

Master's Thesis

| | |
|----------------------------------|---|
| Title of Master's Thesis: | Portfolio selection with Expected Shortfall and regular vine copulae with EVT marginals |
| Author: | Pedro Rangel Mejía |
| Student ID number: | 12122657 |
| Degree program: | Master of Science in Quantitative Finance |
| Examiner: | PD Dr. Ronald Hochreiter |

I hereby declare that:

1. I have written this master's thesis myself, independently and without the aid of unfair or unauthorized resources. Whenever content has been taken directly or indirectly from other sources, this has been indicated and the source referenced. I am familiar with the regulations specified in the Directive on Plagiarism and Other Types of Academic Fraud in Academic Theses.
2. This master's thesis has not been previously presented as an examination paper in this or any other form in Austria or abroad. *
3. This master's thesis is identical with the thesis assessed by the examiner.

September 6, 2023

Date

Signature

*This does not apply to master's theses written as part of WU cooperation programs (joint or double degrees).

Portfolio selection with Expected Shortfall and regular
vine copulae with EVT marginals



MASTER'S THESIS

for the attainment of the degree of **Master of Science**
at the Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien

Submitted by:

Pedro Rangel Mejía

Matriculation number:

12122657

Supervised by:

PD Dr. Ronald Hochreiter

Vienna, September 2023

Abstract

Extreme events are the main drivers of any successful asset allocation criterion. Extreme value theory (EVT) and copula theory provide sound statistical foundations to better account for extreme events and tail dependence in the univariate and multivariate case respectively. This thesis aims to combine both EVT and copula modeling techniques to improve scenario generation for one period stochastic portfolio optimization problems with expected shortfall (ES). The new methodology for scenario generation is tested using data from financial assets that proxy the composition of a pension fund. EVT is used to estimate the marginal distributions of the multivariate returns distribution using semi-parametric methods. In turn, the multivariate distribution is modeled via regular vine copulae. The performance of our technique is evaluated through the computation of ES optimal efficiency frontiers. Our technique yields more stable frontiers than other alternatives and does not necessarily sacrifice expected returns at the cost of stricter assumptions.

Keywords: Stochastic portfolio optimization; financial scenario generation; dependence modeling; extreme value theory; vine copula.

Acknowledgements

In the first place, I would like give a great "Danke schön" to my supervisor, Prof. Hochreiter, for the time and energy he devoted to supervising my thesis. I am deeply grateful for the open mind he had when I first told him I wanted to do something EVT-related, for the way he helped me steer my thesis and for the liberty he gave me to do so. I would also like to thank Prof. Hornik for his suggestion of using vine copulae—which quickly became a core element of the thesis—and for his patience answering my questions on the topic after the DTM lectures. Furthermore, I would like to thank all the QFin Professors I had in the last two years. This master's degree has been intense and intellectually challenging, but all I learned from them made it worth it. A special mention is dedicated to my colleagues at SMN Investment Services, who have given me a great playground to apply what I had learned in the last two years and who had immense patience with me while I finished my studies.

At last, I do not think there are sufficient words to express how indebted I am to my father, Luis Carlos, for his support and for always breeding my intellectual curiosity, and to my mother, Bertha Lucía, who, even despite the ocean that lies between us, has been my source of strength, reason and tranquility in the most difficult moments of my student life.

Contents

| | | |
|-------------------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Stochastic portfolio optimization with expected shortfall | 4 |
| 2.1 | Set-up | 5 |
| 2.2 | Discretization, efficient frontier and linearization | 6 |
| 3 | EVT marginal distributions and vine copulae | 7 |
| 3.1 | EVT and the Peaks-Over Threshold Method | 8 |
| 3.1.1 | Convergence of maxima and the Block Maxima Method | 9 |
| 3.1.2 | Convergence of threshold exceedances and the POT method | 10 |
| 3.1.3 | Semi-parametric mixture distribution | 12 |
| 3.2 | A primer on vine copulae | 14 |
| 3.2.1 | Basic properties of copulae | 15 |
| 3.3 | Pair copula decompositions | 16 |
| 3.3.1 | Regular vine copulae | 17 |
| 4 | Case study: a multi-asset class portfolio | 20 |
| 4.1 | Data | 20 |
| 4.2 | Fitting of the marginal distributions | 21 |
| 4.2.1 | Threshold selection | 22 |
| 4.3 | Scenario generation techniques | 26 |
| 4.4 | Selection and fitting of R-vine copulae | 28 |
| 4.5 | Optimal portfolios and efficiency frontiers | 29 |
| 4.6 | Results | 31 |
| 4.6.1 | Minimum risk portfolios | 31 |
| 4.6.2 | Efficiency frontiers | 34 |
| 5 | Conclusions | 39 |
| A | Appendix | 41 |
| A.1 | Additional figures | 41 |
| References | | 48 |

List of Figures

| | | |
|------|---|----|
| 1 | R-vine structures for $d = 3$ and $d = 4$ | 19 |
| 2 | Encoded R-vine structures with SPD marginals | 29 |
| 3 | Efficiency frontiers for different scenario set generation methods | 35 |
| 4 | Efficiency frontiers for different threshold selection methods | 36 |
| 5 | Efficiency frontiers using different marginal distributions for scenario generation . | 37 |
| 6 | Efficiency frontiers using different optimization methods | 38 |
| A.7 | Q-Q plots of sample returns | 41 |
| A.8 | Sample mean excess plots for upper tail exceedances | 42 |
| A.9 | Sample mean excess plots for lower tail exceedances | 43 |
| A.10 | Shape plots for upper tail exceedances | 44 |
| A.11 | Shape plots for lower tail exceedances | 45 |
| A.12 | R-vine structure with SPD marginals fitted using the $n_u = 150$ rule | 46 |
| A.13 | R-vine structure with SPD marginals fitted using the 10th/90th quantile rule . . | 47 |

List of Tables

| | | |
|---|--|----|
| 1 | List of assets | 21 |
| 2 | Descriptive statistics | 22 |
| 3 | Sample ordinal correlations | 23 |
| 4 | Estimated parameters - 150 observations beyond u | 25 |
| 5 | Estimated parameters - 10th/90th quantiles of $F(x_i)$ | 26 |
| 6 | Copula families of selected R-vine structures | 27 |
| 7 | Minimum risk portfolios - Different scenario generation techniques | 33 |

1 Introduction

The question of how to take decisions under uncertainty in realms where unlikely events have extreme consequences—and it is not possible to predict them *ex ante*—is an old philosophical problem dubbed the induction problem or, more recently, the black swan problem. Although the philosophical ideas of decision making in the face of black swans can be traced back to the skeptical school of Sextus Empiricus, all the way up to David Hume, having even reached popular audiences in recent years through works like Taleb (2010) or Spitznagel (2021), the development of sound mathematical and statistical methods to understand how extreme events behave is relatively recent. On the one hand, Extreme Value Theory (EVT) has provided a rigorous framework to understand the behavior of tail distributions, delivering straightforward and relatively simple techniques to model and estimate them, particularly in the univariate case. On the other hand, for the multivariate case, copula theory makes it possible to recover the multivariate distribution of random vectors by eliminating the influence of their respective marginal distributions, effectively delivering the actual dependence structure between random variables as a result. The dependence structure captures the pure co-movement of random variables with different marginals, making it possible to understand and model the joint interaction between extreme events in a multivariate framework. Therefore, when trying to account for extreme events in such a framework, the union between EVT and copula theory is a natural one, which is why the present work will take advantage of that organic correspondence in order to improve scenario generation for stochastic portfolio optimization problems.

In portfolio management, the main aim of any asset allocation criterion should be to maximize the compound average growth rate (CAGR) of the portfolio over time. To achieve this, the first and foremost necessary condition is to avoid extreme losses, as these have a non-linear, negative impact on the CAGR of a portfolio, making robustness against extreme losses a priority for any sensible portfolio risk management approach. In this particular work, the allocation criterion is determined by means of stochastic portfolio optimization with expected shortfall (ES). The defining characteristic of a stochastic optimization problem is the use of a discrete probability distribution (the so-called scenario set) or realizations from it (the so-called scenarios) as objective or constraint function, which inevitably leads to the question of how to model scenarios and scenario sets. There are two broad families of approaches to solve this problem: the first one, which is mainly found in multistage stochastic optimization, is to approximate the underlying stochastic process of the assets' returns in order to generate scenarios and paths that serve as input for optimal decision making at different nodes of a decision tree; the second one is best suited for one period optimization problems and consists in estimating the multivariate distribution of asset returns and use it as scenario generation engine by drawing random samples from it (which can be used to generate both scenarios and full scenario sets). Overall, both approaches have a common aim, which is to discretize samples of a multivariate distribution or realizations from a stochastic process.

The first family of approaches is supported in the duality problem and the optimality conditions for multistage stochastic programming laid out by Rockafellar and Wets (1977), who demonstrate that these problems can be solved using Lagrange multipliers. In order to generate scenarios for these kind of optimization problems in a financial setting, a first thorough

discussion of different methods can be found in Dupačová et al. (2000), who outline a general procedure to generate scenarios and discuss the aptitude of different stochastic processes (random walk, binomial and trinomial processes and autoregressive models) to simulate scenarios and scenario trees. Høyland and Wallace (2001) formulate the scenario generation problem as an optimization problem, where the distance between a discrete distribution (which is the scenario set used as input in the optimization problem) and a continuous approximation is minimized. To do this, the first four moments of the discrete distribution are matched with those of a continuous candidate that solves the minimization problem. A heuristic algorithm to implement the moment matching method can be found in Høyland et al. (2003). However, one problem of this approach is that the resulting continuous approximations can be unstable and inaccurate. As Hochreiter and Pflug (2007) show in a simple example, the same discrete distribution could be approximated with equal optimality by a normal mixture or by a uniform distribution, which is why Hochreiter and Pflug (2007) reformulate the optimization problem and suggest the minimization of Wasserstein distance instead. Furthermore, an application of sparse grid methods is found in Chen et al. (2015), whose application is a significant improvement with respect to Monte Carlo and quasi Monte Carlo methods. Lastly, a survey of scenario generation methods and fields of application can be found in Di Domenica et al. (2007).

The second family of approaches is rather straightforward in comparison and falls into the broader category of multivariate distribution modeling. Accordingly, the main concern of this approach is to represent the dependence structure of the asset universe adequately and is not affected by the complexities inherent to simulating stochastic processes for multistage programming. With that in mind, the first and most simple idea would be to use the multivariate normal, which Konno and Yamazaki (1991) and Rockafellar and Uryasev (2000) use to illustrate their respective approaches to portfolio optimization—using mean-absolute-deviation in the former and expected shortfall in the latter. Although these two works use the multivariate normal to explain an approach to optimization and not as a normative prescription, elliptical distributions should be avoided by all means, as their pitfalls when tail risk is a concern are well-understood and documented (Embrechts et al., 2002). Another intuitive idea would be to use the historical distribution of returns—again used by, e.g., Krokmal et al. (2001) to exemplify how their optimization framework works—, however, using historical data alone for real-world applications is strongly warned against because the historical distribution only assigns a non-zero probability to events that have been observed, such that the occurrence of extreme events is systematically underestimated.

To avoid the dangers of using elliptical distributions or pure historical data, a recent strain in the literature for portfolio optimization has started to deal with copula and dependence modeling techniques to overcome the aforementioned shortcomings. A first work of this kind can be found in Hatherley and Alcock (2007), who use the Clayton copula with normally distributed marginals in the three dimensional case to compute ES optimal portfolios. Low et al. (2013) also perform ES optimization and use the canonical Clayton vine copula with normal and skewed Student's t marginal distributions to make in- and out-of-sample tests in the 12-dimensional case. Furthermore, González-Pedraz et al. (2015) use data on three commodity futures contracts to estimate Gaussian, t and skewed t copulae with Student's t marginals and with time-varying

parameters, where optimal portfolios are computed using numerical methods. Bekiros et al. (2015) use five different risk measures (including ES) and canonical, drawable and regular vines with empirically distributed marginals to perform portfolio optimization on 20 mining stocks. Yew Low et al. (2016) use the Gaussian copula with marginals described through a GARCH-GJR process in a mean-variance optimization setting. Gijbels and Herrmann (2018) go back to ES optimization, but in a 6-dimensional setting, and use a more flexible vine copula approach, where they allow the marginals and the dependence structures to belong to a wider array of distribution and copula families. Sahamkhadam et al. (2018) use an ARMA-GARCH-EVT copula model (with special focus on elliptical and Achimedean copulae) to perform optimization on a portfolio of 10 stock indexes. At last, it is worth noting that Bayesian approaches can be found in Harvey et al. (2010), who use the skewed multivariate normal as prior multivariate distribution; and in Sahamkhadam et al. (2022), who use vine copulae to model the posterior distribution of returns in a Black-Litterman optimization framework (Black and Litterman, 1992) for the 50 constituents of the EuroStoxx 50.

In summary, the second approach to scenario generation tries to solve a problem of sampling from the best possible distribution and, since the aim of this work is to better take extreme events and tail dependence into account, this thesis will continue the line of work started by Low et al. (2013), Bekiros et al. (2015) and Gijbels and Herrmann (2018) to model dependence structures through vine copulae. We aim to improve the cited studies by addressing shortcomings in two areas: the selection of marginal distributions and the choice of copula families for the pair copulae that form the trees of the vine copula. With respect to the first point, Gijbels and Herrmann (2018) improve the choice of marginal distributions over Low et al. (2013) and Bekiros et al. (2015) by allowing the marginals to have a normal, a Laplace, a Student's t or a skewed logistic distribution and by choosing a wider array of copula families. However, if the objective is to better account for extreme events in the univariate case, the rigorous path is to implement the methods from EVT for the choice and modeling of the marginals. To this end, we model the marginal distributions using a semi-parametric distribution whose tails are estimated using the Generalized Pareto Distribution (GPD) and whose center is entirely non-parametric. The advantage of this procedure is that it combines the best out of parametric and non-parametric methods, as it assigns a non-zero probability for extreme events that have not yet been seen in the data, while dropping any distributional assumptions for the center of the distribution. Besides, the use of the GPD is by no means an arbitrary assumption, for its use is well-founded in one of the fundamental theorems of EVT. Concerning vine copulae, we use the so-called regular vines, first developed by Bedford and Cooke (2001, 2002) and implemented in a computationally efficient way by Difmann et al. (2013). This class of vines are more flexible than the alternatives mentioned thus far because the bivariate components used to build the vine can belong to virtually any copula family and, more importantly, because the way the multivariate copula is decomposed into a tree structure of bivariate copula pairs results in a higher number of pair permutations to choose from. The regular vine approach makes it possible to model dependence in a very accurate and flexible way, without the need for strong distributional assumptions.

Our approach is tested in-sample by computing the efficiency frontier of ES optimal portfolios

using our scenario generation method as input and is compared to other scenario generation techniques and to the mean-variance efficient frontier. For our empirical implementation, we use a multi-asset class data set of 13 assets which were chosen so as to mimic the portfolio selection problem of a pension fund. The aim of choosing such an asset universe is to test our method in an environment where the dependence between assets is not clear cut (as is when staying in one asset class, which was the case of the mentioned works), in the hope that more sophisticated dependence modeling procedures make a tangible difference in the robustness of asset allocation criteria. Ultimately, the question we will try to answer is whether our approach yields higher efficiency frontiers (meaning that higher reward is attainable for the same level of risk) or if it is just an all too conservative approach.

The contribution of this thesis is twofold: on the one hand, our technique to estimate the multivariate distribution of returns with vine copulae makes a contribution to the field of multivariate extreme value modeling, which has been mainly concerned with the formulation of parametric multivariate distributions and not with the use of copulae to recover the multivariate distribution through them. On the other hand, the joint use of EVT and regular vines to model the marginal distributions and the multivariate distribution will contribute to the literature on scenario generation for single stage stochastic optimization problems, especially in the recent literature strain concerned with the use of vine copulae for multivariate modeling.

The rest of the thesis is structured as follows: Section 2 lays out the set-up for stochastic portfolio optimization with ES; Section 3 recaps the key theorems of extreme value theory, explains the Peaks-Over-Threshold method and how it is used to construct the semi-parametric mixture used as main marginal distribution, and gives a brief introduction to vine copulae; Section 4 explains the implementation of our scenario generation approach to the portfolio selection problem of a multi-asset class portfolio, presents the main results and performs sensitivity analyses. To conclude, Section 5 summarizes the most important results of the thesis, its shortcomings and its possible expansions.

2 Stochastic portfolio optimization with expected shortfall

The choice of risk measure to evaluate our scenario generation method in a portfolio optimization problem is not a trivial one. The seminal framework of Markowitz (1952) uses standard deviation as risk measure, which results in a quadratic program. Until recently, one of the main drawbacks of this approach was its computational intensity as it is necessary to optimize over the covariance matrix of returns. Although this is no longer a concern because of the availability of efficient optimization engines and the possibility to approximate the problem with a linear program that uses mean-absolute deviation instead (Konno and Yamazaki, 1991), the Markowitzian approach is not appropriate for the goals of this work because of three reasons: firstly, the optimization is done over the second moment of the distribution—which implicitly assumes joint normality as the decision is defined solely by mean and variance—; secondly, because the assumption of joint normality implicitly leads to the assumption that tail events are symmetric in magnitude and, thirdly, because standard deviation is not a coherent risk measure in the sense of Artzner et al. (1999), for it is neither monotone, nor translation invariant (Hofert et al., 2020, p.117-f). An

attempt to improve the usability of Markowitz (1952) for real-world decision making was done by Black and Litterman (1992), however, it assumes that the equilibrium distribution of returns is normal and still relies on covariance to arrive at a decision vector. In that sense, it does not get around the weaknesses of classical mean-variance optimization that make it unsuitable for our purposes.

Consequently, in order to tackle the problem of accounting for heavy tails and to overcome the limitations of optimizing over the second moment alone, quantile-based risk measures should provide a useful alternative. The first candidate could be Value-at-Risk (VaR); however, as Pflug (2000) proves, VaR is not a convex risk measure and is thereby not coherent. In contrast to standard deviation, which is convex, the use of VaR for portfolio optimization leads to the formulation of a non-convex optimization problem, which is problematic because of the existence of multiple local minima. As Gaivoronski and Pflug (2005) show, the efficiency frontiers of VaR optimal portfolios are unstable and need a fairly complicated method to smooth the local irregularities of historic VaR. As a consequence and, due to the emphasis of this work on downside risk, expected shortfall is left as the best candidate. Unlike VaR, ES is a measure of tail risk that is computed over the entire loss tail (therefore containing more information about the nature of a tail event of a given distribution) and is coherent, as demonstrated by Pflug (2000).

In this vein, we choose the classic approach to portfolio optimization with ES laid out in the seminal works of Rockafellar and Uryasev (2000) and Krockmal et al. (2001). The main contribution of these works is to characterize VaR and ES in one function that can be discretized and linearized, which allows to use discrete realizations from probability distributions (e.g. to use historical data or to take random samples from some multivariate distribution) and results in a linear program. Although there exist fairly sophisticated technologies to solve for ES optimal portfolios (see, e.g., Hochreiter (2008) for an application of algorithms derived from natural computing), the solving of optimization problems in our work will be entirely based on the framework of Krockmal et al. (2001) due to its simplicity and we will limit ourselves to introduce minor changes in notation based on the textbook of Mansini et al. (2015).

2.1 Set-up

Let $\mathbf{x}_t = (x_1, \dots, x_J)^T \in \mathbb{R}^J$ be a random row vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ denoting the vector of returns for assets $j = 1, \dots, J$ at time point t . \mathbf{x}_t is a so-called scenario. There are $t = 1, \dots, T$ scenarios which occur with equal probability π_t , such that $\sum_{t=1}^T \pi_t = 1$. The matrix of returns for all J assets and all T scenarios, $\mathbf{X}_{T \times J}$, is called a scenario set, where x_{tj} is the return of asset j at time t , hence, the column vector of returns for an asset j over time is $\mathbf{x}_j = (x_1, \dots, x_t)$. The scenario set is a discrete realization of the distribution of returns, which has distribution function $F(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J) = F(\mathbf{X})$ and density $f(\mathbf{X})$. Furthermore, let $\mathbf{w} = (w_1, \dots, w_J)$ with $\sum_{j=1}^J w_j = 1$ be a vector of portfolio weights corresponding to the random variable

$$R(\mathbf{w}) = \sum_{j=1}^J x_j w_j, \quad (2.1.1)$$

which is the return of a portfolio for some arbitrary set of weights. Its expected value is a linear function given by

$$\mathbb{E}(\mathbf{w}) = \mu(\mathbf{w}) = \sum_{t=1}^T \pi_t \left(\sum_{j=1}^J x_j w_j \right). \quad (2.1.2)$$

A trading loss is assumed to have positive sign (and gains negative). Then, the loss distribution associated with some loss L , some decision vector \mathbf{w} and some scenario set \mathbf{X} is given by $f_L(\mathbf{w}, \mathbf{X})$. The cumulative distribution function of losses is thus

$$F_L(\mathbf{w}, \mathbf{X}) = \int f_L(\mathbf{w}, \mathbf{X}) f(\mathbf{X}) d\mathbf{X}. \quad (2.1.3)$$

It is assumed by (Krokhmal et al., 2001) that $F_L(\mathbf{w}, \mathbf{X})$ is non-decreasing and everywhere continuous with respect to \mathbf{X} .

Given some confidence level $\alpha \in (0, 1)$, the Value-at-Risk of the loss distribution for some portfolio with weights \mathbf{w} is given by

$$VaR_\alpha(\mathbf{w}) = \inf\{\mathbf{y} \in \mathbb{R} : F_L(\mathbf{w}, \mathbf{X}) \geq \alpha\}, \quad (2.1.4)$$

which is just the α -quantile of the loss distribution. Expected shortfall (ES) is defined as the expected loss in case threshold α is exceeded. In other words, it is the conditional expectation of exceeding VaR_α :

$$ES_\alpha(\mathbf{w}) = \mathbb{E}[F_L(\mathbf{w}, \mathbf{X}) | F_L(\mathbf{w}, \mathbf{X}) \geq VaR_\alpha],$$

and its integral form is given by

$$ES_\alpha(\mathbf{w}) = \frac{1}{1 - \alpha} \int_\alpha^1 f_L(\mathbf{w}, \mathbf{X}) f(\mathbf{X}) d\mathbf{X}. \quad (2.1.5)$$

Ideally, the risk constraint of the optimization problem should be convex (so as to eliminate the possibility of local minima, which is the drawback of using VaR alone) and it should be possible to turn it into a linear and discrete problem in order to solve a simple linear program. To this end, Krokhmal et al. (2001) characterize $ES_\alpha(\mathbf{w})$ and $VaR_\alpha(\mathbf{X})$ through a convex and continuously differentiable function $F_\alpha(\mathbf{w}, \beta)$, given by

$$F_\alpha(\mathbf{w}, \beta) = \beta + \frac{1}{1 - \alpha} \int_\alpha^1 [f_L(\mathbf{w}, \mathbf{X}) - \beta]^+ f(\mathbf{X}) d\mathbf{X}, \quad \beta \in (0, 1). \quad (2.1.6)$$

2.2 Discretization, efficient frontier and linearization

The integral in $F_\alpha(\mathbf{w}, \beta)$ can be discretized by sampling scenarios from the probability distribution $F(\mathbf{X})$ using its quantile function $F^{-1}(\mathbf{q})$, where $\mathbf{q} = (q_1, \dots, q_J)$ is a random vector of length J and $q_j \sim U(0, 1) \forall j$. This yields an array of scenarios $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ (which, when combined into a matrix would yield a scenario set) that can be used to discretize $F_\alpha(\mathbf{w}, \beta)$ by

multiplying each scenario with the equiprobability π_t defined before. The approximation is then

$$\tilde{F}_\alpha(\mathbf{w}, \beta) = \beta + \frac{1}{1-\alpha} \sum_{t=1}^T \pi_t [f_L(\mathbf{w}, \mathbf{X}) - \beta]^+, \quad (2.2.1)$$

which is piecewise linear and convex if and only if the loss distribution $f_L(\mathbf{w}, \mathbf{X})$ is linear w.r.t. \mathbf{w} .

Now, to compute the minimal ES portfolio and the efficiency frontier of the optimization problem, we set $\tilde{F}_\alpha(\mathbf{w}, \beta)$ as the risk function and the return of a portfolio $R(\mathbf{w})$ as reward function. Moreover, we let φ be some arbitrary rate of return and ω some arbitrary level of risk to get the dual problem described by

$$\min_{\mathbf{w}} \tilde{F}_\alpha(\mathbf{w}, \beta) \quad \text{s.t.} \quad R(\mathbf{w}) \geq \varphi \quad (2.2.2)$$

$$\min_{\mathbf{w}} -R(\mathbf{w}) \quad \text{s.t.} \quad \tilde{F}_\alpha(\mathbf{w}, \beta) \leq \omega. \quad (2.2.3)$$

This problem can further be simplified by introducing dummy variables z_t , $t = 1, \dots, T$ instead of the loss distribution in function $\tilde{F}_\alpha(\mathbf{w}, \beta)$ to get the linear function $\beta + (1-\alpha)^{-1} \sum_{t=1}^T \pi_t z_t$, which can replace the risk constraint in Equation 2.2.3 under the set of linear constraints

$$z_t \geq f_L(\mathbf{w}, \mathbf{X}) - \beta, \quad z_t \geq 0 \quad \forall t, \quad \beta \in \mathbb{R}.$$

Additionally, we do not allow short selling and leverage, so the full specification of the portfolio optimization problem is given by the following LP:

$$\begin{aligned} \min_{\mathbf{w}} & -R(\mathbf{w}) \quad \text{s.t.} \quad \omega \geq \beta + (1-\alpha)^{-1} \sum_{t=1}^T \pi_t z_t \\ & z_t \geq f_L(\mathbf{w}, \mathbf{X}) - \beta, \quad \beta \in \mathbb{R} \\ & z_t \geq 0, \quad t = 1, \dots, T \\ & \mu(\mathbf{w}) \geq \varphi \\ & \varphi \geq 0 \\ & w_j \geq 0, \quad j = 1, \dots, J \\ & \sum_{j=1}^J w_j = 1. \end{aligned} \quad (2.2.4)$$

3 EVT marginal distributions and vine copulae

The previous section gave an overview of the literature on portfolio optimization and gave a representation of the stochastic portfolio optimization problem under ES. Key to this optimization approach is the definition of the multivariate probability distribution $F(\mathbf{x}_1, \dots, \mathbf{x}_J) = F(\mathbf{X})$, from which scenarios \mathbf{x}_t or full scenario sets \mathbf{X} can be sampled from. With this in mind, this section is concerned with modeling the multivariate distribution of returns using extreme value theory for the marginal distributions and regular vine copulae to represent the dependence

structure between the marginals and to recover the multivariate distribution as per Sklar's theorem (Sklar, 1959). For the first step, we fit a semi-parametric distribution on the marginals. The Peaks-Over-Threshold (POT) method is used to model the left and right tails of the univariate return distributions and the empirical or kernel distribution for the center. The POT method fits a Generalized Pareto Distribution (GPD) on the observations above a sufficiently high threshold and is supported by one of the fundamental theorems of EVT: the Pickands-Balkema-de Haan theorem. The resulting semi-parametric marginals are used to choose the bivariate copula pairs that serve as building blocks for the regular vine copula from which we ultimately will sample scenarios and scenario sets.

With respect to the literature on multivariate extremes, a first approach to multivariate extreme modeling can be found in Haan and Resnick (1977), who formulate necessary and sufficient conditions for the convergence of maxima to a non-degenerate distribution function. Coles and Tawn (1991) give a first account of a multivariate POT method and suggest logistic and Dirichlet models that can be fitted via Maximum-Likelihood (ML) on higher dimensions. In a similar spirit as ours, Ledford and Tawn (1996) use semi-parametric marginals with GPD tails to arrive at a ML estimator for the bivariate GPD. A d-dimensional expansion of Ledford and Tawn (1996) can be found in Heffernan and Tawn (2004), who review different estimation methods with semi-parametric marginals whose tails are modeled with extreme value distributions other than the GPD. A Bayesian approach to multivariate POT can be found in Sabourin and Renard (2015), who model the dependence structure using a semi-parametric Dirichlet mixture model with GPD marginals and apply it to estimate the joint probability of floods in France. At last, Rootzén et al. (2018) investigate and formalize the properties of the class of multivariate GPD distributions. The main results of this work are the description of multivariate extreme events as a multivariate point process and the proof that the multivariate GPD is threshold stable, meaning that changes in the threshold above which a conditional distribution is fitted lead to the same distribution, up to location changes. Our multivariate approach differs from the existing literature on multivariate extremes in that we do not aim to formulate or use an explicit multivariate distribution function for joint threshold exceedances, but estimate univariate tail distributions first and then use copulae to sample random vectors from them.

3.1 EVT and the Peaks-Over Threshold Method

As we mentioned before, our work uses the univariate POT method instead of other more sophisticated multivariate techniques because of its simplicity and because of its status as "textbook" application for which best practices have been amply developed. The univariate POT has its origins in hydrology, but has found applications in a very wide selection of fields. For example, Davison and Smith (1990) use it in a regression framework to estimate radioactivity concentration in air and give a good account of the properties of the GPD. McNeil (1997) develops what could be seen as the standard procedure for POT implementation at the example of a data set on fire insurance losses. Besides, McNeil and Frey (2000) find an econometric application to model the innovations of a GARCH model with a two-stage procedure that uses EVT and the POT method. A time series application to threshold exceedances can further be found in Smith et al. (1997), who use Markov chains to model extreme values in a time series. In the present

work, we have based our implementation on the best practices outlined in the textbooks by McNeil et al. (2015) and Carmona (2014), who give a good account of "vanilla" applications (such as the found in McNeil (1997)).

We now give an overview of the relevant theorems upon which EVT is founded and the empirical methods derived from these so as to justify our choice of the POT method. This serves as basis to explain how the univariate semi-parametric mixture used for the marginals comes to existence. Our notation, explanations and theorems on EVT are taken from the textbooks by Embrechts et al. (2012) and McNeil et al. (2015); and our semi-parametric approach is entirely based on the procedure suggested in the second chapter of the textbook by Carmona (2014).

3.1.1 Convergence of maxima and the Block Maxima Method

Before going into threshold exceedances, it is necessary to delve into the initial matter of study of EVT, namely, maxima and their convergence in distribution. The most important result of classical EVT—the Fisher-Tippett-Gnedenko theorem—states that maxima converge in distribution towards the generalized extreme value distribution (GEV) and it could be said that this theorem has the same relevance in the study of maxima as the central limit theorem (CLT) has in the study of sums. Accordingly, the GEV would have the same place in the study of extremes as the normal distribution has in central limit theory.

Definition 1. (*Generalized Extreme Value (GEV) distribution*). *The standard GEV is given by*

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & \xi \neq 0 \\ \exp(-e^{-x}), & \xi = 0, \end{cases}$$

where ξ is the shape parameter and $1 + \xi x > 0$. The GEV can be expanded to a three-parameter family by introducing a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\lambda > 0$ by defining $H_{\xi,\mu,\lambda}(x) := H_\xi((x - \mu)/\lambda)$. The value of ξ defines to which type of extreme value distribution does $H_\xi(x)$ belong to: for $\xi > 0$ the distribution is a Fréchet distribution; for $\xi = 0$ it is a Gumbel distribution; and for $\xi < 0$ it is a Weibull distribution.

Now, let x_1, x_2, \dots, x_n be a random sample of n i.i.d. random variables with unknown distribution function $F(x) = P(X \leq x)$. The maximum of the first n observations is denoted by $M_n = \max(x_1, \dots, x_n)$ and is a so-called block maximum. Then, the sequence of normalized maxima $(M_n - d_n)/c_n$, where d_n and c_n are constants with $c_n > 0 \forall n$, converges in distribution to some non-degenerate distribution function $H(x)$ as $n \rightarrow \infty$. Since $P(M_n \leq x) = F(x)$, the convergence of normalized maxima can be expressed as

$$\lim_{n \rightarrow \infty} P\left(\frac{(M_n - d_n)}{c_n} \leq x\right) = \lim_{n \rightarrow \infty} F(c_n x + d_n) = H(x). \quad (3.1.1)$$

The relevance of the GEV as limiting distribution for maxima is summarized by the following definition and the Fisher-Tippett-Gnedenko theorem.

Definition 2. (*Maximum domain of attraction*). *F is said to be in the maximum domain of attraction (MDA) of H , i.e. $F \in MDA(H)$, if the equality in 3.1.1 holds.*

Theorem 1. (Fisher-Tippett-Gnedenko). *If $F \in MDA(H)$ for some non degenerate distribution function H , then H is an extreme value distribution of type $H_{\xi,\mu,\lambda}$.*

This theorem is the support of the first consistent empirical application of EVT by means of the so-called Block Maxima method. This method consists in dividing a sample of random variables x_1, \dots, x_n into m blocks of size k . The maximum of each of the m blocks is taken to form the sequence of independent maxima M_{k1}, \dots, M_{km} onto which a GEV of the form $H_{\xi,\mu,\lambda}$ is fitted. Although this method has found successful uses in areas like hydrology because there are organic criteria to determine block size, it is not well-suited for financial time series. Since it is necessary to make an arbitrary choice of block size, the number of observations onto which the GEV is fitted will be limited *a priori*. This is particularly problematic if data has not been available for a long period of time. If that is the case, the limited amount of blocks might lead to biased estimates of parameters due to the small sample size. For example, if one were to choose yearly maxima in a 20 year data set, only 20 observations would be available to perform induction. Besides, if tail events happen to occur in clusters (as is the case in financial time series), the maximum of the j th block M_{kj} will not be informative if its corresponding block contains clustered extremes because events that actually belong to the right tail of $F(x)$ —but are smaller than M_{kj} —are left out. Conversely, blocks without tail events will report maxima that do not belong to the right tail of $F(x)$, hence making block maxima induction even more biased. One naïve idea to counteract this problem could be to reduce the block size until the desired number of observations is retrieved, however, this is not recommendable from an entropy point of view as the amount of blocks without tail events would increase and the information content of the sample of maxima would thereby decrease as well. In consequence, since our work uses financial time series, a well-suited method for the empirical study of extremes should take all tail events into account, regardless of their frequency and clustering. For this reason, the POT method, which is derived from the study of the limiting behavior of threshold exceedances, is better suited for our purposes.

3.1.2 Convergence of threshold exceedances and the POT method

The study of threshold exceedances is concerned with finding and describing the conditional distribution of excesses (i.e. the peaks) over a sufficiently high threshold u . Analytically, this is given by

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad (3.1.2)$$

with $0 \leq x < x_F - u$, where $x_F \leq \infty$ is the right endpoint of $F(x)$. Intuitively, Equation 3.1.2 gives the probability of observing a threshold exceedance of length $X - u$, given that the threshold is exceeded. A key element of this distribution is the so-called mean excess function (also called residual life function in survival analysis), which gives the expected size of the threshold exceedance as a function of u . It is denoted by

$$e(u) = \mathbb{E}(x - u | x > u), \quad (3.1.3)$$

and its empirical estimator is used to have a visual check for the selection of threshold u , as will be explained in Section 4.2.

The main parametric distributional model used to characterize $F_u(x)$ is the Generalized Pareto distribution (GPD). Moreover, when viewed from the time series, the occurrence of peaks is assumed to follow a Poisson process—but the analysis of this stochastic process is not relevant for our work.

Definition 3. (Generalized Pareto Distribution (GPD)). *The distribution function of the GPD is given by*

$$G_{\xi,\lambda}(x) = \begin{cases} 1 - (1 + \xi x / \lambda)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-x/\lambda), & \xi = 0, \end{cases} \quad (3.1.4)$$

where $\lambda > 0$ and

$$\begin{aligned} x \geq 0, & \quad \text{if } \xi \geq 0, \\ 0 \leq x \leq -\lambda/\xi, & \quad \text{if } \xi < 0. \end{aligned}$$

The domain of the GPD is given by

$$x \in D(\xi, \lambda) = \begin{cases} [0, \infty), & \text{if } \xi \geq 0 \\ [0, -\beta/\xi], & \text{if } \xi < 0. \end{cases}$$

The second moment of the GPD is finite only for $\xi < 1/2$ and, as well as with the GEV, different families can be derived from the GPD: when $\xi > 0$, $G_{\xi,\lambda}(x)$ is an ordinary Pareto distribution; for $\xi = 0$ it is an exponential distribution; and, if $\xi < 0$, it is a short-tailed Pareto type II distribution. The GPD can also be expanded to a three-parameter family by introducing a location parameter $\mu \in \mathbb{R}$ and defining $G_{\xi,\mu,\lambda} := G_{\xi,\lambda}((x - \mu)/\lambda)$. At last, it holds that $G_{\xi,\mu,\lambda} \in MDA(H_\xi) \forall \xi > 0$.

The choice of the GPD as canonical distribution for the analysis of threshold exceedances is by no means arbitrary. Its is supported by the following theorem:

Theorem 2. Pickands-Balkema-de Haan (Pickands III (1975); Balkema and de Haan (1974)). *Under the MDA condition of Equation 3.1.1 it is possible to find a positive-measurable function $\lambda(u)$ such that*

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - G_{\xi,\lambda(u)}(x)| = 0 \iff F \in MDA(H_\xi). \quad (3.1.5)$$

Theorem 2 provides the theoretical grounds to do consistent statistical induction with the Peaks-Over-Threshold method. To understand how the POT method works at the example of financial time series, let $F(x)$ be a univariate distribution of losses such as the one defined in Equation 2.1.3 of Section 2. Furthermore, we assume that the limiting distribution of excesses $F_u(x)$ over some high threshold u is characterized by a GPD, i.e. $F_u(x) = G_{\xi,\lambda(x)}$. Then, given the i.i.d. sequence of losses x_1, \dots, x_n with $x_j \sim F(x) \forall j$, the losses that exceed threshold u are taken to build a separate sample $\bar{x}_1, \dots, \bar{x}_{n_u}$, where n_u is the number of exceedances over

u . Since F_u gives the conditional probability for the size of an exceedance being less than X , it is necessary to compute the size of the excess loss by transforming \bar{x}_j into $y_j = \bar{x}_j - u$. Then, a GPD can be fitted over the resulting n_u exceedances using maximum likelihood (ML), the method of moments (MOM) or probability weighted moments (PWM). We choose ML, one caveat of using it, however, is that the i.i.d. assumption does not hold for financial time series due to, e.g., the clustering of extremes. Nonetheless, the suggestion of McNeil et al. (2015) to overcome this problem is to simply ignore it and classify the estimation as quasi-maximum likelihood (QML).

3.1.3 Semi-parametric mixture distribution

Now, having explained the role of the GPD in EVT and its use to estimate the excess distribution F_u of $F(x)$ in a parametric way, we formulate a univariate semi-parametric distribution that can be fitted to the empirical loss distribution of financial time series. We use the compound probability distribution suggested in Chapter 2 of Carmona (2014) to model $F(x)$. The right and left tails of $F(x)$ are modeled using two separate generalized Pareto distributions, while the center is modeled using the empirical distribution. The lower threshold below which the GPD is fitted to the lower tail is denoted by u^- and the upper threshold, respectively, by u^+ . The analytical form of the compound distribution for $F(x)$ can be summarized as

$$F(x) = \begin{cases} P(X < u^-)P(X \leq x | X < u^-) \\ P(u^- \leq X \leq u^+)P(X \leq x | u^- \leq X \leq u^+) \\ P(X > u^+)P(X \leq x | X > u^+) \end{cases}$$

To arrive at an explicit equation for the compound distribution, we start by finding an expression for its right tail distribution (McNeil et al., 2015, p.154), which is denoted by the survival function $\bar{F}(x) = 1 - F(x)$:

$$\begin{aligned} \bar{F}(x) &= P(X > u^+)P(X > x | X > u^+) \\ &= \bar{F}(u^+)P(X - u > x - u^+ | X > u^+) \\ &= \bar{F}(u^+)\bar{F}_u(x - u^+) \\ &= \bar{F}(u^+)\bar{G}_{\xi^+, u^+, \lambda^+} \\ 1 - F(x) &= \bar{F}(u^+)\left(1 + \xi^+\frac{(x - u^+)}{\lambda^+}\right)^{-1/\xi^+}. \end{aligned} \tag{3.1.6}$$

Note that we have made use of the GPD family with location parameter defined in Definition 3, but substituted μ with the upper threshold u^+ . Besides, a small notation change is introduced to emphasize that ξ^+ and λ^+ are the parameters corresponding to the GPD fitted to exceedances above u^+ . Now, by subtracting one and solving for $F(x)$ we get

$$F(x) = 1 - \bar{F}(u^+)\left(1 + \xi^+\frac{(x - u^+)}{\lambda^+}\right)^{-1/\xi^+}, \quad x > u^+. \tag{3.1.7}$$

We use the estimator n_u^+/n to estimate $\bar{F}(u^+)$ empirically, where n_u^+ is the number of ex-

ceedances above u^+ . The estimator for $\hat{F}(x)$ is denoted by:

$$\hat{F}(x) = 1 - \frac{n_u^+}{n} \left(1 + \hat{\xi}^+ \frac{(x - u^+)}{\hat{\lambda}^+} \right)^{-1/\hat{\xi}^+}, \quad x > u^+. \quad (3.1.8)$$

An estimator for the GPD fitted on the left tail of $F(x)$ can be attained through analogous reasoning. Nevertheless, since the GPD has a non-negative domain and the observations below u^- will be negative, it is necessary to take the absolute value of the sample of exceedances $\tilde{x}_1, \dots, \tilde{x}_{n_u^-}$ and of the threshold u^- itself in order to project the observations onto the line of positive real numbers \mathbb{R}^+ . By letting ξ^- and λ^- be the shape and scale parameters for the GPD fitted below u^- , the estimator for the left tail is

$$\hat{F}(x) = \frac{n_u^-}{n} \left(1 + \hat{\xi}^- \frac{(|x| - |u^-|)}{\hat{\lambda}^-} \right)^{-1/\hat{\xi}^-}, \quad x < u^-, \quad (3.1.9)$$

where n_u^- is the number of observations below threshold u^- .

Finally, the combination of Equations 3.1.8, 3.1.9 and the empirical distribution leads to a semi-parametric mixture with truncated components onto a disjoint support. We denote the distribution function of this model with $\hat{F}^*(x)$. Its estimator is then defined by

$$\hat{F}^*(x) = \begin{cases} \frac{n_u^-}{n} \left(1 + \hat{\xi}^- \frac{(|x| - |u^-|)}{\hat{\lambda}^-} \right)^{-1/\hat{\xi}^-}, & x < u^- \\ \frac{i-0.5}{n}, & x_{(i)} \leq x < x_{(i+1)} \wedge u^- \leq x \leq u^+ \\ 1 - \frac{n_u^+}{n} \left(1 + \hat{\xi}^+ \frac{(x - u^+)}{\hat{\lambda}^+} \right)^{-1/\hat{\xi}^+}, & x > u^+. \end{cases} \quad (3.1.10)$$

Additionally, random samples of $\hat{F}^*(x)$ can be computed using its quantile function $\hat{F}^{*-1}(q)$:

$$\hat{F}^{*-1}(q) = \begin{cases} \left(\frac{\left((q \frac{n}{n_u^-})^{-\hat{\xi}^-} - 1 \right) \hat{\lambda}^-}{\hat{\xi}^-} + |u^-| \right) (-1), & q < q_{u^-} \\ x_{\lfloor qn - 0.5 \rfloor}, & q_{u^-} \leq q \leq q_{u^+} \\ \frac{\left(((1-q) \frac{n}{n_u^+})^{-\hat{\xi}^+} - 1 \right) \hat{\lambda}^+}{\hat{\xi}^+} + u^+, & q > q_{u^+} \end{cases} \quad (3.1.11)$$

where $q_{u^-} = \hat{F}^*(u^-)$ and $q_{u^+} = \hat{F}^*(u^+)$ are the quantiles of the lower and upper thresholds; and $x_{\lfloor qn - 0.5 \rfloor}$ is the $\lfloor qn - 0.5 \rfloor$ -th observation of the empirical CDF. Of course, any non-parametric or parametric model can be used for the center of the distribution (such as kernel density estimation, which we also use in Section 4.5) as long as the non-parametric component is correctly escalated with the unconditional probability $P(u^- \leq X \leq u^+)$.

3.2 A primer on vine copulae

Now that we have derived an explicit estimator for the univariate distributions as semi-parametric compound distributions, the next step is to make the jump to the multivariate world with the help of copulae, especially of a form of copulae called vine copulae. Although copula theory is a well established field, founded by the famous theorem of Sklar (1959), multivariate applications of parametric copulae have lacked and have been mostly available for elliptical copulae (Gaussian and Student's t) and Archimedean copulae (e.g. Gumbel, Frank or Clayton). If we were to limit ourselves to use the available multivariate families, very strong assumptions would need to be made and the full potential of using an EVT-based semi-parametric mixture in the marginals would not be exploited. Using a multivariate Gaussian copula would come at the cost of having no tail dependence, which not only would render our choice of marginals useless, but is dangerous for decision making under uncertainty where extremes do matter. For instance, the use of the Gaussian copula for credit default modeling as developed by Li (2000), went as far as being discussed as one of the causes of the 2008 financial crisis. Although Student's t copula does induce tail dependence, it does it symmetrically. Its use would impose a very strict assumption, namely, that the magnitude of extremes in both tails is perfectly symmetric, which, as the empirical results in Section 4.2 will show, is entirely unrealistic. Archimedean copulae do allow asymmetry in tail dependence—which is what is attempted to induce by modeling the tails with the GPD—but only in either the lower or upper tail, thereby assuming that only one kind of tail dependence is present for all marginals. Its use would disregard the possibility of dependence in both tails and would impose dependence among pairs of random variables that may actually be jointly independent.

Vine copulae were developed in order to overcome the inflexibility and the strong implicit assumptions that arise from the choice of one multivariate family. The main idea behind a vine copula is to decompose a multivariate copula into a cascade of bivariate copulae (each one called pair copulae), which are modeled individually with some parametric bivariate copula. By using graph theory, the pairs can be connected in such a way that they form a fully consistent multivariate distribution. The advantage of this method is that, by modeling the pair copulae separately, it is possible to choose from the wider array of bivariate copulae, where tail dependence can be modeled in more complex ways —e.g., by using the biparameter bivariate or BB class, which includes different parameters for upper and lower tail dependence in the same copula (Joe et al., 1992, chapter 5.2)—, and the parsimony of the model can be improved by modeling some copula pairs with the independence copula.

The decomposition of a copula into a sequential cascade of conditional bivariate copulae was first articulated by Joe (1996), who showed that a d -dimensional copula can be decomposed into bivariate pairs with $d(d - 1)/2$ parameters and gave a formulation in terms of distribution functions. The use of this decomposition to form vine distributions as graphical models was then developed by Bedford and Cooke (2001) and Bedford and Cooke (2002), who define the three main classes of vine copulae, explain how to use them for simulation and give a formulation of vines in terms of densities. A survey of financial applications for vine copulae (including scenario generation for ES optimization) can be found in Aas (2016). More generally, Czado and Nagler (2022) give a comprehensive summary and review of how vine copulae are built,

how they are selected and to which kind of problems they can be applied to. The choice of vine structures and parametric forms of pair copulae is not trivial and has a literature strain of its own, but these issues will be discussed in Section 4.4.

3.2.1 Basic properties of copulae

Before we explain how a vine copula is built, it is necessary to recall the definition and most relevant properties of copulae in general and why they are suited for multivariate statistical modeling. For this section, we base our explanations, notation and proofs on chapter 7 of McNeil et al. (2015). First, we start by defining a copula:

Definition 4. (Copula). A copula $C(\mathbf{u}) = C(u_1, \dots, u_d)$ is a d -dimensional distribution function on the unit hypercube $[0, 1]^d$ with standard uniform marginal distributions.

Thus, C is a mapping of the unit hypercube to the unit interval $C : [0, 1]^d \rightarrow [0, 1]$. The copula density is obtained through partial differentiation and is given by

$$c(u_1, \dots, u_d) := \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1, \dots, \partial u_d} \quad \forall \mathbf{u} \in [0, 1]^d.$$

The following theorem summarizes why all multivariate distributions contain copulae and why multivariate distributions can be represented through them.

Theorem 3. (Sklar's theorem (Sklar, 1959)). Let $\mathbf{x}_d = (x_1, \dots, x_d)$ be a random vector with joint distribution function F and marginal distributions F_1, \dots, F_d . Then, $\exists C : [0, 1]^d \rightarrow [0, 1]$ such that $\forall x_1, \dots, x_d \in \mathbb{R} = [-\infty, \infty]$,

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \tag{3.2.1}$$

with density

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d) \tag{3.2.2}$$

The proof of Theorem 3 is remarkably simple and we deem its inclusion helpful to aid further understanding of the scenario generation methods that will be explained in Section 4.3.

Proof. Let $\mathbf{x}_d = (x_1, \dots, x_d)$ be a random vector with joint distribution function F , continuous marginal distributions F_1, \dots, F_d and quantile functions $F_1^{-1}, \dots, F_d^{-1}$ with $i = 1, \dots, d$. Since F_i is bijective, strictly increasing and a map of the form $F : [-\infty, \infty] \rightarrow [0, 1]$, it is trivially true that $F(x_i) = u_i$, where $U_i \sim U(0, 1) \quad \forall i$, and that $F_i^{-1}(u_i) = x_i$ almost surely. Let C be a copula according to Definition 4. Then, the multivariate joint distribution function F computes the following probability:

$$\begin{aligned} F(x_1, \dots, x_d) &= P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= P(F_1^{-1}(U_1) \leq x_1, \dots, F_d^{-1}(U_d) \leq x_d) \\ &= P(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) \\ &= C(F_1(x_1), \dots, F_d(x_d)), \end{aligned} \tag{3.2.3}$$

which is the identity of Theorem 3. Evaluating 3 at $x_i = F^{-1}(u_i), 0 \leq u_i \leq 1, i = 1, \dots, d$ yields

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \quad (3.2.4)$$

which represents copula C in terms of F the marginals F_i , thereby proving uniqueness. \square

In conclusion, by choosing some univariate marginal distributions F_1, \dots, F_d , there is a unique distribution function given by copula C which represents the joint multivariate distribution of the random vector x . This result justifies the procedure followed in our work, where we model the univariate marginals separately and Theorem 3 allows us to bring them together into a multivariate expression via copulae.

3.3 Pair copula decompositions

Our explanations and examples on how a copula can be decomposed into bivariate pair copulae and how these are used to build vine distributions are entirely based on Aas et al. (2009) and, especially, on Czado and Nagler (2022). Following Theorem 3 and assuming $d = 2$ for simplicity, it is possible to express the conditional distribution function $F_{1|2}$ and the conditional density $f_{1|2}$ with

$$F_{1|2}(x_1|x_2) = \frac{\partial C_{12}(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)} = \frac{\partial C_{12}(F_1(x_1), v)}{\partial v} \Big|_{v=F_2(x_2)}, \quad (3.3.1)$$

$$f_{1|2}(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1), \quad (3.3.2)$$

where $C_{12}(\cdot, \cdot)$ and $c_{12}(\cdot, \cdot)$ are the bivariate copula distribution and density functions corresponding to $F_1(x_1)$ and $F_2(x_2)$. Overall, a univariate conditional density $f_{(d-1)|d}(x_{d-1}|x_d)$ can be expressed as a function of the bivariate copula $c_{(d-1)d}$ and a marginal density $f_{d-1}(x_{d-1})$:

$$f_{(d-1)|d}(x_{d-1}|x_d) = c_{(d-1)d}(F_{d-1}(x_{d-1}), F_d(x_d))f_{d-1}(x_{d-1}) \quad (3.3.3)$$

To illustrate how the decomposition works in a higher dimension, we set $d = 3$, denote the conditional distribution and the conditional density of x_1 given \mathbf{x}_D with $F_{j|D}$ and $f_{j|D}$ respectively; and call the conditional bivariate copula $c_{ij,k}(\cdot, \cdot; x_k)$ a pair copula. The density $f(x_1, x_2, x_3)$ can be factorized into

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f_1(x_1). \quad (3.3.4)$$

The first element to be decomposed is $f_{3|12}(x_3|x_1, x_2)$, which can be done using the bivariate conditional density $f_{13|2}(x_1, x_3|x_2)$. Denoting the copula density by $c_{13;2}(\cdot, \cdot; x_2)$, it is possible to use Equation 3.2.2 to get

$$f_{13|2}(x_1, x_3|x_2) = c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)f_{1|2}(x_1|x_2)f_{3|2}(x_3|x_2). \quad (3.3.5)$$

Applying Equation 3.3.3 to 3.3.5 yields an expression for the univariate conditional density

$f_{3|12}(x_1|x_2, x_3)$:

$$f_{3|12}(x_1|x_2, x_3) = c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) f_{3|2}(x_3|x_2). \quad (3.3.6)$$

The last elements to be decomposed are $f_{2|1}(x_2|x_1)$ and $f_{3|2}(x_3|x_2)$, which can be easily determined using Equation 3.3.3 again:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2), \quad (3.3.7)$$

$$f_{3|2}(x_3|x_2) = c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3) \quad (3.3.8)$$

Finally, putting all together back to Equation 3.3.4 results into

$$\begin{aligned} f(x_1, x_2, x_3) &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) \times c_{23}(F_2(x_2), F_3(x_3)) \\ &\quad \times c_{12}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1), \end{aligned} \quad (3.3.9)$$

which expresses the joint density in terms of bivariate copulae, conditional distribution functions and marginal densities.

One drawback of this decomposition, nonetheless, is that it is not unique and makes it difficult to derive a general factorization of $f(x_1, \dots, x_d)$. For instance, if we would have started by decomposing $f_{2,3|1}$ or $f_{1,2|3}$, two different decompositions would have been attained. The problem of choice of decomposition can be solved by minimizing some information criterion on the log-likelihood of the decomposition and is the core problem of vine copula selection. A detailed explanation on how the log-likelihood function for a decomposition is built can be found in Section 2 of Czado and Nagler (2022).

3.3.1 Regular vine copulae

Following the contributions of Bedford and Cooke (2001, 2002), it is possible to organize all possible factorizations of some joint density $f(x_1, \dots, x_d)$ with a graphical tool called *regular vine* (or R-vine for simplicity). A vine is a graphical structure consisting of nodes, edges and trees, where the edges in one tree sequentially become the nodes of the next one. The R-vine is the most general and flexible model. The relevant subclasses of R-vines are the drawable (D-) vine and the canonical (C-) vine. The following two definitions characterize an R-vine (Czado and Nagler, 2022, p.461).

Definition 5. (Tree). A tree is a connected acyclic graph $T = (N, E)$ with node set N and edge set E .

Definition 6. (Regular vine tree). A set of graphs $\mathcal{V} = (T_1, \dots, T_{d-1})$ is called a d -dimensional R-vine tree sequence if

1. T_1 is a tree with node set $N_1 = 1, \dots, d$ and edge set E_1 .
2. For $j \geq 2$, T_j is a tree with node set $N_j = E_{j-1}$ and edge set E_j .
3. For $j = 1, \dots, d-1$ and $\{a, b\} \in E_j$ it must hold that $|a \cap b|=1$.

The last condition, also called proximity condition, ensures that an edge between two nodes in tree T_j is only feasible if the corresponding edges in T_{j-1} share a common node. To fully characterize an R-vine distribution, a notation for the edges is necessary. In that spirit, the notation for any edge $e \in E_i$ is defined by the complete union

$$A_e = \{j \in N_1 | \exists e_1 \in E_1, \dots, e_{i-1} \in E_{i-1} \text{ s.t. } j \in e_1 \in \dots \in e_{i-1} \in e\}.$$

The conditioning set D_e of an edge $e = \{a, b\}$ is denoted as $D_e := A_a \cap A_b$ and the conditioned sets $\mathcal{C}_{e,a}$ and $\mathcal{C}_{e,b}$ are given by

$$\mathcal{C}_{e,a} = A_a \setminus D_e \text{ and } \mathcal{C}_{e,b} = A_b \setminus D_e.$$

Since $\mathcal{C}_{e,a}$ and $\mathcal{C}_{e,b}$ are singletons (Kurowicka and Cooke, 2010), the edge e can be abbreviated as $e = (C_{e,a}, C_{e,b}; D_e)$. To exemplify how the notation works, we use the R-vine tree structures of Figure 1, which represent Equation 3.3.9 on the left and, on the right, the four dimensional decomposition given by

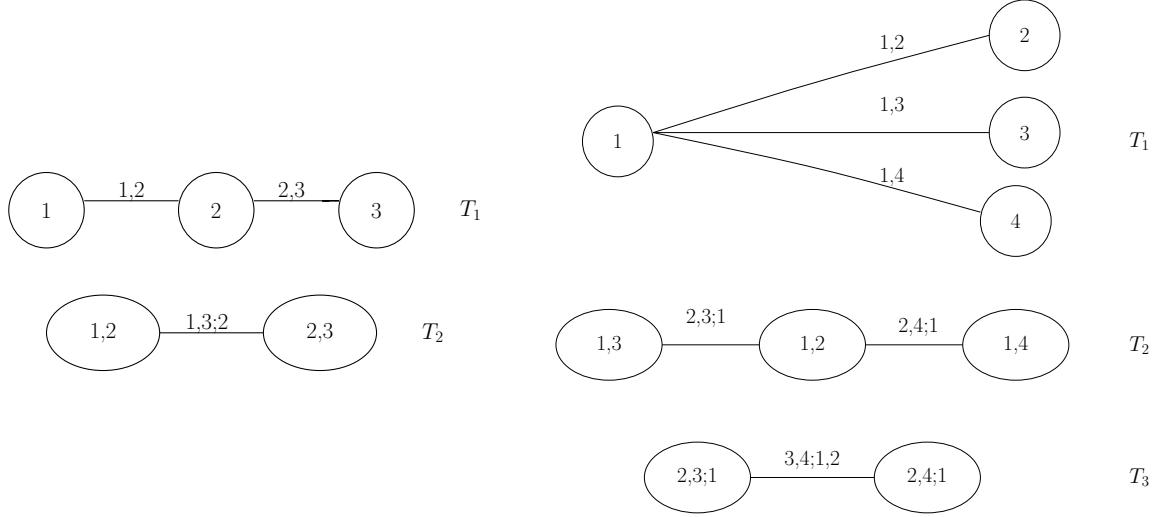
$$\begin{aligned} f(x_1, x_2, x_3, x_4) = & c_{34;12}(F_{3|12}(x_3|x_1, x_2), F_{4|12}(x_4|x_1, x_2); x_1, x_2) \\ & \times c_{23;1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1); x_1) \\ & \times c_{24;1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4, x_1); x_1) \\ & \times c_{14}(F_1(x_1), F_4(x_4)) \times c_{13}(F_1(x_1), F_3(x_3)) \times c_{12}(F_1(x_1), F_2(x_2)) \\ & \times f_4(x_4)f_3(x_3)f_2(x_2)f_1(x_1). \end{aligned} \tag{3.3.10}$$

The most simple example could be to form the only edge of T_2 in the left R-vine. We start by defining the edges of T_1 (which are the nodes of T_2) as $A_a = \{1, 2\}$ and $A_b = \{2, 3\}$. The conditioning set is then $D_e = \{2\}$, such that $C_{e,a} = \{1\}$ and $C_{e,b} = \{3\}$. Plugging into e yields $e = (1, 3; 2)$. For the R-vine with $d = 4$, the edge of T_3 is formed with $A_a = \{2, 3; 1\}$ and $A_b = \{2, 4; 1\}$. Then, $D_e = \{2, 1\}$, $C_{e,a} = \{3\}$ and $C_{e,b} = \{4\}$. Plugging the conditioning set D_e into e yields then $e = (3, 4; 1, 2)$.

The only open question now is how to translate an R-vine structure into distribution and density functions. The R-vine distribution of some random vector $\mathbf{x} = (x_1, \dots, x_d)$ is determined by the triplet $(F, \mathcal{V}, \mathcal{B})$. F is a vector of continuous marginal distributions $F = (F_1, \dots, F_d)$ corresponding to the univariate distributions of the random vector \mathbf{x} . \mathcal{V} is a d -dimensional R-vine tree sequence. \mathcal{B} is a set $\mathcal{B} = \{C_e | e \in E_i; i = 1, \dots, d-1\}$, where C_e is a bivariate copula, E_i is the edge set of tree T_i in some R-vine tree sequence \mathcal{V} and $C_e(\cdot, \cdot)$ is the copula of the conditional distribution $F_{C_{e,a}C_{e,b}|D_e}(x_{C_{e,a}}, x_{C_{e,b}}|X_{D_e})$. The last conditional distribution can be expressed as copula through Theorem 3, such that the copula C_e of edge e is denoted by $C_{C_{e,a}C_{e,b};D_e}$. If all marginal distributions are standard uniform, the following theorem supports the existence of an R-vine distribution.

Theorem 4. (Existence of an R-vine distribution (Czado and Nagler, 2022, p.464)). Let $(F, \mathcal{V}, \mathcal{B})$ be a satisfactory vine specification and $F_i(x_i), i = 1, \dots, d$ be the margins of F .

Figure 1: R-vine structures for $d = 3$ and $d = 4$



This figure depicts the R-vine structures of Equations 3.3.9 and 3.3.10. The R-vine on the left is a D-vine as well because no node is connected to more than two edges. Source: (Czado and Nagler, 2022, p.415)

Then, there exists a d -dimensional distribution function F with density

$$f_{1,\dots,d}(x_1, \dots, x_d) = f_1(x_1) \times \dots \times f_d(x_d) \\ \times \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{C_{e,a}, C_{e,b}; D_e} (F_{C_{e,a}|D_e}(x_{C_{e,a}} | \mathbf{x}_{D_e}), F_{C_{e,b}|D_e}(x_{C_{e,b}} | \mathbf{x}_{D_e})). \quad (3.3.11)$$

The distribution function of $x_{C_{e,a}}$ and $x_{C_{e,b}}$ conditional on D_e is given by

$$F_{C_{e,a}C_{e,b}|D_e}(x_{C_{e,a}}, x_{C_{e,b}} | \mathbf{x}_{D_e}) = C_e(F_{C_{e,a}|D_e}(x_{C_{e,a}} | D_e), F_{C_{e,b}|D_e}(x_{C_{e,b}} | D_e)), \quad (3.3.12)$$

for all $e \in E_i, i = 1, \dots, d-1$ and $e = \{a, b\}$.

Random d -dimensional vectors can be drawn from an R-vine distribution using sequential random sampling. We limit ourselves to a simple example, as understanding how random vectors are sampled from an R-vine is key for the simulation of scenario sets in Section 4. A more detailed explanation can be found in Section 5 of Bedford and Cooke (2001). Suppose we want to draw a random vector $\mathbf{x} = (x_1, x_2, x_3)$ from the R-vine on the left of Figure 1. The first step is to sample x_1 from its distribution function F_1 (e.g., using its inverse transform). Then, since $F_{2|1}$ can be recovered through C_{12} , it is possible to sample x_2 from C_{12} using the already sampled x_1 . Next, $C_{13|2}$ (which is the edge of tree T_2) recovers the joint distribution of $F_{1|2}$ and $F_{3|2}$, who jointly determine $F_{3|12}$. By conditioning on x_1 and x_2 , x_3 can be sampled from the pair copula $C_{13|2}$, yielding the desired random vector. The procedure is analogue for higher dimensions and always involves sequential sampling from pair-copulae.

4 Case study: a multi-asset class portfolio

The preceding sections have outlined how stochastic portfolio optimization with ES works and how EVT and copula theory can be used to generate the scenarios that serve as input for the stochastic program. Now, this section knits all elements together and implements our statistical modeling technique into a portfolio optimization framework using time series data on financial assets. The asset universe comprises thirteen assets belonging to five different asset classes proxying the portfolio selection problem of a pension fund. Mimicking a pension fund is especially interesting in our context because neither the correlations, nor the dependence structure are evident *ex ante*, which would be a drawback of limiting the universe to a single asset class. Our asset universe makes it possible to test how optimization under ES mitigates extreme events when complex dependence structures (and particularly tail dependence) are present; and to test the flexibility and accuracy of vine copulae methods to capture the subtle dependence structures that escape the eye of correlation.

4.1 Data

Our data set consists of time series of returns for thirteen financial assets. The list of assets used for the portfolio selection problem is found on Table 1, which reports their name, ticker and asset class. The data set covers five asset classes: Equities, Fixed Income, Commodities, Alternative Investments and Real Estate. Nine out of thirteen components are ETFs, two are direct quotes on the yield of government bonds and one is an index that tracks the performance of so-called Commodity Trading Advisors (CTAs). The price time series for government bonds are built by computing the price of a zero-coupon bond with the yield quoted on a given day. We rely heavily on ETFs to reduce dimensionality because ETFs are tradable (thus providing actual price data) and reflect the behavior of a whole market segment without needing further assumptions (as would be the case when using other dimension reduction techniques, such as Principal Component Analysis). Nonetheless, one drawback of ETFs is that many of them are a relatively recent product and ETFs on certain market segments have been traded with a reasonable degree of liquidity for less than two decades. For this reason, our sample period is limited from January 2011 to May 2023, yielding 2929 observations in total. And, although this period was characterized by the effects of the Great Moderation and Quantitative Easing, it still contains the crash due to the COVID-19 pandemic in 2020 and the first year of the interest rate hike cycle that started in 2022. The presence of different market regimes is fundamental to make a robust extrapolation of our results into a real-world decision making framework.

Table 2 contains descriptive statistics for all the thirteen constituents of the asset universe. 9 out of the 13 assets show negative skewness and 11 out of 13 are leptokurtic. Although the German and US bonds seem to be mesokurtic and symmetric, assuming a normal distribution for them would be dangerous as their prices were greatly influenced by the monetary and fiscal policy carried out during the sample period. To address the question of normality, we compute the test statistic of the Shapiro-Wilk test for normality, where the null hypothesis states that the data points are normally distributed, and see that the null is rejected for all assets at the 1% level. The daily VaR and the daily ES are computed at the 95% level, where the riskiest

Table 1: List of assets

| Asset | Mnemonic | Asset Class |
|--|----------|-------------------------|
| 1 iShares Core S&P 500 ETF | IVV | Equities |
| 2 iShares MSCI Eurozone ETF | EZU | Equities |
| 3 iShares JPX-Nikkei 400 ETF | JPXN | Equities |
| 4 iShares MSCI Pacific ex Japan ETF | EPP | Equities |
| 5 Lyxor MSCI Eastern Europe ex Russia UCITS ETF Acc | CEC.PA | Equities |
| 6 iShares MSCI Emerging Markets ETF | EEM | Equities |
| 7 Vanguard Intermediate-Term Corporate Bond Index Fund | VCIT | Fixed Income |
| 8 iShares J.P. Morgan USD Emerging Markets Bond ETF | EMB | Fixed Income |
| 9 VanEck Global Real Estate UCITS ETF | TRET.L | Real Estate |
| 10 iShares S&P GSCI Commodity-Indexed Trust | GSG | Commodities |
| 11 Germany Bund 10 Year | DE10Y | Fixed Income |
| 12 U.S. 10 Year Treasury | USFI10Y | Fixed Income |
| 13 Société Générale CTA Index | NEIXCTAT | Alternative Investments |

This table presents the list of assets used for the empirical implementation. The left column reports the commercial name of the asset or index, the middle column reports a mnemonic abbreviation or ticker and the third column reports the asset class to which each component belongs to. The data was retrieved from Yahoo Finance (assets 1 to 10) and Bloomberg (assets 11 to 12 and index 13). The sample period starts in June 2011 and ends in May 2023.

asset according to these risk measures is EZU (VaR of 2.33% and ES of 3.77%), followed by CEC.PA (VaR of 2.12% and ES of 3.45%) and EEM (VaR of 2.17% and ES of 3.36%). On top of that, Table 3 gives a first overview of the dependence structure by reporting sample ordinal correlations for the daily log returns of all assets. We choose Kendall’s τ correlation coefficient as it is a copula-based measure of dependence and is thereby more related to our work than the standard Pearson’s correlation, which only measures linear dependence. The highest correlation is between EEM and EPP (0.64) while the lowest is between IVV and USFI10Y (-0.21).

4.2 Fitting of the marginal distributions

The first step to fit the semi-parametric distribution (SPD for simplicity) defined in Section 3.1.2 is to choose the lower and upper thresholds that mark the beginning of the parametric part of the distribution. The choice of threshold for the POT method is an open question in EVT and there is no first-best approach that has managed to definitely overcome the bias-variance trade-off. Overall, the higher the threshold, the better the parameter estimation, but the higher the variance and the uncertainty around it. For this reason, successful threshold selection should rather aim at parameter stability instead of bias reduction. The usual methodologies range from pragmatic visual checks to minimization of various norms. Lang et al. (1999) do a first review of three threshold selection techniques at the example of hydrological applications: mean number of observations above a threshold, mean excess above a threshold (which is the one we use as diagnostic tool for our automatic selection criteria) and Poisson model fitting, where the optimal threshold is that for which the frequency in the occurrence of peaks best resembles a Poisson process. Solari et al. (2017) choose the threshold that minimizes the complement of the p-value of the Anderson-Darling goodness of fit test. A more recent development by

Table 2: Descriptive statistics

| Asset | CAGR (in %) | Ann. SD (in %) | Skewness | Kurtosis | 0.95 VaR (in %) | 0.95 ES in (%) | Shapiro-Wilk |
|----------|-------------|----------------|----------|----------|-----------------|----------------|--------------|
| IVV | 12.92 | 18.56 | -0.73 | 13.20 | 1.79 | 2.89 | 0.88*** |
| EZU | 4.62 | 24.02 | -0.96 | 9.77 | 2.33 | 3.77 | 0.91*** |
| JPXN | 5.27 | 18.39 | -0.60 | 6.66 | 1.72 | 2.76 | 0.94*** |
| EPP | 3.77 | 21.24 | -0.77 | 13.80 | 1.91 | 3.16 | 0.89*** |
| CEC.PA | -1.85 | 22.80 | -1.03 | 12.06 | 2.12 | 3.45 | 0.92*** |
| EEM | 0.55 | 22.74 | -0.63 | 7.02 | 2.17 | 3.36 | 0.94*** |
| VCIT | 3.38 | 5.94 | 0.10 | 30.92 | 0.52 | 0.85 | 0.84*** |
| EMB | 2.73 | 9.90 | -2.50 | 42.21 | 0.81 | 1.47 | 0.79*** |
| TRET.L | 4.30 | 17.89 | 0.12 | 16.29 | 1.60 | 2.70 | 0.88*** |
| GSG | -5.32 | 22.30 | -0.86 | 7.55 | 2.21 | 3.38 | 0.94*** |
| DE10Y | 0.60 | 7.48 | 0.04 | 3.14 | 0.75 | 1.10 | 0.96*** |
| USFI10Y | -0.43 | 8.47 | 0.01 | 2.83 | 0.85 | 1.17 | 0.98*** |
| NEIXCTAT | 3.83 | 11.53 | -0.80 | 4.03 | 1.16 | 1.74 | 0.96*** |

This table presents the descriptive statistics for the daily returns of twelve financial assets and one index sourced from Yahoo Finance and Bloomberg. The sample period starts in June 2011 and ends in May 2023, yielding 2929 observations. The reported statistics are the compound annual growth rate (CAGR), annualized standard deviation (Ann. SD), skewness, kurtosis, Value-at-Risk (VaR) at the 95% confidence level, Expected Shortfall (ES) at the 95% confidence level and the test statistic for the Shapiro-Wilk test of normality, where the null hypothesis is that the data points are normally distributed. ***, * and * indicate statistical significance at the 1%, 5% and 10% level respectively.

Kiran and Srinivas (2021) maps the GPD random variable to a standard exponential random variable. Then, the optimal threshold is that which minimizes the Mahalanobis distance between the L-moments of the transformed random variable and those of the population. In the financial realm, Chavez-Demoulin et al. (2014) make a non-parametric, Bayesian expansion of the POT method which is contrasted against the classical POT estimation. For the latter, they suggest that a good initial threshold guess is that which recovers 10% of the observations. Our work follows this heuristic criterion and leaves the door open to more sophisticated improvements.

4.2.1 Threshold selection

There is no gold-standard approach to threshold selection and any rule for automatic threshold selection will eventually fail, especially in the realm of financial time series. Since threshold selection is a whole research field on its own, we resort to a pragmatic approach. Our first rule is based on Carmona (2014), who selects the upper threshold u^+ and the lower threshold u^- such that 150 observations are available beyond each cutoff value. For our sample, this amounts to approximately choosing the 5th and 95th quantiles as cutoffs. Our second approach is to fit the GPD below the 10th and above the 90th quantile of the empirical distribution, which is the approach followed by Chavez-Demoulin et al. (2014), and results in 293 exceedances on either side of the distribution. The best practices to evaluate and diagnose goodness of fit suggested by McNeil et al. (2015) are nonetheless used.

We start by inverting the sign of the return distributions to get a loss distribution as defined in Equation 2.1.3 and in Section 3.1.3. A visual check of the normality of the loss distributions can be found in figure A.7 of the Appendix, where it is possible to see that, without exception,

Table 3: Sample ordinal correlations

| | IVV | EZU | JPXN | EPP | CEC.PA | EEM | VCIT | EMB | TRET.L | GSG | DE10Y | USFI10Y | NEIXCTAT |
|----------|-------|-------|-------|-------|--------|-------|-------|-------|--------|-------|-------|---------|----------|
| IVV | 1.00 | | | | | | | | | | | | |
| EZU | 0.60 | 1.00 | | | | | | | | | | | |
| JPXN | 0.48 | 0.47 | 1.00 | | | | | | | | | | |
| EPP | 0.57 | 0.58 | 0.46 | 1.00 | | | | | | | | | |
| CEC.PA | 0.29 | 0.36 | 0.28 | 0.33 | 1.00 | | | | | | | | |
| EEM | 0.54 | 0.56 | 0.45 | 0.64 | 0.36 | 1.00 | | | | | | | |
| VCIT | 0.01 | 0.02 | 0.03 | 0.03 | -0.01 | 0.04 | 1.00 | | | | | | |
| EMB | 0.30 | 0.32 | 0.26 | 0.32 | 0.22 | 0.35 | 0.33 | 1.00 | | | | | |
| TRET.L | 0.30 | 0.24 | 0.28 | 0.29 | 0.24 | 0.25 | 0.10 | 0.23 | 1.00 | | | | |
| GSG | 0.25 | 0.25 | 0.19 | 0.29 | 0.16 | 0.27 | -0.02 | 0.16 | 0.09 | 1.00 | | | |
| DE10Y | -0.13 | -0.18 | -0.09 | -0.13 | -0.16 | -0.13 | 0.39 | 0.12 | 0.02 | -0.11 | 1.00 | | |
| USFI10Y | -0.21 | -0.19 | -0.14 | -0.17 | -0.14 | -0.16 | 0.60 | 0.16 | -0.02 | -0.14 | 0.47 | 1.00 | |
| NEIXCTAT | 0.09 | 0.05 | 0.06 | 0.06 | 0.04 | 0.07 | 0.01 | -0.01 | 0.09 | 0.02 | 0.06 | 0.02 | 1.00 |

This table presents sample unconditional Kendall's τ correlation coefficients between daily log returns for twelve assets and one index. The sample period starts in June 2011 and ends in May 2023, yielding 2929 observations.

the right tail (i.e. losses) is especially heavy with respect to the left one. The first diagnostic tool to be used is the sample mean excess plot (MEP), which is the empirical estimation of the mean excess function defined in Equation 3.1.3. This plot shows the average size of exceedances for a sequence of positive loss data x_1, \dots, x_n as a function of the threshold u . Its empirical estimator is given by the sample mean excess function:

$$e_n(u) = \frac{\sum_{i=1}^n (x_i - u) I_{\{x_i > u\}}}{\sum_{i=1}^n I_{\{x_i > u\}}}. \quad (4.2.1)$$

The mean excess plot is constructed as $\{(x_{i,n}, e_n(x_{i,n})) : 2 \leq i \leq n\}$, where $x_{i,n}$ is the descending i th order statistic. One of the properties of the mean excess function is that it should become increasingly linear for higher values of u (McNeil et al., 2015, p.150-151), implying that a good candidate for u should be chosen around the section of the plot where the mean excesses start to be linear. If the linear trend goes upwards, the fit of the GPD will yield $\xi > 0$; if the trend is horizontal, the shape parameter will be $\xi \approx 0$, suggesting an exponential excess distribution; and, for a negative linear trend, the shape parameter will be negative $\xi < 0$, indicating a thin tailed excess distribution.

Figures A.8 and A.9 show the mean excess plots for the upper and lower threshold exceedances of the thirteen loss distributions. The solid line is the threshold value u beyond which 150 observations are available and the dashed line to the threshold value u corresponding to the 90th quantile of the loss distribution (10th quantile for lower threshold exceedances). The plots for upper threshold exceedances give a very good general depiction of the expected behavior of a MEP. All assets, with the exception of DE10Y, are an increasing function of u and do start to straighten up at some point. The fact that DE10Y is downward sloping may be concerning, yet this just indicates that the resulting GPD will be thin tailed, which was already implied in table 2. Our threshold choices seem to be acceptable in almost all cases,

especially when using the $n_u = 150$ rule. There are assets where it could be set marginally higher, nonetheless. The most notable cases are EZU, JPXN, EMB and USF10Y. The erratic behavior of the last observations should not be a cause for concern as this is a consequence of the very few observations recovered for too high values of u . In our opinion, this justifies the preemptive selection of a threshold for which enough observations are available so as to reduce variance at the cost of some bias in the parameter estimation, which is what is attempted with the two rules used. The behavior of the MEPs is not as clear cut for lower threshold exceedances, however. Only IVV, EPP, CEC.PA, VCIT and TRET.L exhibit a distinct upward slope, while the remaining assets show either a rather horizontal slope or erratic fluctuations after some point. Although the threshold selection criteria are acceptable for the stabler MEPs, the remaining MEPs are not ideal for GPD estimation. Yet, since the focal point of the thesis is the avoidance of extreme losses, parameter uncertainty for the tail distribution of gains should not have a significant impact on optimization under a measure of tail-risk, such as ES. Even so, one of the sensitivity checks we make in Section 4.5 tries to address this problem, where only the right tail of the loss distribution is modeled with the GPD.

Our threshold selection criteria are more concerned with finding an overall good heuristic criterion that yields sufficiently stable parameters. To check whether our criteria achieve this, figures A.10 and A.11 present the shape plots of the right and left tails of all studied assets. These plots show the estimated shape parameter ξ of the GPD as a function $\xi(u)$ of threshold values u . Ideally, the optimal threshold should be in a region of the plot where ξ is relatively stable w.r.t. changes in u . Due to the bias-variance trade-off, the instability of $\xi(u)$ should be increasing in u and, overall, that is the observed behavior in the plots for both right and left tails. As for the choice of threshold, our pragmatic criteria seem to work remarkably well in the selection of u for the right tail. The only exception where it does not seem appropriate is for DE10Y, where the instability of the parameter is global and the 95% confidence interval could not be reported due to problems in the computation of the covariance matrix. Regarding the lower tail (i.e. profits), our criteria have an acceptable performance again, but there is more instability in general and more values of ξ are nearer to zero than in the right tails (implying a rather exponential decay for profits). Additionally, a confidence interval could not be reported for VCIT, USFI10Y and NEIXCTAT. This instability was present in the MEPs as well, hinting that there is an inherent difficulty of using the POT method for fixed income assets. To sum up, both rules select thresholds that tend to be in stable regions of the shape plot. Since, per definition, the $n_u = 150$ rule will select higher thresholds than the 10th/90th quantile rule, and higher thresholds reduce parameter bias, the $n_u = 150$ rule is more effective than the 10th/90th rule for most of the analyzed time series.

Tables 4 and 5 contain the estimated parameters for the GPD tails of the SPD defined in Equation 3.1.10. The upper and lower threshold values, u^+ and u^- , are the location parameters of their respective GPD and are interpreted as the daily log return beyond which an observation is classified as tail event. The estimates for the shape parameter ξ tend to be higher using the $n_u = 150$ rule (8 out of 13 cases), while there is no definitive trend regarding the scale parameter λ (6 out of 13 cases). The shape parameters are, without exception, smaller than 1/2, meaning that the second moment of all the estimated GPD tails is finite. The shape

Table 4: Estimated parameters - 150 observations beyond u

| Criterion: 150 observations beyond u | | | | | | |
|--|------------------|-------------------|--------|------------------|-------------------|-------|
| Mnemonic | $\hat{\xi}^-$ | $\hat{\lambda}^-$ | u^- | $\hat{\xi}^+$ | $\hat{\lambda}^+$ | u^+ |
| IVV | 0.353 (0.112) | 0.006 (0.001) | -0.016 | 0.233 (0.098) | 0.008 (0.001) | 0.018 |
| EZU | 0.075 (0.102) | 0.010 (0.001) | -0.021 | 0.203 (0.088) | 0.011 (0.001) | 0.023 |
| JPXN | 0.059 (0.099) | 0.008 (0.001) | -0.016 | 0.157 (0.090) | 0.009 (0.001) | 0.017 |
| EPP | 0.344 (0.114) | 0.007 (0.001) | -0.018 | 0.304 (0.100) | 0.009 (0.001) | 0.019 |
| CEC.PA | 0.100 (0.082) | 0.008 (0.001) | -0.021 | 0.268 (0.101) | 0.009 (0.001) | 0.021 |
| EEM | 0.219 (0.106) | 0.007 (0.001) | -0.021 | 0.241 (0.096) | 0.009 (0.001) | 0.021 |
| VCIT | 0.358 (NA) | 0.002 (NA) | -0.005 | 0.308 (0.090) | 0.002 (0.000) | 0.005 |
| EMB | 0.099 (0.094) | 0.005 (0.001) | -0.007 | 0.362 (0.103) | 0.004 (0.000) | 0.008 |
| TRET.L | 0.488 (0.122) | 0.005 (0.001) | -0.015 | 0.161 (0.091) | 0.009 (0.001) | 0.016 |
| GSG | 0.057 (0.094) | 0.008 (0.001) | -0.021 | 0.259 (0.096) | 0.008 (0.001) | 0.022 |
| DE10Y | 0.009 (0.085) | 0.004 (0.000) | -0.007 | -0.165 (NA) | 0.004 (NA) | 0.007 |
| USFI10Y | 0.045 (0.090) | 0.004 (0.000) | -0.008 | 0.137 (0.085) | 0.003 (0.000) | 0.008 |
| NEIXCTAT | 0.000 (0.084) | 0.003 (0.000) | -0.011 | 0.204 (0.100) | 0.005 (0.001) | 0.011 |

This table presents the estimated shape, scale and location (i.e. threshold) parameters for the upper and lower GPD tails of the semi-parametric distribution rounded to three decimal places. Standard errors are reported below parameter estimates (so far as available). The selected thresholds for upper and lower exceedances were those beyond which 150 observations were available.

parameter for the upper tail (i.e. for losses) tends to be bigger than that of the lower tail (i.e. for profits) independently of the selection rule, which goes in line with the stylized fact that losses of financial time series are heavy tailed. On the other hand, the loss tail of fixed income securities seems to be particularly light, as ξ^+ is negative for DE10Y using both rules, and is not significantly different from zero for USFI10Y when using the 10th/90th quantile rule. This exponential and superexponential decay can be partially explained because of the expansive monetary policy that took place during the majority of the sample period and its direct effect on taming extreme fluctuations of bond yields. We would expect heavier loss tails when expanding the sample period. Lastly and, as hinted by the shape plots, no standard errors could not be reported under all specifications for some fixed income securities (VCIT and DE10Y) and for the NEIXCTAT index.

Table 5: Estimated parameters - 10th/90th quantiles of $F(x_i)$

| Criterion: 10th/90th quantile of $F(x_i)$ | $\hat{\xi}^-$ | $\hat{\lambda}^-$ | u^- | $\hat{\xi}^+$ | $\hat{\lambda}^+$ | u^+ |
|---|-------------------|-------------------|--------|-------------------|-------------------|-------|
| IVV | 0.281 (0.071) | 0.005 (0.000) | -0.012 | 0.153 (0.062) | 0.009 (0.001) | 0.011 |
| EZU | 0.129 (0.071) | 0.009 (0.001) | -0.015 | 0.203 (0.068) | 0.010 (0.001) | 0.015 |
| JPXN | 0.201 (0.077) | 0.006 (0.000) | -0.013 | 0.208 (0.071) | 0.007 (0.001) | 0.012 |
| EPP | 0.248 (0.070) | 0.007 (0.001) | -0.013 | 0.306 (0.076) | 0.007 (0.001) | 0.013 |
| CEC.PA | 0.056 (0.056) | 0.008 (0.001) | -0.015 | 0.230 (0.068) | 0.008 (0.001) | 0.015 |
| EEM | 0.110 (0.062) | 0.008 (0.001) | -0.015 | 0.206 (0.066) | 0.008 (0.001) | 0.016 |
| VCIT | 0.340 (NA) | 0.002 (NA) | -0.004 | 0.292 (NA) | 0.002 (NA) | 0.004 |
| EMB | 0.311 (0.081) | 0.003 (0.000) | -0.005 | 0.343 (0.075) | 0.003 (0.000) | 0.005 |
| TRET.L | 0.287 (0.068) | 0.005 (0.000) | -0.011 | 0.198 (0.068) | 0.007 (0.001) | 0.011 |
| GSG | 0.009 (0.062) | 0.008 (0.001) | -0.014 | 0.205 (0.065) | 0.008 (0.001) | 0.016 |
| DE10Y | 0.228 (0.071) | 0.002 (0.000) | -0.005 | -0.014 (0.065) | 0.003 (0.000) | 0.005 |
| USFI10Y | 0.165 (0.067) | 0.003 (0.000) | -0.006 | 0.049 (0.052) | 0.003 (0.000) | 0.006 |
| NEIXCTAT | -0.031 (0.057) | 0.004 (0.000) | -0.008 | 0.171 (0.066) | 0.004 (0.000) | 0.008 |

This table presents the estimated shape, scale and location (i.e. threshold) parameters for the upper and lower GPD tails of the semi-parametric distribution rounded to three decimal places. Standard errors are reported below parameter estimates (so far as available). The selected thresholds for upper and lower exceedances were the 90th and 10th quantile of the returns empirical distribution.

4.3 Scenario generation techniques

After having estimated the univariate SPDs, the next step is to combine them into a multivariate distribution from which random samples can be taken and used as input for the optimization problem of Section 2. We use three techniques to sample scenario sets: the first one is to use the historical data set as scenario set without any further assumption. The second one is to bootstrap the empirical copula with semi-parametric marginals to then recover the scenarios using the quantile function of each marginal (i.e. through inverse transform sampling). The third technique estimates a parametric copula over the empirical one using R-vines, as explained in Section 3.2. Once estimated, it is possible to use the parametric vine to perform inverse transform sampling again.

The first method does not require any further explanation, so we focus on the second and third one. The first step is to define the empirical copula, so we start by recalling the notation from Section 2, where \mathbf{x}_j is a column vector of returns for asset j containing t observations. To translate it into the context of the present section, \mathbf{x}_j is a column vector of negative log

Table 6: Copula families of selected R-vine structures

| | $n_u = 150$ rule | 10th/90th rule | Normal marginals |
|----------------------|------------------|----------------|------------------|
| Clayton-Gumbel (BB1) | 6 | 6 | 1 |
| Joe-Clayton (BB7) | 3 | 3 | 1 |
| Clayton | 3 | 3 | 2 |
| Frank | 7 | 8 | 7 |
| Gaussian | 2 | 2 | 1 |
| Gumbel | 10 | 10 | 6 |
| Independence | 11 | 12 | 8 |
| Joe | 1 | 1 | 0 |
| t | 35 | 33 | 52 |

This table reports the number of pair-copulae belonging to a certain family within an R-vine. The first two columns belong to R-vines with semi-parametric distributions where the GPD was fitted using one of the two baseline rules and the third column belongs to an R-vine with normally distributed marginals.

returns, for which its semi-parametric distribution function $\hat{F}_j^*(x_j)$ has been estimated in the preceding section. Trivially, plugging \mathbf{x}_j into $\hat{F}_j^*(x_j)$ yields a standard uniform vector $\hat{F}_j^*(\mathbf{x}_j) = \hat{\mathbf{u}}_j$, $\hat{u}_{tj} \sim U(0, 1) \forall t$. Recall that $\mathbf{X}_{T \times J}$ is a scenario set with T observations and J assets, so, if each column vector of $\mathbf{X}_{T \times J}$ is plugged into its respective distribution function, the resulting mapping $\hat{F}^* : \mathbf{X}_{T \times J} \rightarrow [0, 1]^J$ is the empirical copula \hat{C} of $\mathbf{X}_{T \times J}$:

$$\begin{aligned}\hat{F}^*(\mathbf{x}_1, \dots, \mathbf{x}_j) &= \hat{C}(\hat{F}_1^*(\mathbf{x}_1), \dots, \hat{F}_j^*(\mathbf{x}_j)) \\ &= \hat{C}(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_j).\end{aligned}\tag{4.3.1}$$

The empirical copula \hat{C} is subsequently the building block for the bootstrap and vine copula techniques.

To generate a scenario set using the bootstrap, T random row vectors $\tilde{\mathbf{u}}^B = (\tilde{u}_1^B, \dots, \tilde{u}_j^B)$ are drawn from the empirical copula \hat{C} with equal probability and with replacement. The resulting $T \times J$ matrix is denoted by $\tilde{\mathbf{U}}_{T \times J}^B$. Then, plugging each of the J column vectors of $\tilde{\mathbf{U}}^B$ into the respective quantile functions \hat{F}_j^{*-1} of the marginal distributions of \hat{C} results into a new scenario set:

$$\tilde{\mathbf{X}}_{T \times J}^B = \hat{C}^{-1}(\hat{F}_1^{*-1}(\tilde{\mathbf{u}}_1^B), \dots, \hat{F}_j^{*-1}(\tilde{\mathbf{u}}_j^B)).\tag{4.3.2}$$

The procedure using vine copulae is very similar. After building the empirical copula \hat{C} , an R-vine \hat{C}^V is estimated following the procedure explained in Section 3.2 and performed below in Section 4.4. Then, T random vectors $\tilde{\mathbf{u}}^V = (\tilde{u}_1^V, \dots, \tilde{u}_j^V)$ are drawn from \hat{C}^V , which again results in a $T \times J$ matrix $\tilde{\mathbf{U}}_{T \times J}^B$ composed by random uniform vectors. By plugging each column vector \mathbf{u}_j into its respective quantile function, we get a new scenario set denoted by

$$\tilde{\mathbf{X}}_{T \times J}^V = \hat{C}^{V-1}(\hat{F}_1^{*-1}(\tilde{\mathbf{u}}_1^V), \dots, \hat{F}_j^{*-1}(\tilde{\mathbf{u}}_j^V)).\tag{4.3.3}$$

4.4 Selection and fitting of R-vine copulae

Before the scenario generation techniques laid out above are applied to portfolio optimization problems, it is necessary to estimate the multivariate distribution of returns via vine copulae, i.e., to estimate \hat{C}^V in the notation of Section 4.3. As it was shown in Section 3.3.1, a multivariate distribution can be decomposed using a cascade of bivariate copula functions called pair copulae. However, decompositions are not unique and the amount of viable decompositions grows with $\frac{d!}{2} \times 2^{\binom{d-2}{2}}$ (Morales-Nápoles, 2011). In our case, where $d = 13$, this results in some 112 septillion of feasible decompositions. For this reason, the automatic selection of pair copulae and vine tree structures are a research field in its own right. A first contribution to this field is found in Aas et al. (2009), who develop an algorithm for automatic selection of D- and C-vines. The expansion to R-vines was done by Dißmann et al. (2013), who formulate a greedy, hierarchical algorithm that starts by modeling the dependence of those bivariate pairs with the strongest dependence between them. Then, the vine tree structure is determined by minimizing some information criterion on the log-likelihood function of the decompositions.

We use the algorithm of Dißmann et al. (2013) for model selection. More specifically, we use the very efficient implementation done in the R package `vinecopulib` (Nagler and Vatter, 2023), which is optimized in C++ and can be run in parallel. With respect to the choice of information criterion to guide the search of R-vine structures, it turns out that the choice of criterion has a significant influence on the so-called sparsity (i.e. parsimony) of the model. In this context, a sparse model is one that chooses the independence copula for the highest possible amount of pair copulae. Having a sparse model is fundamental for the computational efficiency of simulation with vine copulae. As Nagler et al. (2019) showed, the Bayesian Information Criterion (BIC) selects better models than the Akaike Information Criterion (AIC), which is why we use the former in our computations. Nagler et al. (2019) suggest a modified BIC (mBIC) to improve sparsity even further, but it is slightly more intense in computation and a significant improvement of sparsity is only observed for $d > 20$. At last, we do not limit pair copula selection to any particular family and allow the algorithm to choose from the following available bivariate copula families:

- Elliptical copulae (Gaussian and Student's t),
- Archimedean copulae (Clayton, Gumbel, Frank and Joe),
- Biparameter bivariate copulae (Clayton-Gumbel (BB1), Joe-Gumbel (BB6), Joe-Clayton (BB7) and Joe-Frank (BB8)),
- Non-parametric copula (transformation kernel)
- Independence copula.

Table 6 reports the number of pair copulae belonging to one of the allowed copula families for R-vines with different marginal distributions. As it is can be observed, the difference between the two threshold selection rules for the fitting of the tails of the SPDs has very little influence on the selection of pair copulae. The only notable difference is that the 10th/90th rule is marginally sparser (12 against 11 independence pair copulae), it has one more Frank pair and

Figure 2: Encoded R-vine structures with SPD marginals

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------|----|----|----|----|---|---|---|---|---|---------|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|
| | 4 | 12 | 1 | 7 | 8 | 1 | 1 | 2 | 2 | 4 | 6 | 8 | 8 | | 1 | 12 | 7 | 8 | 6 | 4 | 1 | 1 | 2 | 4 | 2 | 4 | 4 |
| | 6 | 7 | 9 | 8 | 6 | 3 | 2 | 4 | 4 | 6 | 8 | 6 | | 9 | 7 | 8 | 6 | 4 | 6 | 3 | 2 | 4 | 2 | 4 | 2 | | |
| | 2 | 8 | 3 | 6 | 4 | 2 | 4 | 6 | 6 | 8 | 4 | | 3 | 8 | 6 | 4 | 2 | 2 | 2 | 4 | 6 | 5 | 5 | | | | |
| | 1 | 6 | 2 | 4 | 2 | 4 | 6 | 1 | 8 | 2 | | 2 | 6 | 4 | 2 | 1 | 1 | 4 | 6 | 5 | 6 | | | | | | |
| | 3 | 4 | 4 | 2 | 1 | 6 | 5 | 8 | 1 | | 4 | 4 | 2 | 1 | 5 | 3 | 6 | 5 | 1 | | | | | | | | |
| | 9 | 2 | 6 | 1 | 5 | 5 | 8 | 5 | | 6 | 2 | 1 | 5 | 3 | 9 | 5 | 3 | | | | | | | | | | |
| $V_1 =$ | 5 | 1 | 5 | 5 | 3 | 8 | 3 | | | $V_2 =$ | 5 | 1 | 5 | 3 | 9 | 5 | 9 | | | | | | | | | | |
| | 8 | 5 | 8 | 3 | 9 | 9 | | | | | 8 | 5 | 3 | 9 | 10 | 10 | | | | | | | | | | | |
| | 7 | 3 | 7 | 9 | 7 | | | | | | 7 | 3 | 9 | 10 | 8 | | | | | | | | | | | | |
| | 12 | 9 | 12 | 12 | | | | | | | 12 | 9 | 10 | 7 | | | | | | | | | | | | | |
| | 11 | 13 | 13 | | | | | | | | 11 | 10 | 12 | | | | | | | | | | | | | | |
| | 13 | 11 | | | | | | | | | 10 | 11 | | | | | | | | | | | | | | | |
| | 10 | | | | | | | | | | 13 | | | | | | | | | | | | | | | | |

This figure depicts the encoded R-vine structures for vine copulae with semi-parametric marginal distributions (defined in Section 3.1.3) using the encoding of Morales-Nápoles (2011). The marginals of the left array were fitted using the $n_u = 150$ rule and the marginals of the right array with the 10th/90th quantile rule. Each number represents the corresponding asset in Table 1.

two less Student's t pairs. The small differences in pairs' selection are a good indicator of the robustness of our automatic threshold selection criteria. Furthermore, we compare the R-vines with SPDs marginals with an R-vine with normally distributed marginals. Here, the differences in pairs' selection are noticeable and, as it could have been expected, elliptical copulae are overrepresented (52 Student's t pairs), while copulae with some kind of asymmetric tail dependence (BB, Clayton, Gumbel, Joe) are systematically underrepresented, which, again, is a clear statement on the unsuitability of normality assumptions for dependence modeling of financial data. Also, the sparsity of the R-vine is worse (8 against 12 independent pairs).

Figures A.12 and A.13 depict the selected structures for R-vines with SPD marginals fitted using the 10th/90th quantile rule. Each number in the nodes corresponds to the number assigned to each asset in Table 1. As it is possible to see, the different threshold selection criteria lead to slightly different choices of tree structures. Although only the nodes of the tree are labeled, the labels of the edges (i.e. the conditional pair copulae between nodes) can be easily determined using the notation explained in Section 3.3.1. Alternatively, it is possible to use the R-vine encoding developed by Morales-Nápoles (2011), who represents a d -dimensional vine structure as a triangular matrix V . The label of the j -th edge of tree t is given by $(v_{d-j+1,j}, v_{t,j}; v_{t-1,j}, v_{t-2,j}, \dots, v_{1,j})$. The triangular matrices for the encoded vine structures we estimated are depicted in Figure 2. The matrix V_1 corresponds to the vine copula with $n_u = 150$ fitted marginals and V_2 to the vine copula with 10th/90th quantile fitted marginals.

4.5 Optimal portfolios and efficiency frontiers

Now that the EVT semi-parametric mixture has been fitted to each asset and the resulting marginal distributions have been used to construct the empirical copula \hat{C} and the vine copula \hat{C}^V , the next step is to compute optimal portfolio weights and efficiency frontiers using the

aforementioned sampling methods. In order to have statistically sound results, Monte Carlo (MC) methods have to be used with the aim of quantifying the uncertainty around each optimal solution and efficiency frontier. The first step to achieve this is to choose one of the sampling techniques defined above and generate a sequence $m = 1, \dots, M$ of scenario sets $\tilde{\mathcal{X}} = (\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_m)$ with T scenarios and J columns each. Then, the solution of the optimization problem defined in 2.2.4 with some scenario set $\tilde{\mathbf{X}}_m$ as input yields the vector of optimal weights $\mathbf{w}_m^* = (w_{m,1}^*, \dots, w_{m,J}^*)$. After computing $\mathbf{w}_m^* \forall m$, the MC estimator of the optimal weight of asset j is computed as the mean weight allocated to j over the M optimization problems and is given by

$$\hat{w}_j^* = \frac{1}{M} \sum_{m=1}^M w_{m,j}^*. \quad (4.5.1)$$

The standard deviation of the estimator is defined as

$$\hat{\sigma}_j^{w^*} = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (w_{m,j}^* - \hat{w}_j^*)^2}, \quad (4.5.2)$$

and it is used to compute the standard error of \hat{w}_j^* with

$$\hat{se}_j^{w^*} = \frac{\hat{\sigma}_j^{w^*}}{\sqrt{M}}. \quad (4.5.3)$$

The result of this Monte Carlo simulation is thus the vector of MC optimal weights $\hat{\mathbf{w}}^* = (\hat{w}_1^*, \dots, \hat{w}_J^*)$.

To compute the MC efficiency frontier, it is necessary to begin by computing the efficiency frontier of each scenario set $\tilde{\mathbf{X}}_m \in \tilde{\mathcal{X}}$. To do so, one starts by calculating the average return of all j assets in some scenario set $\tilde{\mathbf{X}}_m$. If the returns are log-returns, the average return of asset j in scenario set $\tilde{\mathbf{X}}_m$ is just its arithmetic mean $\bar{x}_{m,j} = \frac{1}{T} \sum_{t=1}^T x_{m,t,j}$ and the vector of mean returns of $\tilde{\mathbf{X}}_m$ is denoted by $\bar{\mathbf{x}}_m = (\bar{x}_{m,1}, \dots, \bar{x}_{m,J})$. The computation of the efficiency frontiers requires the definition of a sequence of minimum mean constraints that are inputted into the minimum return constraint of Equation 2.2.4. The sequence $k = 1, \dots, K$ of equispaced minimum mean constraints is defined as $(\varphi_1, \dots, \varphi_K)$ where $\varphi_1 = \min\{\bar{\mathbf{x}}_m\}$ and $\varphi_K = \max\{\bar{\mathbf{x}}_m\}$. Solving the optimization problem with some mean constraint φ_k and some scenario set $\tilde{\mathbf{X}}_m$ results in the vector of optimal weights $\mathbf{w}_{m,k}^* = (w_{m,k,1}^*, \dots, w_{m,k,J}^*)$. With the vector $\mathbf{w}_{m,k}^*$, it is possible to build the mean return $\mu(\mathbf{w}_{m,k}^*)$ of the portfolio (as defined in Equation 2.1.2) and the loss distribution $F_L(\mathbf{w}_{m,k}^*, \tilde{\mathbf{X}}_m)$ associated with it. Furthermore, by letting $\rho(F_L)$ be some risk measure associated with the loss distribution $F_L(\mathbf{w}_{m,k}^*, \tilde{\mathbf{X}}_m)$, the risk of the optimal portfolio weights $\mathbf{w}_{m,k}^*$ is $\rho(F_L(\mathbf{w}_{m,k}^*, \tilde{\mathbf{X}}_m))$. For simplicity, we denote the mean return and the risk of a vector of optimal portfolio weights $\mathbf{w}_{m,k}^*$ and a scenario set $\tilde{\mathbf{X}}_m$ by $\mu_{m,k}$ and $\rho_{m,k}$ respectively. Ergo, the efficiency frontier of a scenario set $\tilde{\mathbf{X}}_m$ is characterized by the vector of mean returns $\boldsymbol{\mu}_m = (\mu_{m,1}, \dots, \mu_{m,K})$ and the vector of risk measures $\boldsymbol{\rho}_m = (\rho_{m,1}, \dots, \rho_{m,K})$.

Since the length of the vectors $\boldsymbol{\mu}_m$ and $\boldsymbol{\rho}_m$ is defined *a priori* and these vectors are rank vectors in increasing order, an MC estimator for each of the k elements of the efficiency frontier can be computed. The MC estimator for the k -th element of the mean return vector is given

by

$$\hat{\mu}_k = \frac{1}{M} \sum_{m=1}^M \mu_{m,k}, \quad (4.5.4)$$

and that of the risk vector is analogously defined as

$$\hat{\rho}_k = \frac{1}{M} \sum_{m=1}^M \rho_{m,k}. \quad (4.5.5)$$

Doing this estimation for each of the k elements of the efficiency frontier results in the MC efficiency frontier characterized by the mean return vector $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_k)$ and the risk measure vector $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_k)$. The standard deviation of the estimators is computed as

$$\hat{\sigma}_k^\mu = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (\mu_{m,k}^* - \hat{\mu}_k^*)^2} \quad (4.5.6)$$

and

$$\hat{\sigma}_k^\rho = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (\rho_{m,k}^* - \hat{\rho}_k^*)^2}. \quad (4.5.7)$$

In contrast to the computation of optimal portfolios, we compute confidence intervals instead of standard errors for each of the estimated components of efficiency frontier. Doing so makes it possible to make a consolidated and statistically rigorous comparison of the scenario generation techniques by looking at the efficiency frontiers they produce. Assuming that μ_k and ρ_k are the true values for the k -th element of the efficiency frontier, the confidence intervals for the MC estimators are denoted by

$$\mu_k \in \hat{\mu}_k \pm \frac{z_\alpha \hat{\sigma}_k^\mu}{\sqrt{M}} \quad \text{and} \quad \rho_k \in \hat{\rho}_k \pm \frac{z_\alpha \hat{\sigma}_k^\rho}{\sqrt{M}}, \quad (4.5.8)$$

where z_α is the $(1 - \alpha/2)$ -th percentile of the standard normal distribution.

4.6 Results

The evaluation of the results of our implementation consists of two parts: first, we investigate how robust are the optimal weights of the minimum risk portfolio as a function of scenario generation method and of risk measure. We compare the minimum risk portfolios generated by minimizing variance (i.e. Markowitz) and ES at the 95% and 99% level. Then, we compare the efficiency frontiers that result from using the three scenario generation techniques outlined throughout this work.

4.6.1 Minimum risk portfolios

Table 7 presents the optimal weights for portfolios obtained by either minimizing variance or expected shortfall. The results are grouped by scenario generation and optimization technique, and standard errors are reported for those techniques that require Monte Carlo simulation. The optimal weights are divided by the threshold rule used to fit the marginal distributions of

each scenario set (with exception of the historical, of course). Generally speaking, there are no concerning deviations in allocations between scenario generation techniques when minimizing the same risk measure. The main differences can be spotted when comparing the allocations resulting from the historical scenario set and those resulting from the MC simulations that sampled using the vine copula. Although standard errors when using the bootstrap and the vine copula are not noticeably high, the latter does tend to achieve a reduction in standard errors, regardless of the optimization method. Additionally, the least number of corner solutions (weights equal to zero) occurs when using the simulation-based techniques, especially when vine copulae are combined with 99% ES optimization. While the number of corner solutions when using historical data is as high as 5, the corner solutions under 99% ES optimization are as low as 2 when using the bootstrap and as low as 1 when using vine copulae. This hints that better diversification can be achieved through the joint use of ES optimization and vine copula methods.

Regarding the allocation within the equities asset class, IIV (S&P 500 ETF), JPXN (Nikkei 400 ETF) and CEC.PA (Eastern Europe ETF) are the assets that constantly get a positive allocation. It is interesting to note, however, that IIV is always a corner solution when using the historical scenario set, but never when using simulation-based optimization. Besides, the allocation tends to be the biggest when using vine copulae and stochastic optimization of ES. This same pattern can be observed for JPXN, but not so much for CEC.PA, where allocations are fairly similar except for its weight when minimizing ES at the 99% level (0.3% using the historical scenario set vs. 2.2% and 2.5% using vine copulae). Moreover, EMM (Emerging Markets ETF) is always a corner solution and EPP (Asia-Pacific ETF) only gets a positive weight—however insignificant—in the simulation-based approaches.

For fixed income securities, VCIT (corporate bonds ETF) was by far the largest weight in the portfolio of almost all specifications; consistently achieving allocations between 43% and 53%, and also showing the largest standard errors of any asset. One interesting behavior of VCIT is the sudden drop in weight when optimizing ES over the 99% level, as the weight drops by or more than one half with respect to mean-variance and 95% ES optimization. In this case, the allocation of VCIT is redistributed towards government bonds—especially the German one (DE10Y). Though mostly present with a small allocation, the US bond has an abnormal weight of 19% when using the bootstrap and 99% ES, which is halved when using vine copulae. This is a hint of the higher variance associated with bootstrap methods and this problem will be illustrated in detail in Section 4.6.2. Lastly, and as well as with emerging markets for equities, EMB (Emerging Markets bonds ETF) has a very low weight in the portfolios, being a corner solution in most specifications and getting a slight positive allocation when using vine copulae and optimizing at a high level of ES.

The last asset class included in the portfolio is what could be defined as "alternative investments". This category encompasses TRET.L (real estate ETF), GSG (commodities ETF) and NEIXCTAT (CTA Index). There is not much to say regarding TRET.L other than it is a corner solution in most specifications and only gets an almost zero allocation when using the vine copula with margins fitted using the 10th/90th quantile threshold. GSG does have a consistently positive weight in all set-ups (between 1.5% and 6.5%), but its introduction in a

Table 7: Minimum risk portfolios - Different scenario generation techniques

| Asset | Historical | | | | | | Bootstrap | | | | | | Vine Copula 150 | | | | | | Vine Copula 90/10 10th/90th quantile | | | | | |
|----------|------------------|--------|--------|------------------|--------|--------|--------------------|--------|--------|------------------|--------|--------|------------------|--------|--------|------------------|--------|--------|---|--------|--------|-------------------|--------|---------|
| | $n_u = 150$ | | | $n_u = 150$ | | | 10th/90th quantile | | | Bootstrap | | | $n_u = 150$ | | | Vine Copula 150 | | | $n_u = 150$ | | | Vine Copula 90/10 | | |
| | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% | $\mu - \sigma^2$ | ES 95% | ES 99% |
| IVV | 0.000 | 0.000 | 0.000 | 0.003 | 0.003 | 0.009 | 0.003 | 0.002 | 0.007 | 0.012 | 0.013 | 0.011 | 0.011 | 0.011 | 0.011 | 0.010 | 0.010 | 0.012 | 0.011 | 0.011 | 0.011 | 0.010 | 0.011 | (0.001) |
| EZU | 0.012 | 0.000 | 0.006 | 0.012 | 0.005 | 0.008 | 0.011 | 0.006 | 0.007 | 0.001 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | (0.001) |
| JPXN | 0.016 | 0.015 | 0.072 | 0.015 | 0.022 | 0.057 | 0.015 | 0.023 | 0.062 | 0.020 | 0.022 | 0.023 | 0.019 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.021 | 0.020 | 0.021 | 0.020 | 0.021 | (0.002) |
| EPP | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | 0.002 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | (0.001) |
| CEC.PA | 0.031 | 0.028 | 0.003 | 0.030 | 0.016 | 0.004 | 0.030 | 0.016 | 0.005 | 0.031 | 0.031 | 0.021 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | (0.002) |
| EEM | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | (0.001) |
| VCTT | 0.451 | 0.516 | 0.276 | 0.448 | 0.493 | 0.197 | 0.438 | 0.479 | 0.199 | 0.501 | 0.526 | 0.266 | 0.514 | 0.514 | 0.514 | 0.514 | 0.514 | 0.514 | 0.506 | 0.506 | 0.237 | 0.237 | 0.237 | (0.011) |
| EMB | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.001 | 0.000 | 0.005 | 0.005 | 0.001 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | (0.002) |
| TRET.L | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | (0.001) |
| GSG | 0.039 | 0.021 | 0.015 | 0.038 | 0.030 | 0.034 | 0.039 | 0.030 | 0.034 | 0.045 | 0.047 | 0.059 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | (0.001) |
| DE10Y | 0.263 | 0.234 | 0.338 | 0.255 | 0.216 | 0.291 | 0.258 | 0.220 | 0.288 | 0.235 | 0.235 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | 0.223 | (0.001) |
| USFI10Y | 0.015 | 0.000 | 0.093 | 0.028 | 0.030 | 0.196 | 0.030 | 0.034 | 0.195 | 0.006 | 0.006 | 0.078 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | (0.006) |
| NEIXCTAT | 0.173 | 0.185 | 0.196 | 0.172 | 0.184 | 0.198 | 0.174 | 0.187 | 0.198 | 0.142 | 0.142 | 0.164 | 0.146 | 0.146 | 0.146 | 0.146 | 0.146 | 0.146 | 0.150 | 0.150 | 0.177 | 0.177 | 0.177 | (0.003) |

This table presents the weights for minimum risk portfolios using different scenario generation techniques. Optimization is performed using mean-variance ($\mu - \sigma^2$) optimization and stochastic optimization with ES at the 95% and 99% confidence levels. The results are further grouped by scenario generation technique (either bootstrap or vine copulae) and by the rule used to fit the tails of the marginal distributions ($n_u = 150$ or the 10th/90th quantile rule). Each weight was computed using Monte Carlo simulation over 200 generated scenario sets (except for the results using historical data). Standard errors are reported where applicable.

portfolio should be addressed carefully. As Table 2 shows, its CAGR was negative over the sampling period and its volatility was on the upper end. Although commodities have desirable properties to diversify a portfolio, its use as "long-only" asset (which is the case of GSG) is very unlikely to succeed. Finally, the NEIXCTAT Index gets a rather consistent and stable allocation under all optimization and scenario generation methods, with allocations ranging from 14.2% up to 19.6%; getting invariably higher allocations when minimizing 99% ES.

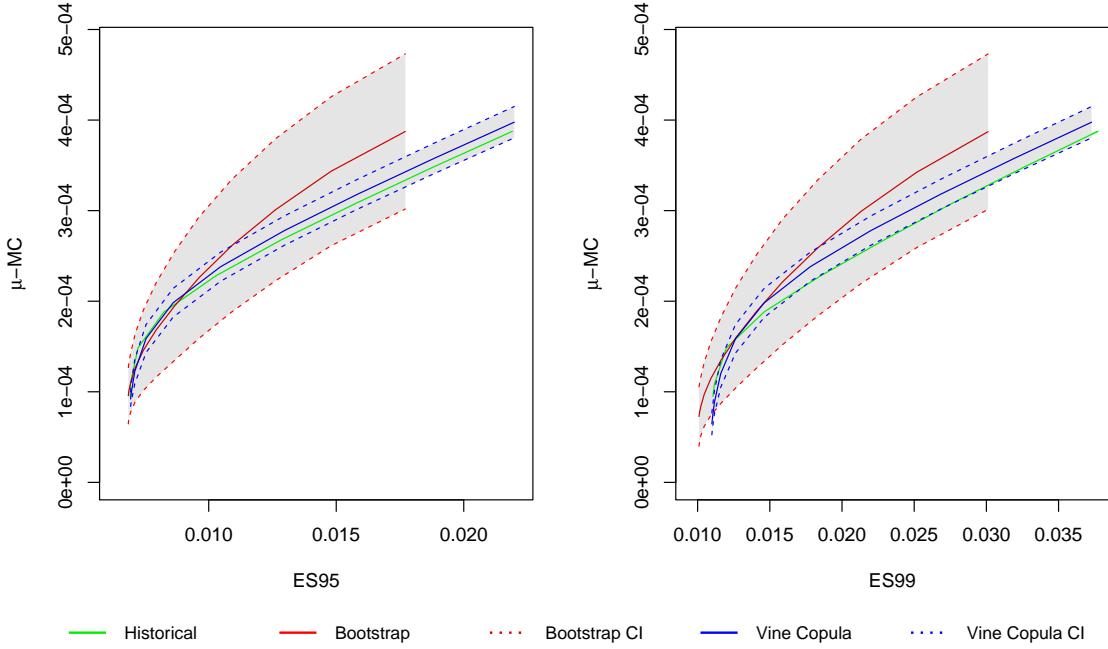
As a side note, the NEIXCTAT Index tracks the performance of so-called Commodity Trading Advisors (CTAs), which are asset managers that specialize in long and short trading with commodities futures and are known to be a high risk-high reward asset class that is significantly uncorrelated with the traditional asset classes. CTAs are usually sought after as source of crisis alpha (i.e. as an asset that pays-off in a non-linear way when the traditional asset classes experience sharp drawdowns). For this reason, it is particularly interesting that NEIXCTAT gets an unmistakably higher allocation (especially when optimizing over the worst 1% range of outcomes) since the optimal solution does choose an asset that is used and designed with the specific aim of protecting against the worst possible realizations. Nonetheless, one shortcoming of the index we used is that it does not fully represent the behavior of a prototypical CTA. As table 2 shows, the CAGR (3.83%) and the standard deviation (11.53%) of NEIXCTAT are not what should be expected from a crisis alpha source. In spite of this, it is indeed highly uncorrelated with all the observed assets, as can be seen in Table 3. In our opinion, the choice of NEIXCTAT is still justified because its status as aggregate index avoids the arbitrary decision of choosing one CTA in particular. Future research could certainly delve into a comprehensive understanding of the role of CTAs in an ES optimal portfolio.

4.6.2 Efficiency frontiers

The main tool to evaluate the performance of scenario generation techniques is the plotting of efficiency frontiers as outlined in Section 4.5. To perform MC estimation, we restrict ourselves to generate $M = 200$ scenario sets due to computational constraints and set the length of the sequence of minimum mean constraints $(\varphi_1, \dots, \varphi_K)$ at $K = 15$. The MC confidence intervals (CI) (defined in Equation 4.5.8) are computed at the 95% level for all the presented applications. Sampling from the estimated vine copula $\tilde{\mathbf{X}}_{T \times J}^V$ was done using quasi-random number generation through the generalized Halton sequence implemented in the R package **rvinecopulib** (Nagler and Vatter, 2023). The efficiency frontier depicts the daily expected return $\hat{\mu}_k$ and the daily level of ES.

Figure 3 presents our baseline application by comparing the efficiency frontiers of the historical scenario set, the bootstrap generated scenario sets and the vine copula generated scenario sets (with 10th/90th quantile marginals) optimized at the 95% and 99% ES level. Both simulation approaches yield a higher efficiency frontier than that of the historical scenario set, the bootstrap lying higher than the vine copula. However, the frontier of the historical scenario set is contained in the confidence interval of both the bootstrap and the vine copula frontiers (being at exactly the lower bound of the vine copula frontier CI for 99% ES optimization). Although

Figure 3: Efficiency frontiers for different scenario set generation methods

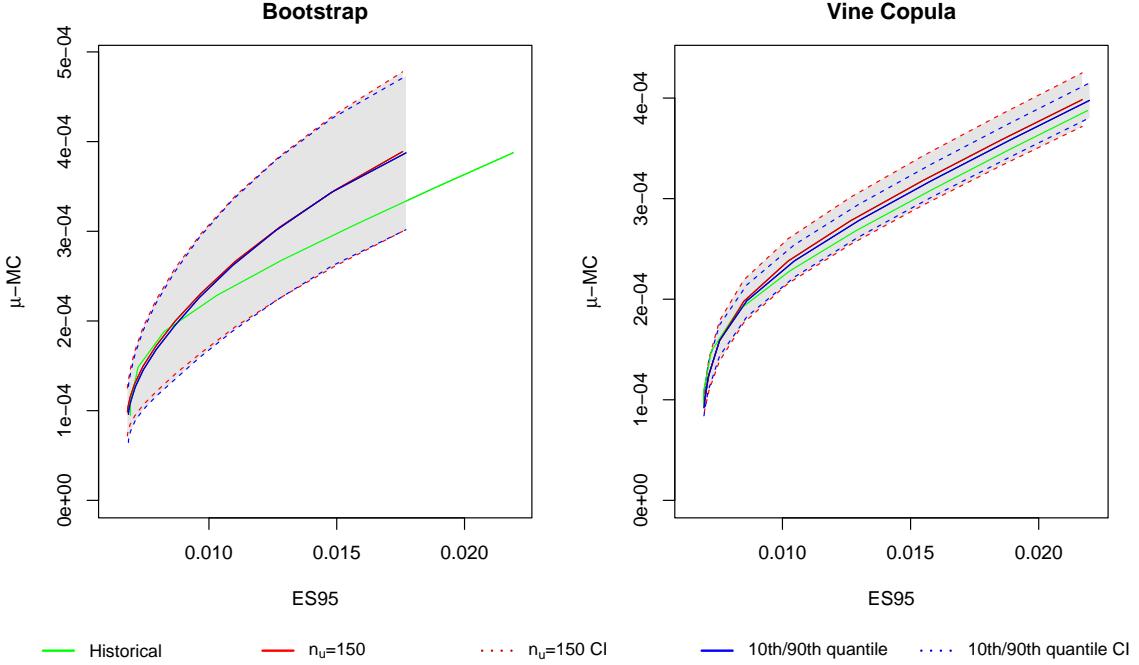


This figure presents the efficiency frontiers of 95% ES and 99% ES optimal portfolios obtained using three different scenario set generation methods. The x-axis is the daily ES and the y-axis the MC expected return $\hat{\mu}_k$. The efficiency frontiers were computed using a sequence of 15 minimum return constraints. The efficiency frontiers of the bootstrap and the vine copula scenario sets were computed using Monte Carlo as described in Section 4.5. Confidence intervals at the 95% level are reported when applicable.

the bootstrap may seem the first-best approach, it has two apparent weaknesses, namely, the very wide confidence interval (whose lower bound lies both below the lower bound of the vine copula CI and below the historical scenario set) and the shorter efficiency frontier. Both weaknesses can be traced back to a problem common to sampling via bootstrap. The shorter frontier indicates that the same efficient portfolio is computed several times either at the beginning or the end of the frontier, implying that the expected returns from the sampled scenario sets are not representative of those of the historical scenario set. This sampling problem, in turn, induces high variance in the MC estimates and leads to the wider CI as well. Reducing variance would require significantly more simulations, but this would come at the cost of a very high computational intensity. By comparison, the efficiency frontier when using vine copulae has a remarkably low variance and its resemblance to the historical frontier is a clear sign of the high degree of representativity achieved when using advanced dependence modeling techniques. In our view, increasing the number of simulations for the bootstrap is not justified as the vine copula approach achieves lower variance and stabler frontiers with the same amount of simulations.

Although the historical frontier is within the CI of the vine copula frontier, the stability of the latter confirms that our approach yields accurate and reliable samples that are able to capture the subtle dependence structure between assets, achieving a frontier that is at least as high as the one obtained through the historical scenario set. Thus, our approach does not result in too conservative allocations (which was our initial concern) despite the fact that our tail modeling assumptions are fairly conservative. Accounting for extreme events and the correct

Figure 4: Efficiency frontiers for different threshold selection methods



This figure presents the efficiency frontiers of 95% ES optimal portfolios obtained using two different rules to select the threshold for tail modeling of the marginal distributions. The x-axis is the daily ES and the y-axis the MC expected return $\hat{\mu}_k$. The efficiency frontiers were computed using a sequence of 15 minimum return constraints. The efficiency frontiers of the bootstrap and the vine copula scenario sets were computed using Monte Carlo as described in Section 4.5. Confidence intervals are reported at the 95% level.

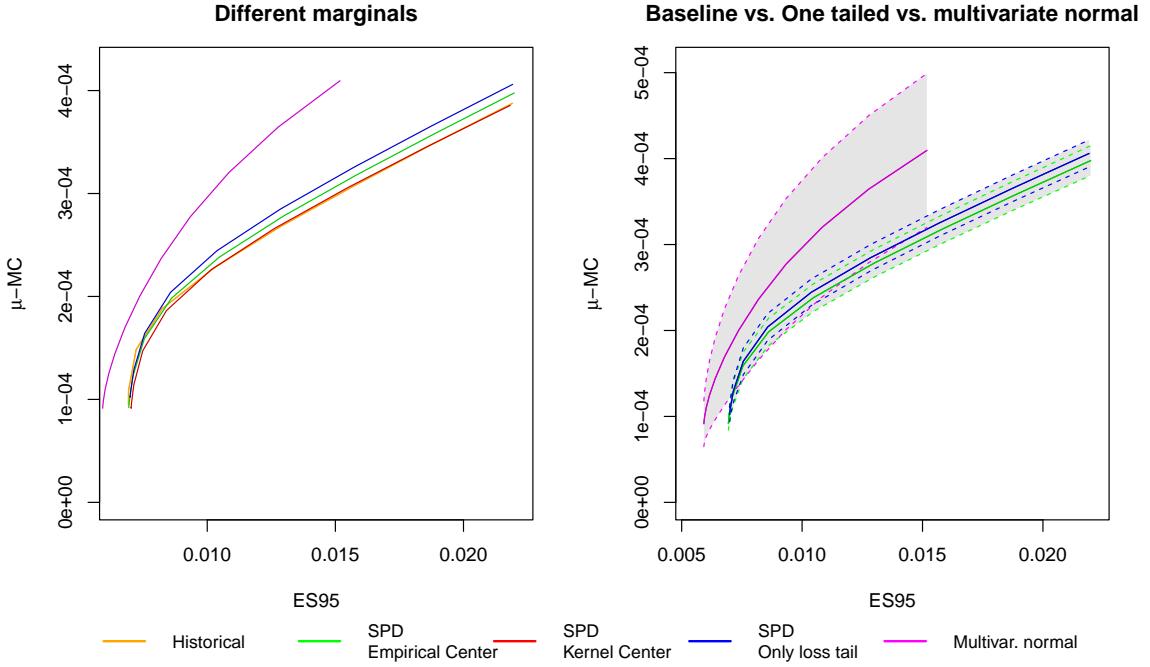
dependence structure does not necessarily come at the cost of sacrificing expected returns. Either way, a word of caution is needed when using our approach for actual decision making: as the confidence intervals hinted, our approach does capture the observed dependence structures, and so, its best use should be as complement to validate other techniques, such as historical simulation. The robustness of our technique over time as stand-alone criterion for portfolio selection should be ideally evaluated through backtests, yet this assessment is outside the scope of this thesis and is left as subject for future inquiries.

Sensitivity analysis

One possible critique of our approach is that it depends on too many parameters and the robustness of the baseline results could be threatened by changes in the risk measure, the threshold to fit the SPD, the overall choice of marginals or the optimization technique. To illustrate how problematic these concerns are, we plot the efficiency frontiers using set-ups that address the different potential shortcomings.

Figure 4 tests the sensitivity with respect to threshold selection in the semi-parametric marginals. The efficiency frontiers correspond to 95% ES optimal portfolios, where the marginals of the multivariate distributions used to generate scenario sets were modeled using the $n_u = 150$ and the 10th/90th quantile rules for threshold selection. The scenario sets were generated using the bootstrap and the vine copula. As it is possible to see, the resulting efficiency frontiers and confidence intervals are almost identical and are an improvement with respect to the historical

Figure 5: Efficiency frontiers using different marginal distributions for scenario generation

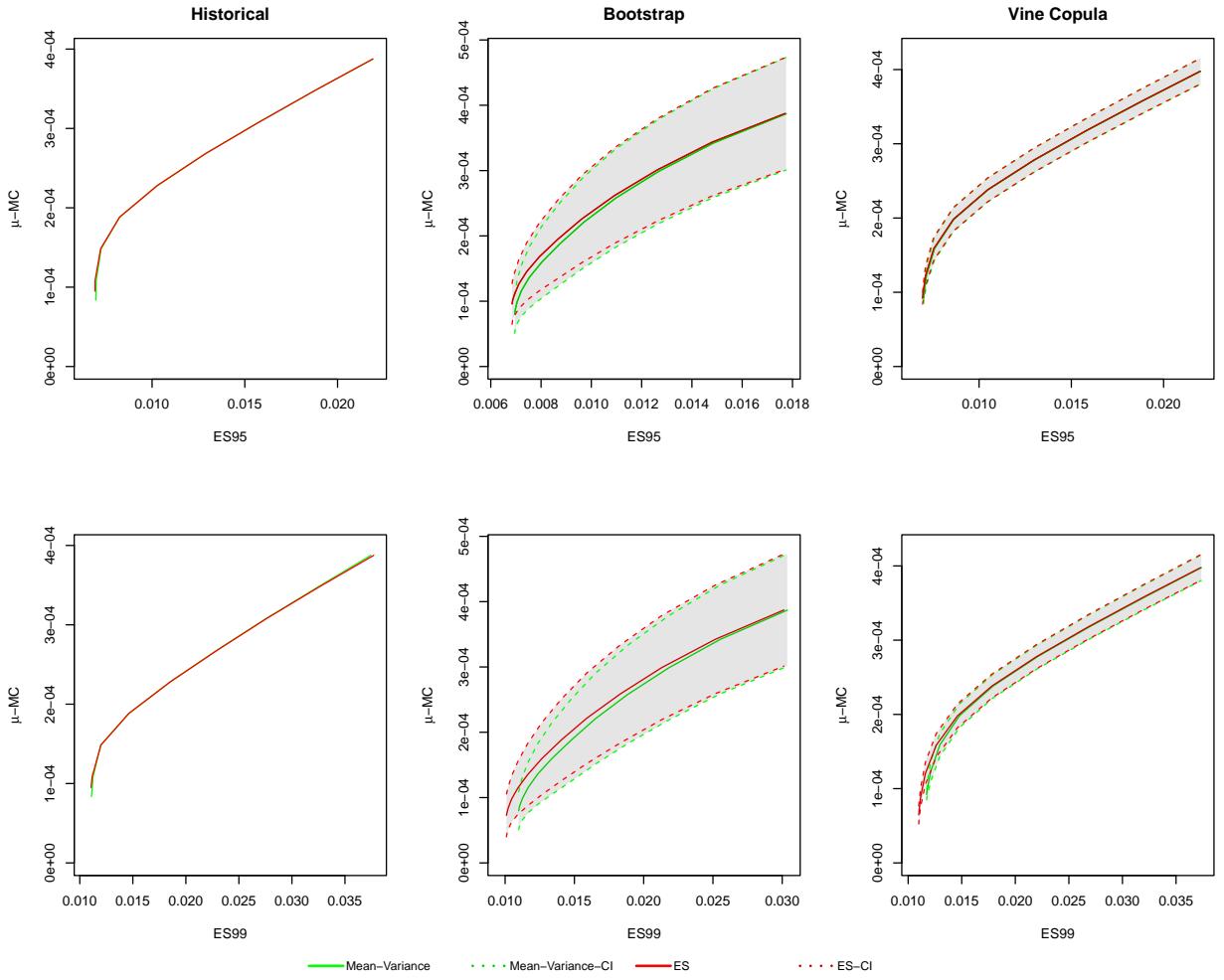


This figure presents the efficiency frontiers of 95% ES optimal portfolios obtained using different specifications for the marginal distributions of the scenario sets. The tails of the SPDs were chosen using the 10th/90th quantile rule. The x-axis is the daily ES and the y-axis the MC expected return $\hat{\mu}_k$. The efficiency frontiers were computed using a sequence of 15 minimum return constraints. All scenario sets were generated via vine copulae. The efficiency frontiers were computed using Monte Carlo as described in Section 4.5. Confidence intervals are reported at the 95% level where applicable.

scenario set. There are no noticeable differences in the frontiers of the bootstrap, while, with vine copulae, the frontier corresponding to the $n_u = 150$ rule is slightly higher. The CI is also wider and fully encompasses that of the 10th/90th quantile rule. This hints at the bias-variance trade-off inherent to the Peaks-Over-Threshold method. Nonetheless, the differences are not statistically significant and show that small changes in the threshold for tail distribution estimation do not compromise the robustness of the baseline results.

The next sensitivity check is with respect to the modeling of the marginals. In figure 5, we make small changes to the SPD and use the multivariate normal distribution (MVN) as well. This figure presents the efficiency frontiers of 95% ES optimal portfolios with scenario sets that were generated using vine copulae and direct samples from the MVN. The first change made is the use of the Epanechnikov kernel instead of the empirical distribution for the non-parametric center of the SPD. The second change was to only model the right tail of the loss distribution (i.e. the losses) with the GPD, where the threshold u was the 90th quantile of the loss distribution. Lastly, we made an excursion into the world of elliptical distributions and computed an efficiency frontier with scenario sets generated from the multivariate normal distribution, which we fitted to the sample via maximum likelihood. The differences between the different set-ups of SPDs are almost unnoticeable, but the frontier corresponding to the SPD marginals with only one parametric tail is the highest one. Conversely, there is no benefit in using kernel density for the center of the SPD as the efficiency frontier is identical to that of the historical scenario set (which is barely noticeable). Regarding the MVN, it seems that it

Figure 6: Efficiency frontiers using different optimization methods



This figure presents the efficiency frontiers of Mean-Variance, 95% ES and 99% ES optimal portfolios obtained using three different scenario set generation methods. The x-axis is the daily ES and the y-axis the MC expected return $\hat{\mu}_k$. The efficiency frontiers were computed using a sequence of 15 minimum return constraints. The efficiency frontiers of the bootstrap and the vine copula scenario sets were computed using Monte Carlo as described in Section 4.5. Confidence intervals at the 95% level are reported when applicable.

achieves a great improvement, but it has similar problems as when using the bootstrap, namely, high variance and a short efficiency frontier. In the plot to the right, where the CI is reported, it becomes clear that the baseline approach even lies within the CI of the MVN. Besides, the use of the MVN in real-world decision problems is strongly advised against, as the pitfalls and dangers of elliptical distributions are well known (Embrechts et al., 2002).

The last sensitivity check performed is with respect to optimization methods. Figure 6 presents the efficiency frontier of mean-variance, 95% ES and 99% ES optimal portfolios using the three scenario set generation methods. When using historical returns and vine copulae, mean-variance and stochastic ES optimization result in the same efficiency frontier. The difference becomes more notable when using the bootstrap but, as has been mentioned enough, this may be attributable to the high variance of this sampling method. The phenomenon of observing almost identical efficiency frontiers with different optimization methods should not be surprising as it has been documented since the seminal contributions of Rockafellar and Uryasev (2000) and Krockmal et al. (2001). As these authors recall, the efficiency frontiers are highly

sensible to asset selection and it is not uncommon to observe a certain convergence between mean-variance and stochastic optimization of ES.

All in all, the sensitivity checks did not show any strong deviation from the baseline results. On the one hand, they confirmed the stability of vine copula against resampling methods and showed that, if else, it may be sensible to only model the loss tail with the GPD instead of both. The influence of this change in a backtest, however, should be negligible as ES is already computed only over the loss tail. The pragmatic choices of threshold we made to model the tails via POT proved to be robust and the induced variance coming from a higher threshold selection was exiguous. Finally, the choice of optimization method does not seem to play a role as great as the choice of scenario generation method but, again, the full implications of this choice over time should be addressed via backtesting.

5 Conclusions

The present thesis develops a new scenario generation technique for one period stochastic optimization problems using EVT and advanced copula modeling. The motivation behind this approach was to improve portfolio risk management by better accounting for extreme events and tail dependence when generating scenarios for portfolio optimization. In that order of ideas, the contribution this work made was twofold: on the one hand, the joint use of regular (R-) vines and univariate modeling techniques from EVT expanded the literature on multivariate extremes modeling and, on the other hand, the scenario generation technique proved to be a meaningful and robust expansion to portfolio selection problems.

Scenario generation problems are ultimately problems of discretizing samples from some multivariate distribution or stochastic process. In our approach, we first model the univariate distribution of returns of each asset using a semi-parametric compound distribution with Generalized Pareto (GPD) tails and non-parametric center. Then, the multivariate distribution of returns is recovered using a novel approach for multivariate extreme modeling called vine copulae (more specifically, R-vine copulae), which allows to model the dependence structure between assets in a complex, accurate and flexible way. The resulting copula distribution can be used to simulate scenarios or entire scenario sets for stochastic optimization problems.

Fundamentally, the question we wanted to answer was whether our approach to scenario generation resulted in too conservative portfolio management decisions, or whether higher expected returns were attainable even after accounting for extreme events and for tail dependence in the univariate and multivariate worlds. To evaluate the performance of our approach, we computed the efficiency frontier of expected shortfall optimal portfolios using three different scenario generation procedures (historical data, bootstrap simulation and Monte Carlo simulation with samples drawn from the vine copula) and different multivariate modeling techniques.

In summary, higher efficiency frontiers were partially attainable using our methodology. Depending on the set-up, our approach yielded frontiers that were at least as good as the empirical one and had a significantly lower variance when compared to bootstrap and multivariate normal sampling. The stability and low variance of the vine copula approach show that, firstly, high quality samples that capture the complex dependence between assets can be generated through

our method (and thereby constitute a valuable contribution to multivariate extremes modeling because of its flexibility and accuracy) and, secondly, that returns do not need to be sacrificed when modeling extreme events rigorously, at least when ES is used as risk constraint. The performed sensitivity checks further confirm the robustness of our results and, again, highlight the low variance associated with vine copula simulation. Additionally, the computation of minimal ES portfolios when using vine copula simulation resulted in fewer corner solutions than when using the bootstrap or empirical data, which is one of the common concerns when performing portfolio optimization, regardless of the risk measure that is used.

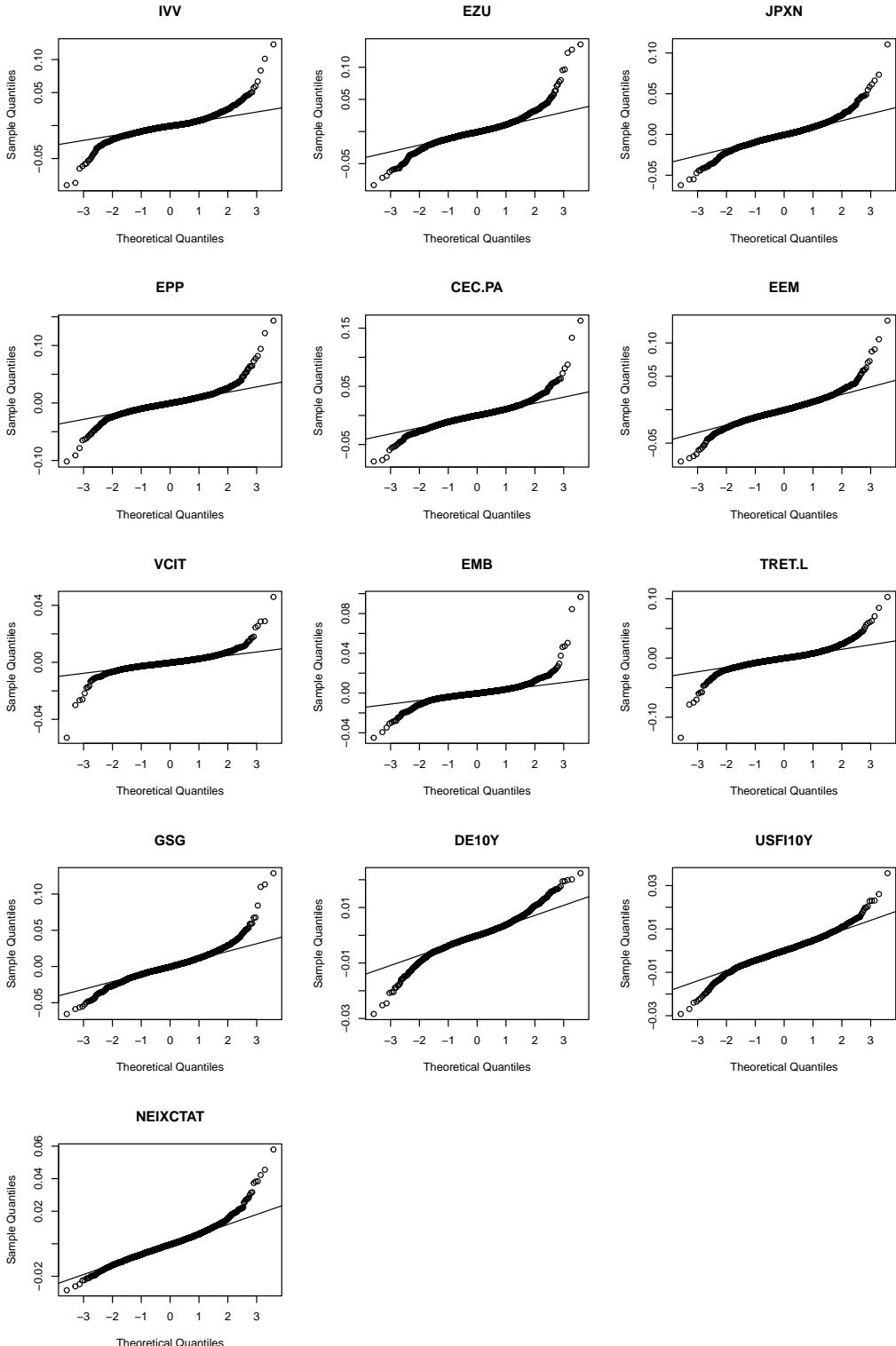
As for the questions that remain open, one possible improvement could be the use of more sophisticated methods for automatic threshold selection when fitting the GPD of the univariate distributions. Nonetheless, as the sensitivity analyses showed, it is likely that the improvement brought about by more complex criteria will, at best, be marginal. Moreover, it can be said that we sacrificed the length of the sample period in order to reduce dimensionality via ETFs. As it was explained, the use of ETFs has desirable properties when reducing dimensionality, but the availability of ETF data for some asset classes is reduced. This leads to a trade-off that could be dubbed as the time-dimensionality trade-off, where the length of the sampling period can only be expanded at the cost of higher dimensionality, where ETFs are substituted with the price data of their individual components. Although we decided to sacrifice time for dimensionality in order to include as many asset classes as possible, future research could take this problem into account and make an asset universe selection for which a longer time horizon is available.

In conclusion, the status of our technique as reliable tool for scenario generation is clear-cut and well supported. Yet, though the efficiency frontiers hint that our technique is suitable for decision making under uncertainty, our evaluation was done exclusively in-sample. For this reason, our results are, at most, a necessary condition for the choice of our methodology as portfolio management criterion. To asses the sufficiency of our approach, future research should focus on out-of-sample backtesting, where it is possible to asses whether our approach does improve the robustness of a portfolio with respect to extreme events without making great sacrifices to the CAGR of the portfolio over time. In our opinion, the greatest strength of the modeling technique herein developed is that it has a wide range of applications, not only limited to portfolio optimization, but to any field requiring the modeling and simulation of complex dependence structures and extreme events.

A Appendix

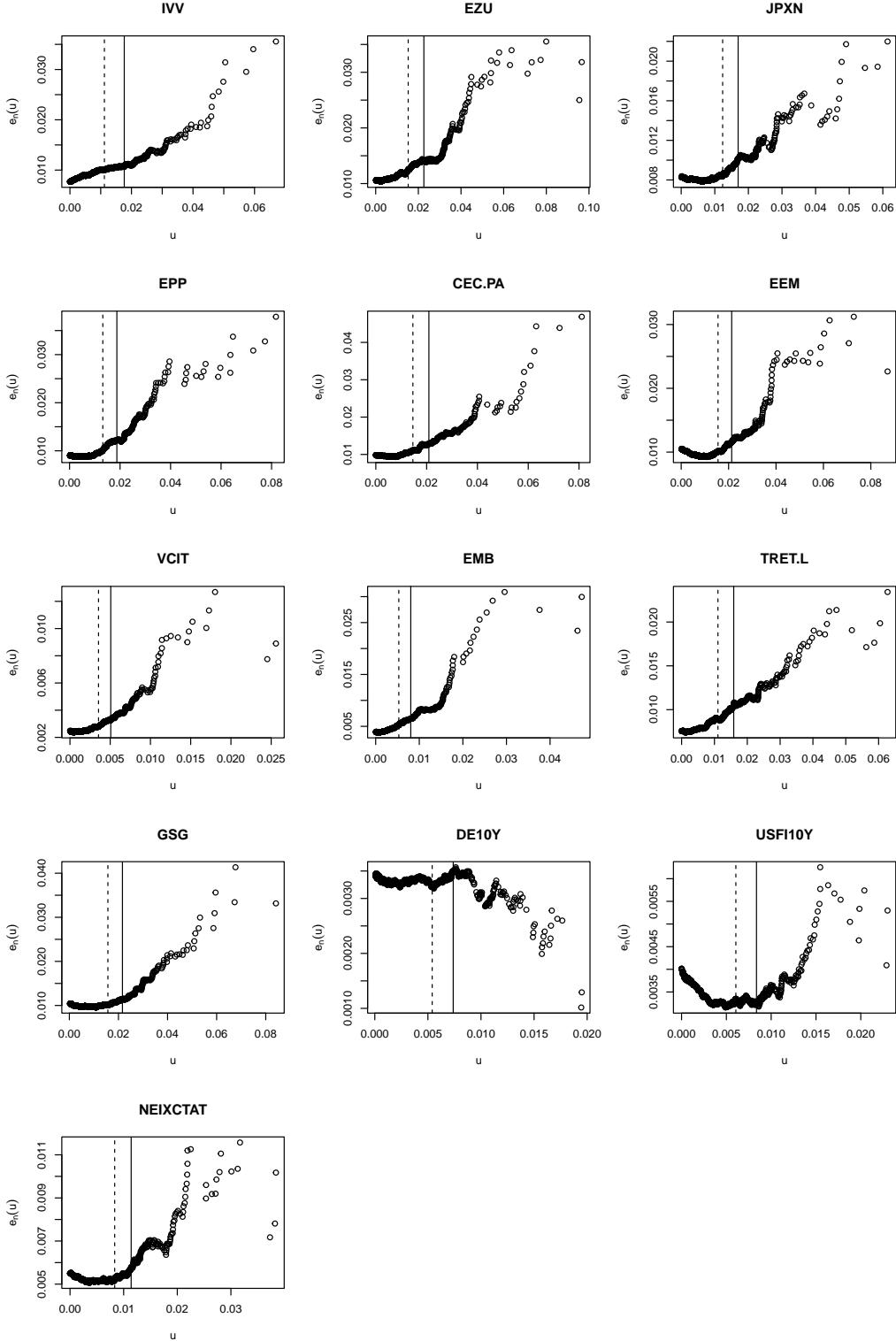
A.1 Additional figures

Figure A.7: Q-Q plots of sample returns



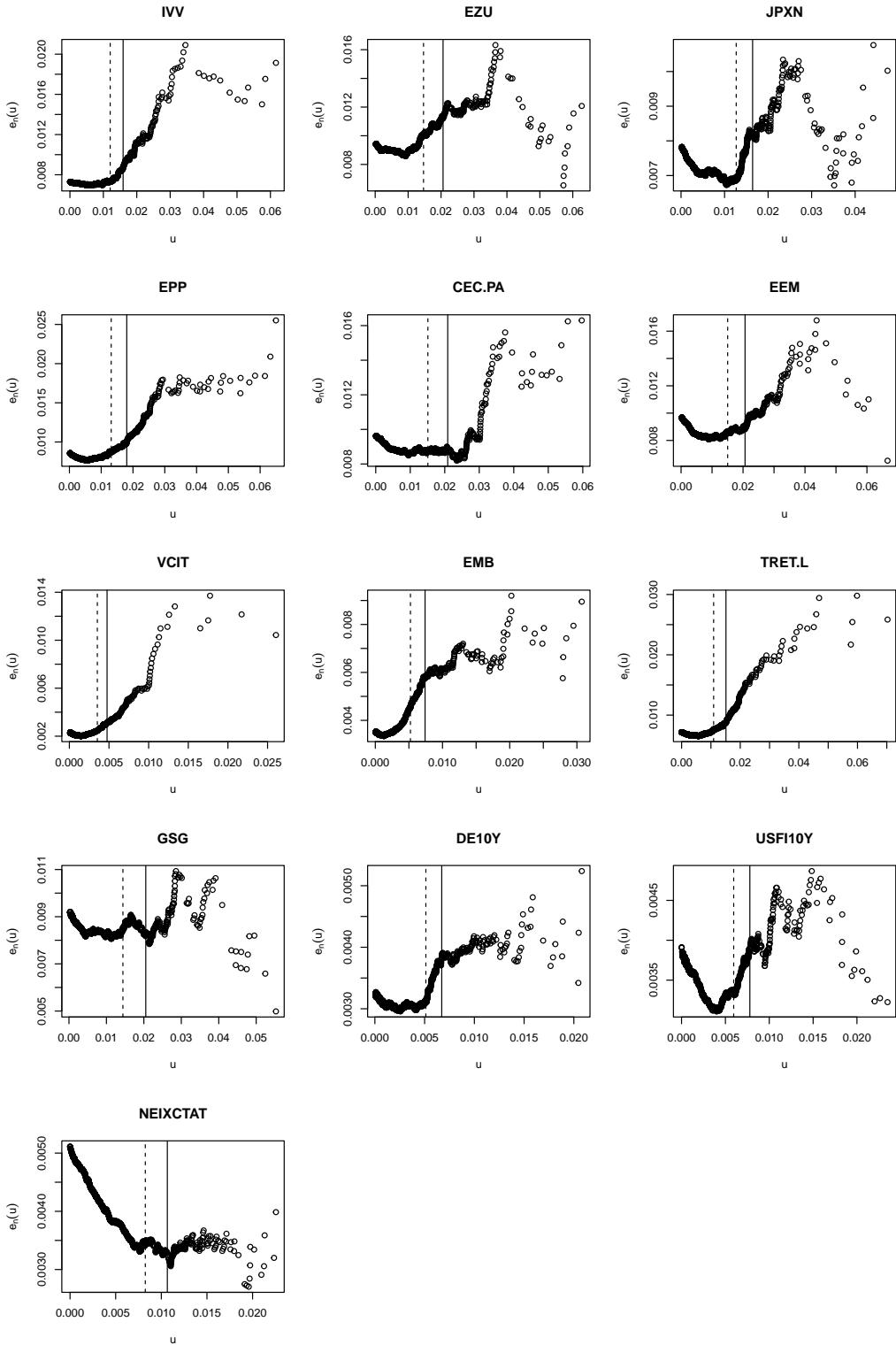
This figure presents the QQ plots for the loss distributions of twelve assets and one index. The loss distribution consists of daily negative log returns. The sample period starts in June 2011 and ends in May 2023, yielding 2029 observations.

Figure A.8: Sample mean excess plots for upper tail exceedances



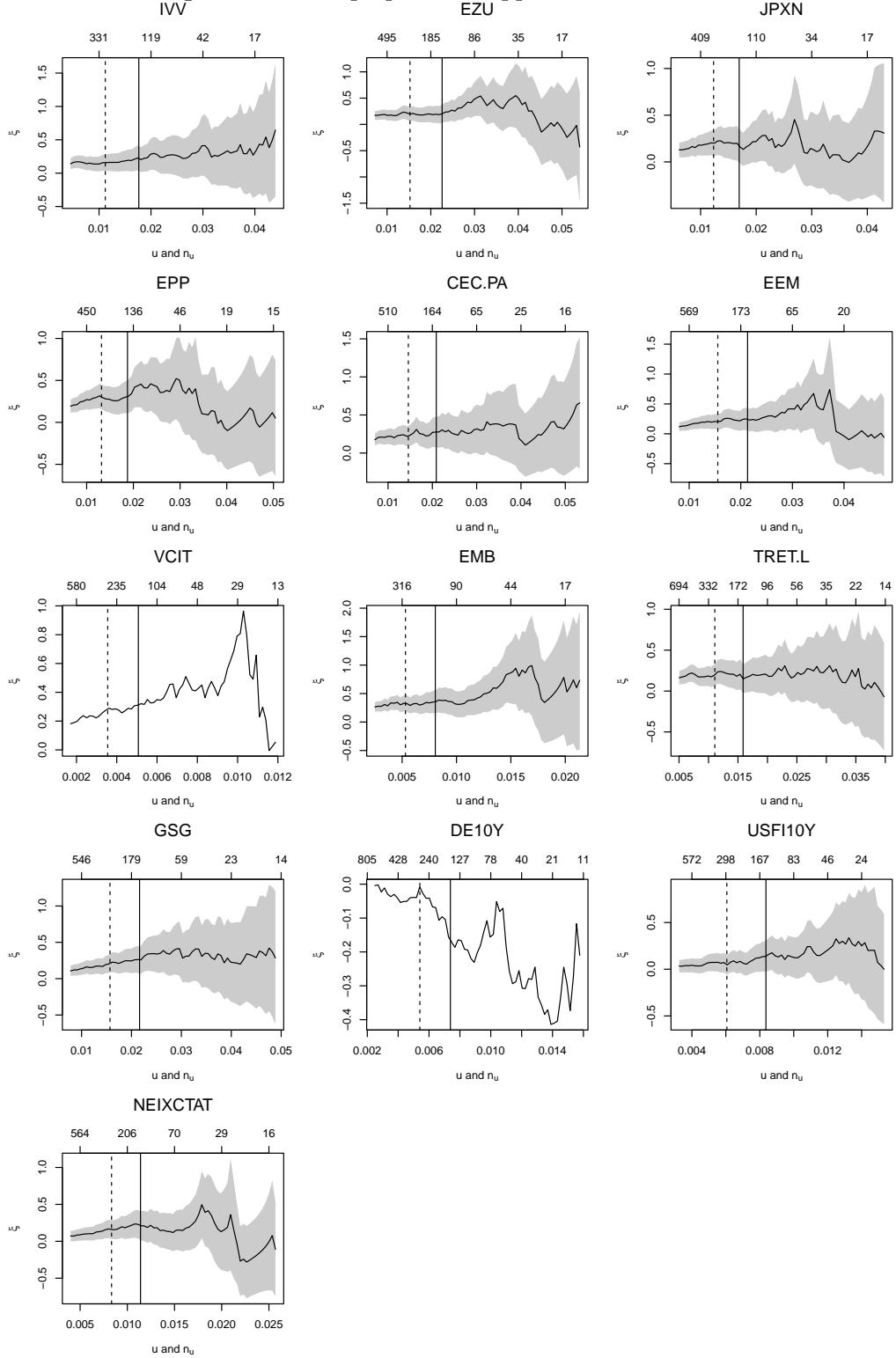
This figure presents the mean excess plots for the upper threshold exceedances of twelve assets and one index. The mean excess for each threshold value was computed using the positive values of the loss distribution for each time series. The loss distribution consists of daily negative log returns and, therefore, its positive values are losses. The solid line is the threshold value u above which there are 150 observations available (ca. 95th quantile of the loss distribution). The dashed line is the threshold value u that corresponds to the 90th quantile of the loss distribution. The graphs were plotted using the built-in functions in the R package `qrntools` (Hofert et al., 2022).

Figure A.9: Sample mean excess plots for lower tail exceedances



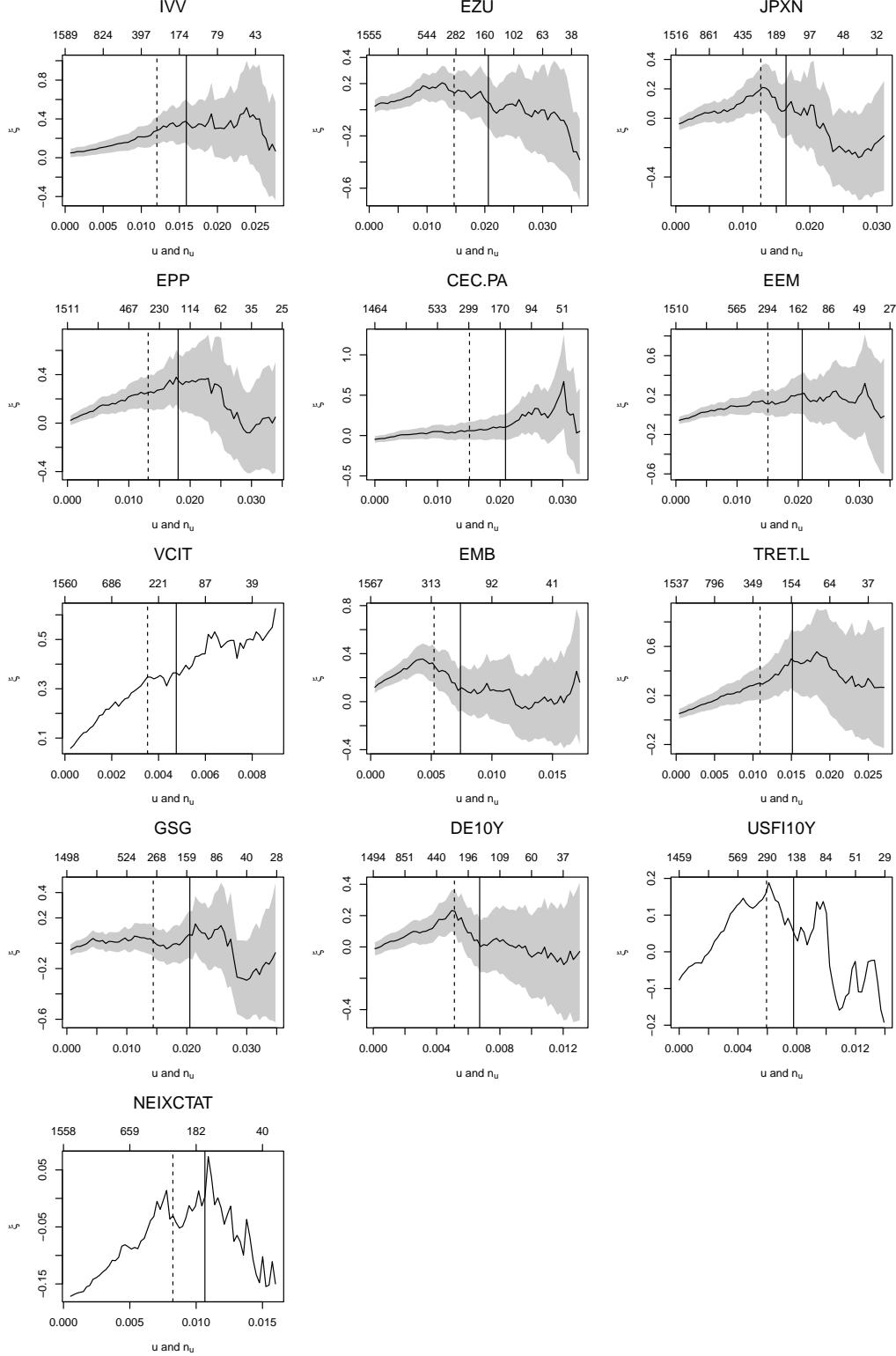
This figure presents the mean excess plots for the lower threshold exceedances of twelve assets and one index. The mean excess for each threshold value was computed using the negative values of the loss distribution for each time series. The loss distribution consists of daily negative log returns and, therefore, its negative values are gains. The solid line is the threshold value u below which there are 150 observations available (ca. 5th quantile of the loss distribution). The dashed line is the threshold value u that corresponds to the 10th quantile of the loss distribution. The graphs were plotted using the built-in functions in the R package `qrntools` (Hofert et al., 2022).

Figure A.10: Shape plots for upper tail exceedances



This figure presents the shape parameters ξ and their corresponding 95% confidence interval (when available) as a function of threshold u . The plots were computed for the right tail of the loss distributions of twelve assets and one index. The lower labels of the x-axis are the value for the threshold u for which ξ was computed. The upper labels indicate how many observations n_u were available for each value of u . The solid, vertical line is the threshold value u above which there are 150 observations available. The dashed, vertical line is the threshold value u that corresponds to the 90th quantile of the loss distribution. The graphs were plotted using the built-in functions in the R package `qrmttools` (Hofert et al., 2022).

Figure A.11: Shape plots for lower tail exceedances



This figure presents the shape parameters ξ and their corresponding 95% confidence interval (when available) as a function of threshold u . The plots were computed for the left tail of the loss distributions of twelve assets and one index. The lower labels of the x-axis are the value for the threshold u for which ξ was computed. The upper labels indicate how many observations n_u were available for each value of u . The solid, vertical line is the threshold value u below which there are 150 observations available. The dashed, vertical line is the threshold value u that corresponds to the 10th quantile of the loss distribution. The graphs were plotted using the built-in functions in the R package `qrmtools` (Hofert et al., 2022).

Figure A.12: R-vine structure with SPD marginals fitted using the $n_u = 150$ rule

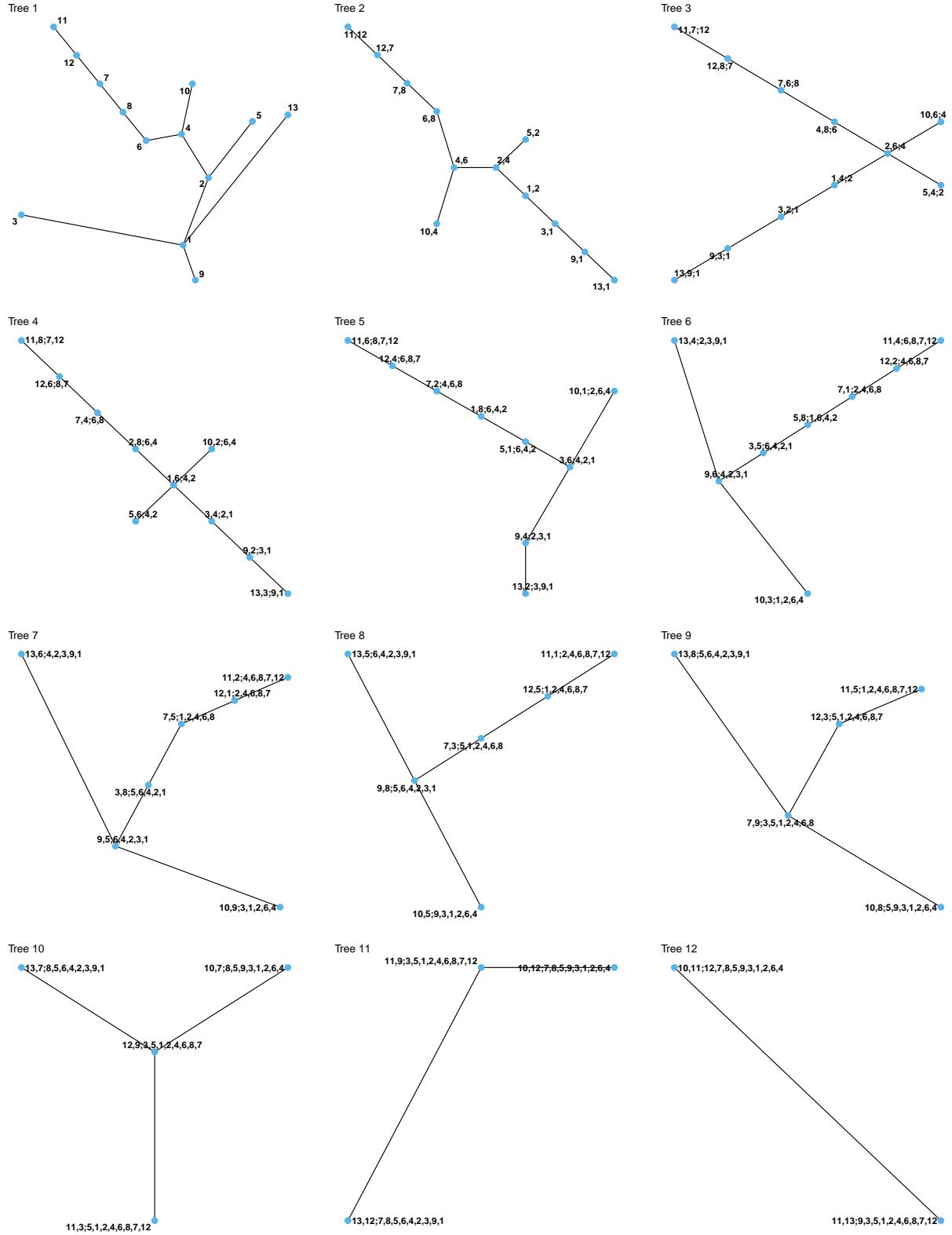
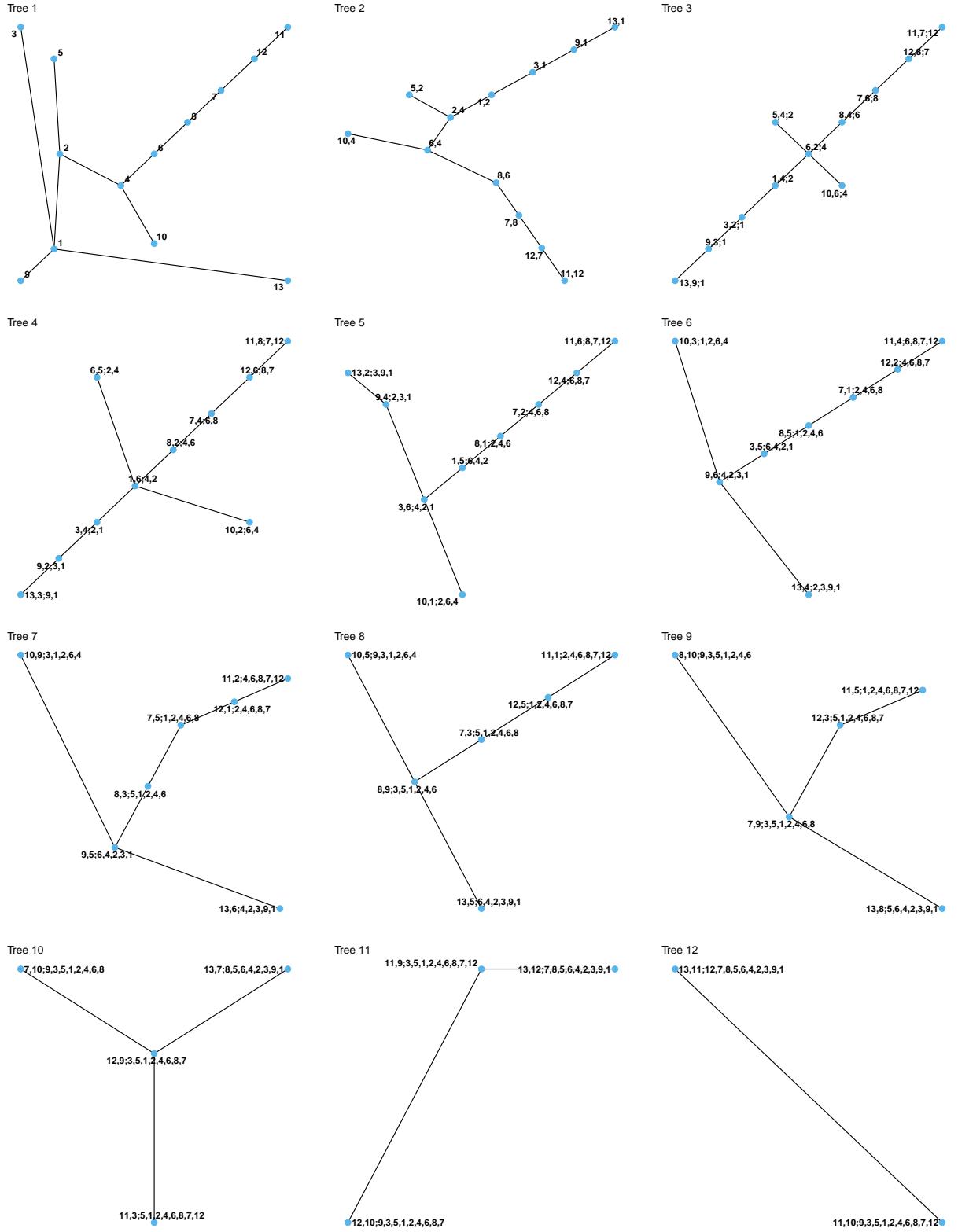


Figure A.13: R-vine structure with SPD marginals fitted using the 10th/90th quantile rule



This figure depicts the R-vine structure for a regular vine with semi-parametric marginal distributions as defined in Section 3.1.3. Model selection was performed with the R package `rvinecopulib` (Nagler and Vatter, 2023).

References

- Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4):43.
- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.
- Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1/4):245–268.
- Bedford, T. and Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068.
- Bekiros, S., Hernandez, J. A., Hammoudeh, S., and Nguyen, D. K. (2015). Multivariate dependence risk and portfolio optimization: An application to mining stock portfolios. *Resources Policy*, 46:1–11.
- Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.
- Carmona, R. (op. 2014). *Statistical analysis of financial data in R*. Springer Texts in Statistics. Springer, New York [etc.], 2nd ed. edition.
- Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52.
- Chen, M., Mehrotra, S., and Papp, D. (2015). Scenario generation for stochastic optimization problems via the sparse grid method. *Computational Optimization and Applications*, 62(3):669–692.
- Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.
- Czado, C. and Nagler, T. (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(1):453–477.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.
- Di Domenica, N., Lucas, C., Mitra, G., and Valente, P. (2007). Scenario generation for stochastic programming and simulation: a modelling perspective. *IMA Journal of Management Mathematics*, 20(1):1–38.

- Dißmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.
- Dupačová, J., Consigli, G., and Wallace, S. W. (2000). Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100(1/4):25–53.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2012). *Modelling extremal events: For insurance and finance*, volume 33 of *Applications of Mathematics : Stochastic Modelling and Applied Probability*. Springer, Berlin and Heidelberg, 4th corr. print. and 9th print edition.
- Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In Dempster, M. A. H., editor, *Risk Management*, pages 176–223. Cambridge University Press.
- Gaivoronski, A. A. and Pflug, G. (2005). Value-at-risk in portfolio optimization: Properties and computational approach. *The Journal of Risk*, 7(2):1–31.
- Gijbels, I. and Herrmann, K. (2018). Optimal expected-shortfall portfolio selection with copula-induced dependence. *Applied Mathematical Finance*, 25(1):66–106.
- González-Pedraz, C., Moreno, M., and Peña, J. I. (2015). Portfolio selection with commodities under conditional copulas and skew preferences. *Quantitative Finance*, 15(1):151–170.
- Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw Gebiete (Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete)*, 40(4):317–337.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., and Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485.
- Hatherley, A. and Alcock, J. (2007). Portfolio construction incorporating asymmetric dependence structures: a user’s guide. *Accounting & Finance*, 47(3):447–472.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3):497–546.
- Hochreiter, R. (2008). Evolutionary stochastic portfolio optimization. In Kacprzyk, J., Brabazon, A., and O’Neill, M., editors, *Natural Computing in Computational Finance*, volume 100 of *Studies in Computational Intelligence*, pages 67–87. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Hochreiter, R. and Pflug, G. C. (2007). Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research*, 152(1):257–272.
- Hofert, M., Frey, R., and McNeil, A. J. (2020). *The Quantitative Risk Management Exercise Book*. Princeton University Press, Princeton - Oxford.

- Hofert, M., Hornik, K., and McNeil, A. J. (2022). Package 'qrmtools'. tools for quantitative risk management.
- Høyland, K., Kaut, M., and Wallace, S. W. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2/3):169–185.
- Høyland, K. and Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307.
- Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141.
- Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1):171–183.
- Kiran, K. G. and Srinivas, V. V. (2021). A mahalanobis distance-based automatic threshold selection method for peaks over threshold model. *Water Resources Research*, 57(1).
- Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management Science*, 37(5):519–531.
- Krokhmal, P., Uryasev, t., and Palmquist, J. (2001). Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk*, 4(2):43–68.
- Kurowicka, D. and Cooke, R. (2010). *Uncertainty analysis with high dimensional dependence modelling*. Wiley, Chichester, England and Hoboken, NJ.
- Lang, M., Ouarda, T., and Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225(3-4):103–117.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.
- Li, D. X. (2000). On default correlation. *The Journal of Fixed Income*, 9(4):43–54.
- Low, R. K. Y., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? *Journal of Banking & Finance*, 37(8):3085–3099.
- Mansini, R., Ogryczak, W., and Speranza, M. G. (2015). *Linear and Mixed Integer Programming for Portfolio Optimization*. EURO Advanced Tutorials on Operational Research. Springer International Publishing and Imprint: Springer, Cham, 1st ed. 2015 edition.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137.
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.

- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton NJ, revised edition edition.
- Morales-Nápoles, O. (2011). Counting vines. In Joe, H. and Kurowicka, D., editors, *Dependence Modeling: Vine Copula Handbook*, pages 189–218. WORLD SCIENTIFIC.
- Nagler, T., Bumann, C., and Czado, C. (2019). Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *Journal of Multivariate Analysis*, 172:180–192.
- Nagler, T. and Vatter, T. (2023). Package ‘rvinecopulib’: High performance algorithms for vine copula modeling.
- Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In Pardalos, P. and Uryasev, S. P., editors, *Probabilistic Constrained Optimization*, volume 49 of *Nonconvex Optimization and Its Applications*, pages 272–281. Springer US, Boston, MA.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.
- Rockafellar, R. and Wets, R.-B. (1977). Measures as lagrange multipliers in multistage stochastic programming. *Journal of Mathematical Analysis and Applications*, 60(2):301–313.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41.
- Rootzén, H., Segers, J., and L. Wadsworth, J. (2018). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.
- Sabourin, A. and Renard, B. (2015). Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data. *Water Resources Research*, 51(12):9646–9664.
- Sahamkhadam, M., Stephan, A., and Östermark, R. (2018). Portfolio optimization based on garch-evt-copula forecasting models. *International Journal of Forecasting*, 34(3):497–506.
- Sahamkhadam, M., Stephan, A., and Östermark, R. (2022). Copula-based black–litterman portfolio optimization. *European Journal of Operational Research*, 297(3):1055–1070.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP*, VIII(3):229–231.
- Smith, R., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.
- Solari, S., Egüen, M., Polo, M. J., and Losada, M. A. (2017). Peaks over threshold (pot): A methodology for automatic threshold estimation using goodness of fit p -value. *Water Resources Research*, 53(4):2833–2849.

Spitznagel, M. (2021). *Safe haven: Investing for financial storms*. John Wiley & Sons Inc, Hoboken New Jersey.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable*. Random House Trade Paperbacks, New York, 2nd ed., random trade pbk. ed. edition.

Yew Low, R. K., Faff, R., and Aas, K. (2016). Enhancing mean–variance portfolio selection by modeling distributional asymmetries. *Journal of Economics and Business*, 85:49–72.