impact of changing the marginal distributions, while keeping the copula fixed to be a Gaussian one. Compared to the jointly normal setting (top left), adding skewed margins (top right) rotates the orientation of the elliptically shaped contours lines. From Table 4, we see that this change in shape translates into a new set of optimal weights. At the same time, the overall level of the surface increases. For the global 5% MESP, this leads to a lower expectation, as well as a lower Value-at-Risk and Expected-Shortfall. Taken together, this indicates a substantially riskier behaviour compared to the jointly normal case.

Adding heavy tails (bottom left and right) again impacts the shape and level of the objective

Vector	Copula	Margins	Skewness
$\mathbf{X}_1$	Gaussian, $\mathbf{R} = \begin{pmatrix} 1 & 0.45 & 0.45 \\ 0.45 & 1 & 0.45 \\ 0.45 & 0.45 & 1 \end{pmatrix}$	$\mathcal{N}(0.55, 1.2^2)$ $\mathcal{N}(0.50, 1.0^2)$ $\mathcal{N}(0.45, 0.8^2)$	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$
$\mathbf{X}_2$	Gaussian, $\mathbf{R}$ as above	$\mathcal{SL}(1.559104, 0.2764488, 0.25)$ $\mathcal{SL}(1.040445, 0.3898484, 0.50)$ $\mathcal{SL}(0.6488672, 0.3909743, 0.75)$	$\mathbb{S}[X_1] = -1.55$ $\mathbb{S}[X_2] = -0.86$ $\mathbb{S}[X_3] = -0.34$
$\mathbf{X}_3$	Gaussian, $\mathbf{R}$ as above	Laplace(0.55, 0.8485281) Laplace(0.50, 0.7071068) Laplace(0.45, 0.5656854)	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$
$\mathbf{X}_4$	Gaussian, $\mathbf{R}$ as above	$t_4(0.55, 0.8485281)$ $t_4(0.50, 0.7071068)$ $t_4(0.45, 0.5656854)$	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$
$\mathbf{X}_5$	Ali-Mikhail-Haq, $\theta = 0.9376$	$\mathcal{N}(0.55, 1.2^2)$ $\mathcal{N}(0.50, 1.0^2)$ $\mathcal{N}(0.45, 0.8^2)$	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$
$\mathbf{X}_6$	Ali-Mikhail-Haq, $\theta = 0.9376$	$\mathcal{SL}(1.559104, 0.2764488, 0.25)$ $\mathcal{SL}(1.040445, 0.3898484, 0.50)$ $\mathcal{SL}(0.6488672, 0.3909743, 0.75)$	$\mathbb{S}[X_1] = -1.55$ $\mathbb{S}[X_2] = -0.86$ $\mathbb{S}[X_3] = -0.34$
$\mathbf{X}_7$	Frank, $\theta = 2.885244$	Laplace(0.55, 0.8485281) Laplace(0.50, 0.7071068) Laplace(0.45, 0.5656854)	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$
$\mathbf{X}_8$	Ali-Mikhail-Haq, $\theta = 0.9376$	$t_4(0.55, 0.8485281)$ $t_4(0.50, 0.7071068)$ $t_4(0.45, 0.5656854)$	$\mathbb{S}[X_1] = 0.00$ $\mathbb{S}[X_2] = 0.00$ $\mathbb{S}[X_3] = 0.00$

Table 2: Specification of random vectors used in the numerical illustrations. The marginal parameters are selected such that the first two moments coincide in all cases. The copula parameters are selected such that Kendall's tau is equal to $\tau_K = 0.297$ for all pairs.

2-Copula	Kendall's tau
Gaussian ($\rho \in [-1, 1]$)	$\tau_K(\rho) = \frac{2}{\pi} \arcsin(\rho)$
Frank ($\theta \in \mathbb{R}$)	$\tau_K(\theta) = 1 - \frac{4}{\theta} \left(1 - \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt\right)$
Ali-Mikhail-Haq ($\theta \in [-1, 1]$)	$\tau_K(\theta) = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2} \ln(1 - \theta)$

Table 3: Kendall's tau for the utilized bivariate copula families. Formulas for the Frank and Ali-Mikhail-Haq copulas can be found in Nelsen (2006). For the Gaussian copula see Fang and Fang (2002) and Lindskog et al. (2003).

functions and therefore the weights of the global 5% MESP. From Table 4 we see that the expectation increases when switching to heavier margins. While the Expected-Shortfall decreases, indicating a higher risk compared to the jointly normal case, the Value-at-Risk increases for both, the Laplace and Student-t margins. This leads to the interesting situation where the chosen risk measures stand in contradiction. Due to the set-up this effect can be attributed solely to the differences in margins, as the dependence structure represented by the copula stays unchanged.

In Figure 2, we keep the same margins as in Figure 1, while changing the underlying copula. This leads to different contour plots compared to their counterparts in Figure 1, but also between the plots in Figure 2. While the skew logistic margins in combination with the Ali-Mikhail-Haq copula lead to the overall lowest Value-at-Risk and Expected-Shortfall for the global 5% MESP, see Table 4, the combination of Laplace margins and a Frank copula results in the overall highest Expected-Shortfall, while Student-t margins and a Gaussian copula lead to the overall highest Value-at-Risk. As in the case of the Gaussian dependence structure, we find combinations for the Ali-Mikhail-Haq copula, namely $\omega^T \mathbf{X}_5$ and $\omega^T \mathbf{X}_8$, where the Expected-Shortfall decreases while the Value-at-Risk increases upon changing marginal information only. Comparing again $\omega^T \mathbf{X}_6$ and $\omega^T \mathbf{X}_7$, we find substantially different optimal weights for the global 5% MESP, as well as the biggest overall difference in the expected portfolio return.

While the weights for the global 5% MESP can differ substantially, a binding minimum return constraint forces the weights to align more for higher minimum returns as shown in Table 5. Once the minimum return constraint is set so high that it effectively excludes the asset with the lowest expected return, the weights are fully determined by the minimum return and budget constraints. This can be seen in Figures 1 and 2 when the restricted optimal weights are placed on the boundary of the feasible region. However, while a minimum return constraint with $r = 0.48$ is binding in most cases, the combination of Laplace margins with a Frank copula (bottom left in Figure 2) still admits the global minimum in this case, see Table 5.

Overall, the examples clearly show that despite matching marginal first and second moments, as well as matching pairwise Kendall's tau, the actual underlying model specification impacts the aggregated quantity Z substantially. As in the familiar Markowitz (1952) model, optimal portfolios usually display diversification over the possible assets. For example, in the case of the global 5% MESP this means that instead of investing 100% in the stock with lowest marginal

Copula	Margin	$Z(\boldsymbol{\omega})$	ω_1	ω_2	ω_3	$\mathbb{E}[Z(\boldsymbol{\omega})]$	$\text{VaR}_{0.05}[Z(\boldsymbol{\omega})]$	$\text{ES}_{0.05}[Z(\boldsymbol{\omega})]$
Gauss.	normal	$\boldsymbol{\omega}^T \mathbf{X}_1$	0.11306	0.27701	0.60993	0.47516	-0.74240	-1.05044
Gauss.	skew log.	$\boldsymbol{\omega}^T \mathbf{X}_2$	0.07751	0.25429	0.66819	0.47047	-0.79849	-1.26871
Gauss.	Laplace	$\boldsymbol{\omega}^T \mathbf{X}_3$	0.14386	0.29294	0.56321	0.47903	-0.72430	-1.18377
Gauss.	Student-t	$\boldsymbol{\omega}^T \mathbf{X}_4$	0.13182	0.28430	0.58388	0.47740	-0.66487	-1.15889
AMH	normal	$\boldsymbol{\omega}^T \mathbf{X}_5$	0.06006	0.23266	0.70727	0.46764	-0.82062	-1.13386
AMH	skew log.	$\boldsymbol{\omega}^T \mathbf{X}_6$	0.01610	0.20480	0.77910	0.46185	-0.87616	-1.36797
Frank	Laplace	$\boldsymbol{\omega}^T \mathbf{X}_7$	0.19483	0.30199	0.50317	0.48458	-0.69303	-1.04304
AMH	Student-t	$\boldsymbol{\omega}^T \mathbf{X}_8$	0.07599	0.24946	0.67455	0.47007	-0.73669	-1.26248

Table 4: Properties of $Z(\boldsymbol{\omega}) = \boldsymbol{\omega}^T \mathbf{X}_i$, $i = 1, \dots, 8$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ are the globally optimal minimum Expected-Shortfall portfolio weights at 5%.

Copula	Margin	$Z(\boldsymbol{\omega})$	ω_1	ω_2	ω_3	$\mathbb{E}[Z(\boldsymbol{\omega})]$	$\text{VaR}_{0.05}[Z(\boldsymbol{\omega})]$	$\text{ES}_{0.05}[Z(\boldsymbol{\omega})]$
Gauss.	normal	$\boldsymbol{\omega}^T \mathbf{X}_1$	0.15512	0.28976	0.55512	0.48000	-0.74453	-1.05420
Gauss.	skew log.	$\boldsymbol{\omega}^T \mathbf{X}_2$	0.15472	0.29056	0.55472	0.48000	-0.81199	-1.29349
Gauss.	Laplace	$\boldsymbol{\omega}^T \mathbf{X}_3$	0.15263	0.29474	0.55263	0.48000	-0.72509	-1.18397
Gauss.	Student-t	$\boldsymbol{\omega}^T \mathbf{X}_4$	0.15456	0.29089	0.55455	0.48000	-0.66794	-1.16022
AMH	normal	$\boldsymbol{\omega}^T \mathbf{X}_5$	0.15928	0.28144	0.55928	0.48000	-0.83982	-1.15303
AMH	skew log.	$\boldsymbol{\omega}^T \mathbf{X}_6$	0.15324	0.29352	0.55324	0.48000	-0.93503	-1.43815
Frank	Laplace	$\boldsymbol{\omega}^T \mathbf{X}_7$	0.19483	0.30199	0.50317	0.48458	-0.69303	-1.04304
AMH	Student-t	$\boldsymbol{\omega}^T \mathbf{X}_8$	0.15713	0.28574	0.55713	0.48000	-0.76061	-1.27762

Table 5: Properties of $Z(\boldsymbol{\omega}) = \boldsymbol{\omega}^T \mathbf{X}_i$, $i = 1, \dots, 8$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ are the optimal minimum Expected-Shortfall portfolio weights at 5% under a minimum return constraint of $\mathbb{E}[Z(\boldsymbol{\omega})] \geq 0.48$.

Expected-Shortfall, the dependence between assets will lead to more favourable combinations.

To explore the impact of an increase in the number of assets, we consider a setting similar to $\mathbf{X}_5$ with an Ali-Mikhail-Haq copula and normal margins. In this example, we gradually introduce new assets starting from $d = 2$ and going up to $d = 5$. The marginal distributions of the assets are $X_1 \sim \mathcal{N}(0.550, 1.2^2)$, $X_2 \sim \mathcal{N}(0.525, 1.1^2)$, $X_3 \sim \mathcal{N}(0.500, 1.0^2)$, $X_4 \sim \mathcal{N}(0.475, 0.9^2)$ and $X_5 \sim \mathcal{N}(0.450, 0.8^2)$. We combine these marginal distributions to form the portfolios $Z_2(\boldsymbol{\omega}) = \omega_1 X_1 + \omega_5 X_5$, $Z_3(\boldsymbol{\omega}) = \omega_1 X_1 + \omega_3 X_3 + \omega_5 X_5$, $Z_4(\boldsymbol{\omega}) = \omega_1 X_1 + \omega_3 X_3 + \omega_4 X_4 + \omega_5 X_5$ and $Z_5(\boldsymbol{\omega}) = \omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3 + \omega_4 X_4 + \omega_5 X_5$. The dependence structure between the margins is in all cases given by an Ali-Mikhail-Haq copula of the appropriate dimension with parameter $\theta = 0.9376$. For each dimension, we compute the weights of the global minimum Expected-Shortfall portfolio at 5% and report the optimal weights and portfolio characteristics in Table 6. As seen from the results, the introduction of additional investment opportunities in general lowers the risk associated with the portfolio by opening up new diversification opportunities. We also take this example to analyse the impact of using C^{-j} in the calculations. To this end, we re-

Figure 1: Contour lines for $\mathbb{E}[Y|Y \geq F_Y^{-1}(0.95)]$ for $Y = -\omega^T \mathbf{X}_1$ (top, left), $Y = -\omega^T \mathbf{X}_2$ (top, right), $Y = -\omega^T \mathbf{X}_3$ (bottom, left) and $Y = -\omega^T \mathbf{X}_4$ (bottom, right); see Table 2 for details. The global minimum is marked by the solid black dot, circles represent optimal weights when a minimum return constraint $r \in \{r_1, \dots, r_{20}\}$, where $r_j = 0.455 + (j - 1)0.005$ for $j = 1, \dots, 20$, is active. The solid line indicates the budget constraint for $r_6 = 0.48$.

Figure 2: Contour lines for $\mathbb{E}[Y|Y \geq F_Y^{-1}(0.95)]$ for $Y = -\omega^T \mathbf{X}_5$ (top, left), $Y = -\omega^T \mathbf{X}_6$ (top, right), $Y = -\omega^T \mathbf{X}_7$ (bottom, left) and $Y = -\omega^T \mathbf{X}_8$ (bottom, right); see Table 2 for details. The global minimum is marked by the solid black dot, circles represent optimal weights when a minimum return constraint $r \in \{r_1, \dots, r_{20}\}$, where $r_j = 0.455 + (j - 1)0.005$ for $j = 1, \dots, 20$, is active. The solid line indicates the budget constraint for $r_6 = 0.48$.

	ω_1	ω_2	ω_3	ω_4	ω_5	$\mathbb{E}[Z_i(\omega)]$	$\text{VaR}_{0.05}[Z_i(\omega)]$	$\text{ES}_{0.05}[Z_i(\omega)]$
$Z_2(\omega)$	0.10684	—	—	—	0.89316	0.46068	−0.85480	−1.18390
$Z_3(\omega)$	0.06029	—	0.23469	—	0.70502	0.46776	−0.82011	−1.13211
$Z_4(\omega)$	0.01459	—	0.17115	0.29258	0.52168	0.46733	−0.76845	−1.01930
$Z_5(\omega)$	0.06893	0.06893	0.06893	0.06893	0.72427	0.46723	−0.77078	−0.78477

Table 6: Optimal weights of $Z_i(\omega)$, $i \in \{2, \dots, 5\}$, for the global minimum Expected-Shortfall portfolio at 5%. Results are obtained via Smolyak quadrature with level $\ell = 6$ for all dimensions.

	$d = 2$		$d = 3$		$d = 4$		$d = 5$	
	time	K	time	K	time	K	time	K
C^{-j}	0.0227	93.31%	1.8909	97.22%	13.0601	88.59%	59.9985	98.98%
density	0.2031	—	7.1341	—	55.5645	—	738.4296	—

Table 7: Computation time in minutes for the global optimal 5% Expected-Shortfall portfolio weights given in Table 6 either based on Theorem 3.2 (i) and (iii)(b) (C^{-j}), or on Proposition 3.1 and Theorem 3.2 (iii)(a) (density). For C^{-j} we report the percentage of overshoots over $K = 1.5$.

compute the optimal portfolios reported in Table 6 without involving C^{-j} . The computations are therefore only based on Proposition 3.1 and Theorem 3.2 (iii)(a). As shown in Table 7 the computation times reported in minutes increase when C^{-j} is not used. This reflects the additional integration that is performed in this case. Although the threshold $K = 1.5$ discussed in Section 4.2 is repeatedly passed over when utilizing C^{-j} , it is important to notice that in the case of overshooting K , only 1 out of d integrals, see Theorem 3.2, is computed with the slower method. This explains the benefits of reducing the integration domain even when violations happen. Although reducing the integration domain improves computation times as shown in Table 7, the utilized numerical integration and optimization are inherently subject to the curse of dimensionality. The usage of the proposed method for higher and high dimensional portfolios is therefore only realistically feasible, if the structure of the underlying model can in some way be exploited to speed up (some part of the) computations, or if advanced computational techniques such as parallelization, or a combination of the two, are utilized.

We further study the effect of the positivity, or no short selling, constraint (C1) by re-considering $\mathbf{X}_8$. In contrast to Figure 2, we show in Figure 3 the optimal weights when short selling, i.e., negative weights are allowed. The feasible region when both, (C1) and (C2) are in effect is indicated by the gray triangle, while the contour lines are now also given outside of the simplex. From Figure 2, we see that the optimal solutions under (C2) are forced to stay inside the simplex, while this is not necessary for the solutions with short selling indicated in

Figure 3: Contour lines for $\mathbb{E}[Y|Y \geq F_Y^{-1}(0.95)]$ for $Y = -\omega^T \mathbf{X}_8$, see Table 2 for details. The global minimum is marked by the solid black dot, while the gray area represents the feasible set under constraints (C1) and (C2). Circles represent optimal weights when only a minimum return constraint $r \in \{r_1, \dots, r_{20}\}$, where $r_j = 0.455 + (j - 1)0.005$ for $j = 1, \dots, 20$, is enforced. The solid line indicates the budget constraint for $r_6 = 0.48$. Results are obtained via Smolyak quadrature with level $\ell = 7$.

min. return	ω_1	ω_2	ω_3	$\mathbb{E}[Z(\omega)]$	$\text{VaR}_{0.05}[Z(\omega)]$	$\text{ES}_{0.05}[Z(\omega)]$
$r_{17} = 0.535$	0.70000	0.30000	0.00000	0.53500	-1.07054	-1.80495
$r_{18} = 0.540$	0.80000	0.20000	0.00000	0.54000	-1.12318	-1.90515
$r_{19} = 0.545$	0.90000	0.10000	0.00000	0.54500	-1.18999	-2.02649
$r_{20} = 0.550$	1.00000	0.00000	0.00000	0.55000	-1.25896	-2.16773

Table 8: Properties of $Z(\omega) = \omega^T \mathbf{X}_8$ for given minimum return constraints $r_j = 0.455 + (j - 1)0.005$ for $j = 17, \dots, 20$. Here $\omega = (\omega_1, \omega_2, \omega_3)$ are the optimal minimum Expected-Shortfall portfolio weights at 5% obtained under the no short selling constraint (C1).

Figure 3. Under (C2), the last four optimal solutions are therefore forced to the boundary of the simplex, while this is not the case with short selling. The results without short selling are reported in Table 8, while Table 9 reports the results in the case of short selling. In line with expectations, short selling decreases the risk while achieving the same minimum return as the constrained solutions; see Table 8 and Table 9.

5.2 Real Data Example: Apple, Microsoft, IBM, Bank of America Corporation, and Pacific Gas and Electric Company

Figure 4 and Figure 5 show consecutive weekly gross returns $R_{i,t} = (S_{i,t} - S_{i,t-1})/S_{i,t-1}$ based on Friday closing prices $S_{i,t}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$), and Pacific Gas and Electric Company ($i = 5$) from July 1997 to January

min. return	ω_1	ω_2	ω_3	$\mathbb{E}[Z(\boldsymbol{\omega})]$	$\text{VaR}_{0.05}[Z(\boldsymbol{\omega})]$	$\text{ES}_{0.05}[Z(\boldsymbol{\omega})]$
$r_{17} = 0.535$	0.64901	0.40199	-0.05099	0.53500	-1.06873	-1.79683
$r_{18} = 0.540$	0.69471	0.41057	-0.10529	0.54000	-1.11265	-1.87062
$r_{19} = 0.545$	0.74454	0.41093	-0.15546	0.54500	-1.15956	-1.94692
$r_{20} = 0.550$	0.79638	0.40723	-0.20362	0.55000	-1.20676	-2.02611

Table 9: Properties of $Z(\boldsymbol{\omega}) = \boldsymbol{\omega}^\top \mathbf{X}_8$ for given minimum return constraints $r_j = 0.455 + (j - 1)0.005$ for $j = 17, \dots, 20$. Here $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ are the optimal minimum Expected-Shortfall portfolio weights at 5% obtained with short selling.

2010³. The vertical dashed line in the figures marks our pick of the beginning of the financial crisis on August 7, 2007. Although the first indications of financial turmoil can be found earlier in 2007, the second week of August 2007 has been identified as a pivotal moment in the development of the financial crisis with equity hedge funds suffering unprecedented losses; see for example Khandani and Lo (2007, 2011). On August 7 and August 8 especially model-driven long/short equity funds suffered severe losses, see Khandani and Lo (2007), which is why we pick August 7, 2007 to indicate the start of the financial crisis. This is also in line with other recent research, see e.g. Beirlant et al. (2016). Our cut off date ensures to cover the most relevant post-crises data.

A successful joint model has to address the apparent volatility clustering within the marginal time series, as well as the dependence between the returns. We therefore consider a Copula-GARCH model, see Patton (2013) for an overview, with the following specifications. Each marginal return time series is modelled by a GARCH(1, 1) process as

$$R_{i,t} = \mu_i + \varepsilon_{i,t}, \quad \text{where} \quad \varepsilon_{i,t} = \sigma_{i,t}\eta_{i,t}, \quad \text{and} \quad \sigma_{i,t}^2 = \gamma_i + \alpha_i\sigma_{i,t-1}^2 + \beta_i\varepsilon_{i,t-1}^2. \quad (11)$$

Finally the independent and identically distributed (iid) innovations $(\eta_{1,t}, \eta_{2,t}, \eta_{3,t}, \eta_{4,t}, \eta_{5,t})$ follow a joint distribution with a Gaussian copula $C_{\mathbf{R}}$, $\mathbf{R} \in \mathbb{R}^{5 \times 5}$, and Student-t margins with ν_i degrees of freedom, appropriately scaled to achieve unit variance. To check for time variation in the parameters, we proceed with the estimation in a rolling window type fashion. For each time series, we thus divide the data into 132 blocks of 523 observations by including and excluding always one observation at a time.

Given a block $B_{i,k}$, $k = 1, \dots, 132$, of data, we fit the GARCH(1, 1) model leading to estimated parameters $(\hat{\mu}_{i,k}, \hat{\gamma}_{i,k}, \hat{\alpha}_{i,k}, \hat{\beta}_{i,k}, \hat{\nu}_{i,k})_{i=1}^5$. Based on the estimated parameters, we obtain

³Data downloaded from the Yahoo finance service <http://finance.yahoo.com/>.

time series $\hat{\eta}_{i,k}$ of 523 iid residuals each, that are then used to estimate the ten copula parameters constituting $\hat{\mathbf{R}}_k$. As an example, the first rolling window to include crisis data is obtained for $k = 6$ and stretches from 08.08.1997 to 10.08.2007.

The dependence between the time series becomes apparent in Figure 6, where we show pairwise scatterplots of $(\hat{\eta}_{1,1}, \hat{\eta}_{2,1}, \hat{\eta}_{3,1}, \hat{\eta}_{4,1}, \hat{\eta}_{5,1})$. The positive association between the errors visible in Figure 6 is further supported by the estimated copula parameter $\hat{\mathbf{R}}_k$ shown in Figure 7. The positive estimates for $\mathbf{R}_k$ translate into positive values for Kendall's tau of the implied bivariate marginal distributions, see Table 3.

In Figures 8–12, we show the estimated GARCH parameters based on the blocks $B_{i,k}$. While Apple consistently has the highest conditional return $\hat{\mu}_{1,k}$ (Figure 8), it also features the highest baseline variance $\hat{\gamma}_{1,k}$ (Figure 9) for the majority of the rolling windows, and at the same time it features the highest dependence on the past shocks $\hat{\beta}_{i,k}$ (Figure 11). For the dependence on the past realized variance $\hat{\alpha}_{1,k}$ (Figure 10) the picture reverses the order of the assets compared to $\hat{\beta}_{i,k}$ (Figure 11). The degrees-of-freedom parameters $\hat{\nu}_{i,k}$ (Figure 12) are generally between 3.5 and 6.5 indicating considerably heavier tails compared to an alternative normal model. From Figure 12, we see that Apple, Microsoft and IBM, representing the technology sector, are generally lighter tailed than Bank of America Corporation and Pacific Gas and Electric Company, representing the financial and utility sector. While the parameters show overall a stable behaviour, all stocks display a downward jump in the evolution of the degrees-of-freedom parameter either going from rolling window $k = 65$ to $k = 66$ (Apple, IBM), or going from rolling window $k = 66$ to $k = 67$ (Microsoft, Bank of America Corporation, Pacific Gas and Electric Company). Given that the data covered by the respective rolling windows are 25.09.1998–26.09.2008 for $k = 65$, 02.10.1998–03.10.2008 for $k = 66$ and 09.10.1998–10.10.2008 for $k = 67$ we can see from the gross returns in Figure 4 and Figure 5 that going from 65 to 66, or from 66 to 67 respectively, incorporates large fluctuations of gross returns into the data. This in turn leads to heavier tails of the considered univariate models.

Based on the estimated parameters, we turn to the construction of optimal portfolios based on the 132 rolling windows. Starting from the implied conditional joint distributions of gross returns, we compute⁴ conditional global minimum Expected-Shortfall portfolios, utilizing Smolyak

⁴Computations are carried out with help of GNU parallel, see Tange (2011), and the C++ libraries Boost, Eigen and NLOpt.

quadrature with nested Gaussian quadrature and $\ell = 6$. Figure 13 shows the resulting optimal relative weights $\omega_{i,k}$ based on the blocks $B_{i,k}$, $k = 1, \dots, 132$, for the global 1% minimum Expected-Shortfall portfolio.

From the gross returns in Figure 4 and Figure 5, we see that the Bank of America Corporation, representing the financial sector, is affected the most by the crisis among the five stocks. The increased risk on a marginal level is also reflected in the parameter estimates. While the return for Bank of America Corporation is fairly stable, the degrees-of-freedom parameter is declining over time while the dependence on the past realized variance $\hat{\alpha}_{1,k}$ shows an increase, especially during the beginning of the crisis. Considering the Expected-Shortfall optimal relative weights in Figure 13, this increased riskiness is reflected by a drastic drop of the associated relative weight right at the beginning of the crisis when going from $\omega_{4,6} = 0.4012$ to $\omega_{4,7} = 0.1675$. For $k \geq 7$ the optimal weights $\omega_{4,k}$ follow a downward trend, only to slightly increase towards the final rolling windows. It is important to notice that the optimal portfolio decomposition is unstable at times, and a drastic change in weights is not specific to Bank of America Corporation. The investment in Bank of America Corporation is however consistently low, suggesting a persistent shift into either the technology or utility sector over the course of the crisis in this example. From a practical perspective, the observed instability of the portfolio composition in Figure 13 is undesirable. While the instability may hint towards model misspecification, this is not necessarily the case. Figure 14 shows that the optimal weights for the Markowitz portfolio tend to be more stable for this dataset. To combat the observed instability of the portfolio decomposition, it would be possible to penalize for large changes in the weights during the optimization. While this would necessarily lead to an increase in portfolio risk, it could avoid larger rearrangements of the portfolio structure from one period to another.

We also transform the (conditional) relative weights $\omega_{i,k}$ into absolute weights $\delta_{i,k}$ that constitute the time t portfolio value via $V_{P,t} = \sum_{i=1}^5 \delta_{i,t} S_{i,t}$. When $\omega_{i,t}$ denotes as usual the weights used for the portfolio gross return $R_{P,t} = \sum_{i=1}^5 \omega_{i,t} R_{i,t}$ it follows directly that $\delta_{i,t} = \frac{\omega_{i,t}}{S_{i,t}} V_{P,t}$. As the total value of the investment at the start of each period is fixed by the investor we set it to $V_{P,t} = 1000$ USD in the following computations. The corresponding optimal number of shares $\delta_{i,k}$ are shown in Figure 15. Due to the scaling by the stock price $S_{i,t}$, the lower investment in Bank of America Corporation over the crisis period is less obvious in absolute terms compared

to relative ones.

To highlight the effect of selecting the Expected-Shortfall optimal portfolio, we consider additionally two investment strategies alternative to the optimization based approach outlined so far. Firstly, we consider an equal investment strategy where we set $\omega_{i,k} = 0.2$ for $i = 1, \dots, 5$ and $k = 1, \dots, 132$. Secondly, we show the effect of model misspecification by utilizing the Markowitz (1952) model, which is a limiting special case of the considered setup. To this end, we re-estimate the marginal time series with normal margins, and then use the obtained residuals to fit a Gaussian copula. As with the more general model, the estimation is carried out for all 132 rolling windows. We then select global 1% minimum Expected-Shortfall weights under this jointly normal model. Utilizing this model will highlight the impact of selecting optimal weights under a simplified model, while evaluating the risk under the full model.

In Figure 16, we show the Expected-Shortfall under the full model (11) when using the three different sets of weights. The first notable difference in the risk profiles is visible when going from rolling window $k = 18$ to $k = 19$.⁵ For the Markowitz weights, the Expected-Shortfall drops by 0.03220039. In comparison, for the optimal weights of the full model the Expected-Shortfall only drops by 0.01244976 when adjusting to the new data. We visualize the difference between the two risk profiles in Figure 17, where we indicate the rolling windows $k = 18$ and $k = 66$ by vertical dashed lines. From Figure 17 it can be seen that from rolling window $k = 19$ on the differences are generally more distinct. Pronounced differences arise again in the rolling windows after $k = 66$, showing that the two risk profiles are not merely shifted by a constant. Given the model setup, where the Markowitz model is a special case of the full model (11), the differences in the risk profile can be attributed to the inflexibility at the marginal level. This view is supported in Figure 12 when looking at the degrees-of-freedom parameter for Microsoft and IBM. In case of Microsoft, the degrees-of-freedom parameter drops when going from $k = 16$ to $k = 17$, while a drop in the parameter for IBM is visible when going from $k = 18$ to $k = 19$. This suggest that the observed difference in the risk profiles is due to the inflexibility of the Markowitz model to capture the increased tail heaviness present in the data.

For all models, we can also see an increase in risk when going from rolling window $k = 66$ to $k = 67$, which is in agreement with the previously identified changes in the degrees-of-freedom

⁵Rolling window $k = 18$ covers the data from 31.10.1997 to 02.11.2007, while rolling window $k = 19$ covers the data from 07.11.1997 to 09.11.2007.

parameters. As expected, the optimal Expected-Shortfall portfolio based on the full model shows the lowest risk for all rolling windows, represented by the highest Expected-Shortfall at all times. While there is less of a difference to the equal investment and Markowitz weights for the early rolling windows, the excess risk increases for later rolling windows. In the case of the Markowitz based weights, we see that the underlying optimization, even though it is based on a simplified model, adapts to the changes in the underlying joint distribution. Of course, this does not hold for the equal weights strategy, where especially for the rolling windows following the risk increase at $k = 67$ the excess risk becomes quite large.

Figure 18 shows the Value-at-Risk associated to the optimal portfolios based on the three different weights. It is interesting to see that the optimal 1% Expected-Shortfall portfolio is not necessarily preferable in terms of the Value-at-Risk. In fact, the equal investment and Markowitz weights both lead to a more preferable Value-at-Risk for certain rolling windows.

To get a better understanding of the implied differences we also use Theorem 3.2 (ii) to compute the density f_Z for the three types of weights for the rolling windows $k = 21$ and $k = 90$, see Figure 19. While there is little to no difference between the implied densities during the less risky regime in place for $k = 21$, there are pronounced differences during the more risky period for $k = 90$.

From an economic point of view, the global minimum Expected-Shortfall portfolio constitutes the choice of a maximally risk averse investor. Due to diversification effects, this is however not equivalent to an investment into the least risky asset only. Taking a look at the expected returns of the global 1% minimum Expected-Shortfall portfolio in Figure 8 (solid-circle black line) shows, that an educated choice of weights dominates the naive strategy by offering substantially more return coupled with lower risk. Implementing an effective minimum return constraint (C3) offers optimal choices for investors willing to accept more risk in exchange for a higher expected portfolio return. The respective optimal portfolios make it possible to reach all returns between those of the global 1% minimum Expected-Shortfall portfolio and the asset with the highest individual return, Apple in our case. Visually, this is represented by the space between the gray solid-circle and the gray solid-cube lines in Figure 8.

Figure 4: Consecutive weekly gross returns for Apple (top), Microsoft (middle) and IBM (bottom) from July 1997 to January 2010. The vertical dashed line marks the beginning of the financial crisis on August 7, 2007.

Figure 5: Consecutive weekly gross returns for Bank of America Corporation (top) and Pacific Gas and Electric Company (bottom) from July 1997 to January 2010. The vertical dashed line marks the beginning of the financial crisis on August 7, 2007.

Figure 6: Scatterplot of the standardized GARCH residuals $(\hat{\eta}_{1,1}, \hat{\eta}_{2,1}, \hat{\eta}_{3,1}, \hat{\eta}_{4,1}, \hat{\eta}_{5,1})$.

Figure 7: Estimated parameters $(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{14}, \hat{\rho}_{15}, \hat{\rho}_{23}, \hat{\rho}_{24}, \hat{\rho}_{25}, \hat{\rho}_{34}, \hat{\rho}_{35}, \hat{\rho}_{45})_{k=1}^{132}$ constituting the off-diagonal elements of $\hat{\mathbf{R}}_k$ for the Gaussian copula based on the standardized GARCH residual series.

Figure 8: Estimated parameter $\hat{\mu}_{i,k}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$. The solid line shows the expected return of the global 1% minimum Expected-Shortfall portfolio for each rolling window.

Figure 9: Estimated parameter $\hat{\gamma}_{i,k}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 10: Estimated parameter $\hat{\alpha}_{i,k}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 11: Estimated parameter $\hat{\beta}_{i,k}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 12: Estimated parameter $\hat{v}_{i,k}$ for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 13: Optimal relative weights $\omega_{i,k}$ for the minimum 1% Expected-Shortfall portfolio for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 14: Relative weights $\omega_{i,k}$ for the Markowitz 1% Expected-Shortfall portfolio for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$.

Figure 15: Optimal absolute weights $\delta_{i,k}$ for the minimum 1% Expected-Shortfall portfolio for Apple ($i = 1$), Microsoft ($i = 2$), IBM ($i = 3$), Bank of America Corporation ($i = 4$) and Pacific Gas and Electric Company ($i = 5$) for the rolling windows $B_{i,k}$. The total amount of investment for each rolling window is set to $V_{P,t} = 1000$ USD.

Figure 16: 1% Expected-Shortfall of the global 1% minimum Expected-Shortfall portfolio, equal weights portfolio and Markowitz portfolio for the rolling windows $B_{i,k}$.

Figure 17: Difference between the 1% Expected-Shortfall of the global 1% minimum Expected-Shortfall portfolio and the Markowitz portfolio for the rolling windows $B_{i,k}$. Vertical lines indicate rolling windows $k = 18$ and $k = 66$.

Figure 18: 1% Value-at-Risk of the global 1% minimum Expected-Shortfall portfolio, equal weights portfolio and Markowitz portfolio for the rolling windows $B_{i,k}$.

Figure 19: The implied density f_Z based on the weights of the global minimum 1% Expected-Shortfall portfolio (solid), equal investment (dashed) and Markowitz model (points). Calculations are based on the weights for $k = 21$ (left) and $k = 90$ (right).

6 Conclusion

In this paper we provide a framework that allows to select optimal portfolio weights with respect to the Expected-Shortfall as the underlying risk measure. The approach we present does not place any restrictions on the margins or copula of the underlying vector of asset returns. This provides a significant gain in flexibility compared to current approaches that perform similar tasks and place constraints on the joint distribution of the asset returns. The optimization objective function, taken from Rockafellar and Uryasev (2000), is guaranteed to be globally convex and is thus a particularly good target for the inevitable numerical optimization. In terms of the numerical computations we provide formulas for the necessary auxiliary functions C^{-j} for popular copula families, thus allowing for a wider application of the algorithms introduced in Gijbels and Herrmann (2014). Compared to Monte Carlo simulation based approaches, our methodology is non-random in the sense that the error committed by our algorithm depends on the given parameters in a purely deterministic fashion only.

For the discussed optimal portfolios, we experience the anticipated diversification effects between assets. In our illustrations, we however find that even joint distributions that are closely comparable in terms of the first two marginal moments and pairwise measures of dependence, can lead to substantially different optimal weights and portfolio Expected-Shortfall. Therefore, the examples provided call for caution in model specification on both, the marginal and the copula level. This highlights the need for flexible models that disentangle marginal modelling from dependence structure modelling instead of providing one fixed framework. Furthermore, we encounter scenarios where upon changing only marginal distributions we experience a rising Value-at-Risk and at the same time a falling Expected-Shortfall. This counter intuitive behaviour emphasizes the importance of choosing a sensible and task-specific underlying risk measure in actuarial and financial computations.

The price we pay for our fully flexible model is the computational cost. Compared to models allowing for (semi-) closed form solutions, our approach relies generally on multivariate numerical integration. Even when considering efficient multivariate integration techniques such as Quasi Monte Carlo or Smolyak integration, it is clear that we stand behind models that can exploit structural assumptions to arrive at (semi-) closed form solutions in terms of computation speed. While this possibly hinders the applicability to higher dimensional problems our real

Distribution of V	$\mathbb{E}[V V \leq z]$	Derived in
skew logistic	$m + s \left(\log \left(\frac{\mathcal{L}_{m,s}(z)}{1 - \mathcal{L}_{m,s}(z)} \right) - \frac{1}{\xi} - \mathcal{L}_{m,s}(z) \mathcal{L}(\mathcal{L}_{m,s}(z), 1, \xi + 1) \right)$	Appendix A
Student-t	$m + s \frac{v\Gamma((v+1)/2)}{F_V(z)\sqrt{v\pi}\Gamma(v/2)(1-v)} \left(1 + \frac{(z-m)^2}{vs^2} \right)^{-\frac{v-1}{2}}$	Appendix B
Laplace	$m - s \frac{\text{sgn}((z-m)/s)e^{- (z-m)/s }((z-m)/s + \text{sgn}((z-m)/s))}{2F_V(z)}$	Appendix C

Table 10: Conditional expectation $\mathbb{E}[V|V \leq z]$ for different marginal distributions.

data application shows the potential of our method.

Considering the results of our illustrative examples and rolling window real data example, we find that model misspecification strongly impacts the choice of optimal weights. For further research, we therefore suggest an in-depth model selection that not only addresses time variation on the marginal level, but possibly also addresses time variation in the dependence structure. Combining such a conditional copula modelling with our methodology would then allow to specifically address the time variation in the portfolio weights and the related risk measures.

From a practical point of view the variation in the optimal portfolio weights might be undesirable as the constant re-balancing raises transaction costs. We leave the introduction of a suitable penalization to safeguard against too frequent re-balancing for further research.

Appendix

A Expected-Shortfall for the type I skew logistic distribution

For $V \sim \mathcal{SL}(\cdot; m, s, \xi)$ we aim at calculating $\mathbb{E}[V|V \leq x] = (F_V(x))^{-1} \int_{-\infty}^x tf_V(t)dt$ and therefore first compute

$$\begin{aligned}
\int_{-\infty}^x tf_V(t)dt &= \int_0^{F_V(x)} F_V^{-1}(p)dp = mF_V(x) + s \int_0^{F_V(x)} \ln \left(\frac{p^{1/\xi}}{1 - p^{1/\xi}} \right) dp \\
&= mF_V(x) + s \left(\frac{1}{\xi} \int_0^{F_V(x)} \ln(p)dp - \int_0^{F_V(x)} \ln(1 - p^{1/\xi})dp \right) \\
&= mF_V(x) + s \left(\frac{1}{\xi} F_V(x)(\ln(F_V(x)) - 1) - \xi \int_0^{(F_V(x))^{1/\xi}} z^{\xi-1} \ln(1 - z)dz \right),
\end{aligned}$$

where substitution is used in the second and last equation. For the last remaining integral we use term-by-term differentiation of power series to get

$$\frac{d}{dz} \left(-z^\xi \sum_{n=1}^{\infty} \frac{z^n}{n^2 + n\gamma} \right) = - \sum_{n=1}^{\infty} \frac{d}{dz} \frac{z^{n+\xi}}{n^2 + n\gamma} = - \sum_{n=1}^{\infty} \frac{z^n}{n} z^{\xi-1} = z^{\xi-1} \ln(1-z),$$

with $\ln(1-z) = -\sum_{n=1}^{\infty} \frac{z^n}{n}$, for $|z| < 1$. This yields $\int z^{\xi-1} \ln(1-z) dz = -z^\xi \sum_{n=1}^{\infty} \frac{z^n}{n^2 + n\gamma} + C$, where the infinite series can be re-expressed by utilizing the identity

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{z^n}{n^2 + \xi n} &= \frac{1}{\xi} \sum_{n=1}^{\infty} \frac{(n+\xi)z^n - nz^n}{(n+\xi)n} = \frac{1}{\xi} \left(- \sum_{n=1}^{\infty} \frac{z^n}{(n+\xi)} + \sum_{n=1}^{\infty} \frac{z^n}{n} \right) \\ &= \frac{1}{\xi} \left(- \sum_{n=0}^{\infty} \frac{z^{n+1}}{(n+\xi+1)} + \sum_{n=1}^{\infty} \frac{z^n}{n} \right) = \frac{1}{\xi} \left(-z \sum_{n=0}^{\infty} \frac{z^n}{(n+\xi+1)} + \sum_{n=1}^{\infty} \frac{z^n}{n} \right) \\ &= \frac{-z\mathcal{L}(z, 1, \xi+1) - \ln(1-z)}{\xi}, \end{aligned}$$

where $\mathcal{L}$ denotes the Lerch transcendent given by $\mathcal{L}(z, s, \xi) = \sum_{n=0}^{\infty} \frac{z^n}{(n+\xi)^s}$. Taking all results together and inserting $G(x) = (F_V(x))^{1/\xi}$ we obtain

$$\begin{aligned} \mathbb{E}[V|V \leq x] &= \frac{1}{F_V(x)} \int_{-\infty}^x t f_V(t) dt \\ &= m + \frac{s}{F_V(x)} \left(F_V(x) \frac{\ln((G(x))^\xi) - 1}{\xi} + \xi F_V(x) \sum_{n=1}^{\infty} \frac{(G(x))^n}{n^2 + \xi n} \right) \\ &= m + s \left(\ln(G(x)) - \frac{1}{\xi} + \xi \frac{-G(x)\mathcal{L}(G(x), 1, \xi+1) - \ln(1-G(x))}{\xi} \right) \\ &= m + s \left(\ln \left(\frac{G(x)}{1-G(x)} \right) - \frac{1}{\xi} - G(x)\mathcal{L}(G(x), 1, \xi+1) \right). \end{aligned}$$

The result in (10) follows by recognizing that $G(x) = \mathcal{L}_{m,s}(x)$.

B Expected-Shortfall for the Student-t distribution

For $V = m + sT$, where $T \sim t_v$ with $v > 1$ and $m \in \mathbb{R}$ and $s > 0$ are location and scale parameters we use substitution to compute

$$\begin{aligned}\mathbb{E}[V|V \leq x] &= m + s \frac{\Gamma((v+1)/2)}{F_V(x)} \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\Gamma(v/2)} \int_{-\infty}^{\frac{x-m}{s}} t \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} dt \\ &= m + s \frac{\Gamma((v+1)/2)}{F_V(x)\sqrt{v\pi}\Gamma(v/2)} \frac{v}{1-v} \left(1 + \frac{t^2}{v}\right)^{-\frac{v-1}{2}} \Big|_{t=-\infty}^{t=\frac{x-m}{s}} \\ &= m + s \frac{v\Gamma((v+1)/2)}{F_V(x)\sqrt{v\pi}\Gamma(v/2)(1-v)} \left(1 + \frac{(x-m)^2}{vs^2}\right)^{-\frac{v-1}{2}}.\end{aligned}$$

C Expected-Shortfall for the Laplace distribution

By partial integration we have $\int s \frac{1}{2} e^{-s} ds = -\frac{1}{2} e^{-s}(s+1) + C$ and $\int s \frac{1}{2} e^s ds = \frac{1}{2} e^s(s-1) + C$.

For the standard Laplace distribution this leads to

$$\begin{aligned}\int_{-\infty}^x s \frac{1}{2} e^{-|s|} ds &= \begin{cases} \int_{-\infty}^x s \frac{1}{2} e^s ds = \frac{1}{2} e^x(x-1) & \text{for } x < 0 \\ \int_{-\infty}^0 s \frac{1}{2} e^s ds + \int_0^x s \frac{1}{2} e^{-s} ds = -\frac{1}{2} e^{-x}(x+1) & \text{for } x \geq 0 \end{cases} \\ &= -\frac{1}{2} \widetilde{\text{sgn}}(x) e^{-|x|} (x + \widetilde{\text{sgn}}(x)),\end{aligned}$$

where $\widetilde{\text{sgn}}(x) = \text{sgn}(x)$ for $x \neq 0$ and $\widetilde{\text{sgn}}(0) = 1$ in the above formula. By substitution it follows for a random variable V following a Laplace distribution with location $m \in \mathbb{R}$ and scale $s > 0$

$$\begin{aligned}\mathbb{E}[V|V \leq x] &= \frac{1}{F_V(x)} \int_{-\infty}^x t f_L(t) dt \\ &= m - s \frac{\widetilde{\text{sgn}}((x-m)/s) e^{-|(x-m)/s|} ((x-m)/s + \widetilde{\text{sgn}}((x-m)/s))}{2F_V(x)}.\end{aligned}$$

D Expected-Shortfall Optimization Gradient

To compute the gradient $\nabla \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = (\frac{\partial}{\partial \omega_1} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha), \dots, \frac{\partial}{\partial \omega_d} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha))^T$ based on (1), we separately compute $\frac{\partial}{\partial \omega_k} F_{Z(\boldsymbol{\omega})}(z)$ and $\frac{\partial}{\partial \omega_k} \mathbb{E}[Z(\boldsymbol{\omega})|Z(\boldsymbol{\omega}) \leq z]$ for $k \in \{1, \dots, d\}$. Starting from Theo-

rem 3.2 (i) we get

$$\frac{\partial}{\partial \omega_k} F_Z(z) = \int_{\mathbb{I}^{d-1}} c(u_1, \dots, u_{j-1}, \tau_z^j(u_1, \dots, u_{d-1}), u_j, \dots, u_{d-1}) \frac{\partial}{\partial \omega_k} \tau_z^j(u_1, \dots, u_{d-1}) d\mathbf{u},$$

where a condition comparable to the one in Theorem 3.2 (ii) is needed to interchange derivatives and integration. Setting $\kappa_{z,j} = \frac{z}{\omega_j} - \sum_{i=1}^{j-1} \frac{\omega_i}{\omega_j} F_i^{-1}(u_i) - \sum_{i=j+1}^d \frac{\omega_i}{\omega_j} F_i^{-1}(u_{i-1})$ we further get

$$\frac{\partial}{\partial \omega_k} \tau_z^j(u_1, \dots, u_{d-1}) = \begin{cases} -f_j(\kappa_{z,j}) \frac{F_k^{-1}(u_\ell)}{\omega_j}, & \text{if } k \neq j \text{ where } \ell = k \text{ if } k < j \text{ and } \ell = k - 1 \text{ if } k > j, \\ -f_j(\kappa_{z,j}) \frac{\kappa_{z,j}}{\omega_j}, & \text{if } k = j. \end{cases}$$

Concerning $\frac{\partial}{\partial \omega_k} \mathbb{E}[Z|Z \leq z]$, we start from Theorem 3.2 (iii)(a) where we only need to concentrate on the integral terms given the above result. For $i \in \{1, \dots, d\}$ we have

$$\begin{aligned} & \frac{\partial}{\partial \omega_k} \int_{B_z^{j_i}} \omega_i F_i^{-1}(u_i) c(u_1, \dots, u_d) du_d \dots du_1 \\ &= \int_0^1 \dots \int_0^1 \frac{\partial}{\partial \omega_k} \int_0^{\tau_z^{j_i}(u_1, \dots, u_{j_i-1}, u_{j_i+1}, \dots, u_d)} \omega_i F_i^{-1}(u_i) c(u_1, \dots, u_d) du_{j_i} du_d \dots du_1, \end{aligned}$$

where we use Fubini's theorem and again the assumed interchangeability of integrals and derivatives. Next, we apply Leibniz's integral rule $\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(t, \theta) dt = \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(t, \theta) dt + f(b(\theta), \theta) b'(\theta) - f(a(\theta), \theta) a'(\theta)$ to the following cases:

1) $i \neq j_i$:

$$\begin{aligned} & \frac{\partial}{\partial \omega_k} \int_0^{\tau_z^{j_i}(u_1, \dots, u_{j_i-1}, u_{j_i+1}, \dots, u_d)} \omega_i F_i^{-1}(u_i) c(u_1, \dots, u_{j_i-1}, t, u_{j_i+1}, \dots, u_d) dt \\ &= \int_0^{\tau_z^{j_i}(\cdot)} \frac{\partial}{\partial \omega_k} [\omega_i F_i^{-1}(u_i) c(u_1, \dots, u_{j_i-1}, t, u_{j_i+1}, \dots, u_d)] dt \\ &+ \omega_i F_i^{-1}(u_i) c(u_1, \dots, u_{j_i-1}, \tau_z^{j_i}(u_1, \dots, u_{j_i-1}, u_{j_i+1}, \dots, u_d), u_{j_i+1}, \dots, u_d) \frac{\partial}{\partial \omega_k} \tau_z^{j_i}(\cdot) \\ &= \begin{cases} \omega_i F_i^{-1}(u_i) c(u_1, \dots, u_{j_i-1}, \tau_z^{j_i}(\cdot), u_{j_i+1}, \dots, u_d) \frac{\partial}{\partial \omega_k} \tau_z^{j_i}(\cdot), & \text{for } i \neq k, \\ \int_0^{\tau_z^{j_k}(\cdot)} F_k^{-1}(u_k) c(u_1, \dots, u_{j_k-1}, t, u_{j_k+1}, \dots, u_d) dt \\ + \omega_k F_k^{-1}(u_k) c(u_1, \dots, u_{j_k-1}, \tau_z^{j_k}(\cdot) \frac{\partial}{\partial \omega_k} \tau_z^{j_k}(\cdot), & \text{for } i = k. \end{cases} \end{aligned}$$

2) $i = j_i :$

$$\begin{aligned}
& \frac{\partial}{\partial \omega_k} \int_0^{\tau_z^i(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_d)} \omega_i F_i^{-1}(t) c(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_d) dt \\
&= \int_0^{\tau_z^i(\cdot)} \frac{\partial}{\partial \omega_k} [\omega_i F_i^{-1}(t) c(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_d)] dt \\
&+ \omega_i F_i^{-1}(\tau_z^i(\cdot)) c(u_1, \dots, u_{i-1}, \tau_z^i(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_d), u_{i+1}, \dots, u_d) \frac{\partial}{\partial \omega_k} \tau_z^i(\cdot) \\
&= \begin{cases} \omega_i F_i^{-1}(\tau_z^i(\cdot)) c(u_1, \dots, u_{i-1}, \tau_z^i(\cdot), u_{i+1}, \dots, u_d) \frac{\partial}{\partial \omega_k} \tau_z^i(\cdot), & \text{for } i \neq k, \\ \int_0^{\tau_z^i(\cdot)} F_i^{-1}(t) c(u_1, \dots, u_{i-1}, t, u_{i+1}, \dots, u_d) dt \\ + \omega_i F_i^{-1}(\tau_z^i(\cdot)) c(u_1, \dots, u_{i-1}, \tau_z^i(\cdot), u_{i+1}, \dots, u_d) \frac{\partial}{\partial \omega_i} \tau_z^i(\cdot), & \text{for } i = k. \end{cases}
\end{aligned}$$

To compute $\frac{\partial}{\partial \omega_k} \mathbb{E}[Z|Z \leq z]$ one needs to utilize the product rule and pull the derivative inside the summation in Theorem 3.2 (iii)(a) using the above formulas. Finally we then have

$$\frac{\partial}{\partial \omega_k} \mathcal{F}_\beta(\boldsymbol{\omega}, \alpha) = - \frac{\alpha \frac{\partial}{\partial \omega_k} F_Z(-\alpha) + \left[\frac{\partial}{\partial \omega_k} F_Z(-\alpha) \right] \mathbb{E}[Z|Z \leq -\alpha] + F_Z(-\alpha) \frac{\partial}{\partial \omega_k} \mathbb{E}[Z|Z \leq -\alpha]}{1 - \beta}.$$

If partial derivatives of the Value-at-Risk are needed, they can be computed via $\frac{\partial}{\partial \omega_k} \text{VaR}(p) = \frac{\partial}{\partial \omega_k} F_Z^{-1}(p) = \left(\frac{\partial}{\partial \omega_k} F_Z(F_Z^{-1}(p)) \right)^{-1}$. Here, $F_Z^{-1}(p)$ needs to be evaluated only once for each set of $\boldsymbol{\omega}$, using e.g. the Newton-Raphson method in conjunction with the numerical computation of $F_Z(\boldsymbol{\omega})$ via Theorem 3.2 (i).

References

- Acerbi, C., Nordio, C. and C. Sirtori. (2001). Expected Shortfall as a Tool for Financial Risk Management. Working Paper (arXiv:cond-mat/0102304).
- Acerbi, C. and D. Tasche. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance* 26(7), 1487–1503.
- Aksenov, S., Savageau, M., Jentschura, U., Becher, J., Soff, G. and P. Mohr. (2003). Application of the combined nonlinear-condensation transformation to problems in statistical analysis and theoretical physics. *Computer Physics Communications* 150, 1–20.
- Artzner, P., Delbaen, F., Eber, J.M., and D. Heath. (1999). Coherent Measures of Risk. *Mathematical Finance* 9(3), 203–228.
- Bargès, M., Cossette, H. and É. Marceau. (2009). TVaR-based capital allocation with copulas. *Insurance: Mathematics and Economics* 45, 348–361.
- Beirlant J., Schoutens W., De Spiegeleer J., Reynkens T. and K. Herrmann. (2016). Hunting for Black Swans in the European Banking Sector Using Extreme Value Analysis. In: Kallsen J.,

- Papapantoleon A. (eds) Advanced Modelling in Mathematical Finance. Springer Proceedings in Mathematics & Statistics, vol 189, Springer.
- Bernardi, M. (2013). Risk measures for skew normal mixtures. *Statistics & Probability Letters* 83(8), 1819–1824.
- Cossette, H., Côté, M.P., Marceau, É and K. Moutanabbir. (2013). Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics* 52, 560–572.
- Czado, C. (2010). Pair-copula constructions of multivariate copulas, in P. Jaworski, F. Durante, W. Härdle and T. Rychlik (Eds.), *Copula Theory and Its Applications*. Berlin: Springer.
- Davis, P. and P. Rabinowitz. *Methods of numerical integration*. Academic Press, 2nd ed. 1984.
- Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A. and S. Vanduffel. (2008). Some results on the CTE-based capital allocation rule. *Insurance: Mathematics and Economics* 42(2), 855–863.
- Embrechts, P., McNeil, A. and D. Straumann. (2002). Correlation and dependency in risk management: properties and pitfalls, in M. Dempster (Ed.), *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press.
- Embrechts, P., A. Höing, and G. Puccetti (2005). Worst var scenarios. *Insurance: Mathematics and Economics* 37(1), 115–134.
- Fang, H. and K. Fang. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* 82, 1–16.
- Gerstner, T. and M. Griebel. (1998). Numerical integration using sparse grids. *Numerical Algorithms* 18, 209–232.
- Gijbels, I. and K. Herrmann. (2014). On the distribution of sums of random variables with copula-induced dependence. *Insurance: Mathematics and Economics* 59, 27–44.
- Hofert, M. (2010). Construction and sampling of nested Archimedean copulas, in P. Jaworski, F. Durante, W. Härdle and T. Rychlik (Eds.), *Copula Theory and Its Applications*. Berlin: Springer.
- Hofert, M., Mächler, M. and A.J. McNeil. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis* 110, 133–150.
- Hofert, M. and D. Pham. (2013). Densities of nested Archimedean copulas. *Journal of Multivariate Analysis* 118, 37–52.
- Jentschura, U., Mohr, P., Soff, G. and E. Weniger. (1999). Convergence acceleration via combined nonlinear-condensation transformations. *Computer Physics Communications* 116, 28–54.
- Johnson, N., Kotz, S., and N. Balakrishnan. *Continuous Univariate Distributions*. Volume 2. John Wiley & Sons, New York, 2nd ed. 1995.
- Kaarnioja, V. Smolyak Quadrature. Master’s thesis, University of Helsinki, 2013.
- Khandani, A.E. and A.W. Lo. (2007). What happened to the quants in August 2007? *Journal of Investment Management* 5, 5–54.

- Khandani, A.E. and A.W. Lo. (2011). What happened to the quants in August 2007? Evidence from factors and transaction data. *Journal of Financial Markets* 14, 1–46.
- Landsman, Z. and E. Valdez. (2003). Tail Conditional Expectations for Elliptical Distributions. *North American Actuarial Journal* 7(4), 55–71.
- Lindskog, F., McNeil, A. and U. Schmock. (2003). Kendall’s tau for elliptical distributions. In Bol, Nakhaeizade et al. (Eds.), *Credit Risk - measurement, evaluation and management*. Physica–Verlag, Heidelberg, 149–156.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* 7(1), 77–91.
- McNeil, A., Frey, R. and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, 2nd ed., 2015.
- Nelsen, R. *An introduction to copulas*. Springer, 2nd ed. 2006.
- Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- Patton, A. (2013). Copula Methods for Forecasting Multivariate Time Series. In G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting*, Volume 2, Springer Verlag.
- Rockafellar, R. and S. Uryasev. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 21–41.
- Shao, Q. (2002). Maximum likelihood estimation for generalised logistic distributions. *Communications in Statistics - Theory and Methods* 31(10), 1687–1700.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229–231.
- Smolyak, S. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics* (4), 240–243. Translation of *Doklady Akademii Nauk SSSR*.
- Stoer, J. and R. Bulirsch (1993). *Introduction to Numerical Analysis* (2nd ed.). Springer.
- Tange, O. (2011). GNU Parallel - The Command-Line Power Tool. ;login: *The USENIX Magazine* 36(1), 42–47.
- Windberger, T., and A. Zeileis. (2014). Structural Breaks in Inflation Dynamics within the European Monetary Union. *Eastern European Economics* 52(3), 66–88.