École polytechnique fédérale de Lausanne

Master Thesis

Diversification effects in asymptotically independent portfolios

Author:
Daniel Bischof

Supervisor:
Prof. Anthony C. Davison
Dr. Sebastian Engelke

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in Mathematics

October 11, 2016

Declaration of Authorship

I, Daniel Bischof, declare that this thesis titled "Diversification effects in asymptotically independent portfolios" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at TU Darmstadt.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"*The trouble with weather forecasting is that it's right too often for us to ignore it and wrong too often for us to rely on it.*"

Patrick L. Young

"*There are* 10^{11} *stars in the galaxy. That used to be a huge number. But it's only a hundred billion. It's less than the national deficit! We used to call them astronomical numbers. Now we should call them economical numbers.*"

Richard Feynman

"*Careful. We don't want to learn from this.*"

Bill Watterson, "Calvin and Hobbes"

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Abstract

Chair of Statistics
Institute of Mathematics

Master of Science

Diversification effects in asymptotically independent portfolios

by Daniel Bischof

When it comes to portfolio management, the assumption of normal distributed returns is one of the biggest misconceptions in modern finance. The sheer number of extreme losses is a strong indicator that this assumption is not reasonable. This also implies that the variance of portfolio losses is not a good risk measure, since variance measures data variation around the mean and does not give a robust estimate of the truly large losses, which tend to be more catastrophic for investors. In this thesis we look into assets with regularly varying loss distributions, which have an extensive analytical framework based upon extreme value theory. In real markets, fitting a regular varying distribution often reveals portfolios with infinite variance, so the classical theory developed by Markowitz cannot be applied. We therefore consider the Value-at-Risk, a quantile of the loss distribution, which gives a better way to estimate large losses. Unfortunately, large losses in some markets tend to become more and more independent of each other in the limit, so the assumed limit distribution can be a very bad approximation for the actual asset dependence on a subasymptotic level. New developments in the theory of regular variation, namely hidden regular variation, help to overcome these weaknesses, provide a better estimate for market dependence, and can be used to calculate portfolio weights which minimize the Value-at-Risk and provide an improvement over the classical theory. We calculate an implicit analytical approximation for the Value-at-Risk based on a tail-equivalent surrogate, which we test on linear factor models with simulated data. We find that the model gives reasonable results in a direct calculation of the Value-at-Risk if the hidden component is heavy enough and it turns out to provide good approximations for implicit portfolio weights at large thresholds.

Acknowledgements

I would like to express my special appreciation and thanks to both Prof. Anthony Davison and Dr. Sebastian Engelke for the useful comments, remarks and engagement through the learning process of this master thesis and the interesting and educational discussions, which made every single one of the numerous trips from Zurich to Lausanne worth it. One simply could not wish for better, more competent or friendlier supervisors. I would also like to thank Prof. Michael Kohler of the Department of Mathematics at TU Darmstadt, the second reader of this thesis, and I am grateful for enabling me to write it externally at EPFL.

My special thanks goes to swissQuant Group AG in Zurich, in particular to the Derivatives Technologies team, which kindly allowed me to come back to Zurich during the summer to work part-time with an exciting team on challenging projects, gain more experience in financial markets and financially enable this adventure. Without them granting me access to their computational facilities and endless supply of coffee and fruit, this work would have been a lot less successful.

Finally, I must express my very profound gratitude to my dear friends and to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

List of Figures

Chapter 1

Introduction

In recent years, the frequency and severity of financial crises have exposed major shortcomings of traditional risk assessment methodologies for capturing the risk of rare but damaging events, and have made the search for better approaches to risk modelling and measurement more crucial than ever. This problem is rooted in several wrong or simplistic assumptions about financial markets that have been adopted as standard. The most famous of these assumptions is that market returns are normally-distributed. This assumption goes back to the early 20th century, when Louis Bachelier proposed his option pricing model, based on the assumption that stock markets itself would follow a Brownian motion, and was one of the first persons to describe financial markets mathematically. While not being taken seriously initially, his work regained attention after the 1929 crisis, when investors became painfully aware that while stocks offered a great opportunity for a company to finance itself and for an investor to gain money based on the company's success, a responsible investor had to take risk into account. Moreover, the crisis showed that it is not the idiosyncratic value of the company alone that determines the stock price of a public company, but other factors too. Factors shared with other companies include exposure to common external factors which would influence the company's performance, like the state of the economy or the industry. This was already known, but it became clear that these systemic factors could also be a risk to the investor. At this time, the term 'diversification' was not a well-known term for reducing an investor's risk — the approach back then was to invest just in companies which were considered safe. It was believed that one should rather invest in fewer companies than in more, so that the exposure to sources of risk was kept minimal.

In 1952, Markowitz published his landmark paper on risk-minimizing portfolio theory (see [35]), coming seemingly out of nowhere. He assumed that not markets, but rather market returns, would follow a normal distribution, and showed for the first time that the prevalent practitioners' opinion to invest into fewer stocks rather than more was wrong, if the term "risk" was assumed to be the variance of asset returns. Instead he proposed that

diversification was generally a good thing and that there existed a unique portfolio minimizing this risk. With this paper he became "the father of modern portfolio theory (MPT)" and started a whole series of quantitative developments on the topic of minimizing the risk of a portfolio.

This approach was of course not without criticism. Just a few years later Mandelbrot urged people not to oversimplify markets and market observations. In one of his many papers and discussions about financial theory (see [34]), he analysed financial returns from the beginning of the century until the 60s and noticed that a normal distribution might not be correct, as returns were "too spiked" around zero, leading to either overestimating the variance or underestimating tail-regions. He proposed to rather focus on a Pareto-like distribution for the tails of the return distribution, but also noted that this would lead to increasing difficulties in more sophisticated problems, since the normal distribution is a very well-understood mathematical object and can be handled very flexibly. Another group of people followed this observation and developed a portfolio theory for so-called α-stable distributions. The best-known development in this field was the observation of Fama and Miller (see [16], Chapter 5) that while in most practical situations diversification was advantageous to the risk profile of a portfolio, there exist situations where this is not true, particularly when the returns are too heavy-tailed.

Heavy-tailed distributions like the Pareto distribution fit many of Mandelbrot's observed "stylized facts". While the normal distribution fitted to the center of the return distribution is a good first approximation of market behaviour, it heavily underestimates the probability of rare, but extreme, losses. Besides the crisis in the late 1920s, another often-discussed example is the Black Monday crash. On October 19, 1987, the world's stock markets crashed, shedding a huge value in a very short time. The Dow Jones Industrial Average dropped by 22.61% that day. How likely was such an event, according to the prevailing models of the time? Under the assumption of normal-distributed-returns, this was apparently a 20σ event, which means that the event was twenty standard deviations away from the expected behaviour. If the assumption of normal distributed returns was true, this event would have had a probability of around 10^{-50}, something that happens once in the lifetime of the universe. Just thirty years later the sub-prime mortgage crisis hit the USA and led to another global financial crisis, resulting in a total stock market loss of 6.9 trillion US$ in addition to a cumulative 3.3 trillion US$ loss in home equity from U.S. homeowners.

From an economic point of view, therefore, variance might not be the

best measure to evaluate market risk. The lasting impacts on financial markets, economical stability and on private portfolios are caused by single large-loss events. For example, an investor can receive a margin call by losing his investments due to one single large crash, independent of the general variance of the returns of the portfolio he was holding to that point. Margin accounts allow investors to make investments with their broker's money. They act as leverage and can thus magnify gains. But they also magnify losses, and, in some cases, a brokerage firm can sell an investor's securities without notification or even sue if the investor does not fulfill a margin call. These calls occur in extreme loss cases and may lead to the default of the investor, who is suddenly unable to compensate his losses. A risk measure needs to be informative on how large extreme losses can bes, which the variance is not constructed for.

A more informative risk measure for extreme losses is the Value-at-Risk, essentially the quantile loss of the return distribution, which has become the dominant regulatory tool in the last twenty years in Basel II and Basel III. Instead of evaluating what is happening around the center of the distribution by calculating the mean distance of a sample point from the mean, the Value-at-Risk ignores everything below a given threshold. The philosophy of the investor therefore switches from reducing noise to stabilize the portfolio against large losses, which has become more and more important to investors, banks and regulators in recent years.

This is the ideal setup for the use of regularly varying distributions, which are essentially semi-parametric generalizations of the Pareto distribution, and on which we focus in this thesis. Such distributions have had a lot of attention in financial modelling in the recent years since they fit "stylized facts" like the observations from Mandelbrot pretty well, they are generally a very flexible and broad class, and tail risk measures like the Value-at-Risk have a very natural representation in this class of distributions. To evaluate the risk of a whole portfolio, not just a single asset, it is also important to understand how the assets in a portfolio are related. As we will explain, regularly varying distributions have very nice stability properties and can be analyzed very well by multivariate extreme value theory. This is both a blessing and a curse. While we will show that the dependence of assets has some stability properties that can be used as an approximate solution for large quantiles, we will see that the limit dependence given by multivariate extreme value theory does not need to be related to the dependence at finite levels.

In this thesis we will focus on regularly varying portfolios which possess

asymptotic independent assets, meaning that the dependence between the assets vanishes in the limit. As mentioned by Poon et al. (see [40]), this is a valid assumption in many markets and provides a very challenging situation for the Quantitative Analyst seeking portfolio optimization strategies. This does not mean that we have weak dependence at finite levels; in fact there are model classes with arbitrarily strong dependence at finite levels, but which converge to independent assets in the limit. Recent developments have enabled statisticians to find solutions to this problem, and one of them will be analyzed in this thesis. Based on these developments, we will propose an analytical solution, which approximates the Value-at-Risk of a portfolio possessing asymptotic independence and which can be used in a portfolio optimization problem.

In Chapter 2 we give a short introduction to regularly varying functions and provide essential tools. We also collect results on univariate extreme value theory and show that these have very useful stability properties. In Chapter 3 will extend this to the multivariate case, where we need to elaborate how the dependence between the components of a random vector affects the limit distribution. Chapter 4 is dedicated to the financial application; here we will transfer the collected results to a portfolio context and develop our main theorem. Chapter 5 is dedicated to a simulation study. We will assume linear factor models as the dependence structure between the assets and show how our model performs and how it behaves in certain situations. In Chapter 6 we will talk about how this framework can be extended and give a view of current research.

Chapter 2

Univariate theory

2.1 A short course on regular variation

2.1.1 Definition and class-uniqueness

The notion of regular variation, which was introduced by J. Karamata in 1930, has been proven to be fruitful and important in many connections and finds an ever-increasing number of applications in fields like the analysis of differential equations and modern probability theory. Reasons for this are that the class of regular varying functions provide useful stability properties that have been used where the behaviour of the limit of a function needs to be assessed.

We are interested in a class of possibly diverging measurable functions, whose scaled limits grow like non-trivial, finite functions. If this limiting behaviour is accessible, we do not need to know the exact properties of those functions in the limit, but rather just their asymptotic behaviour. It turns out that the class of possible limit behaviours is surprisingly small.

Theorem 1 (Regular varying functions)**.** *Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a measurable, eventually positive function, such that*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = \phi(x) \tag{2.1}$$

exists for a finite, positive function ϕ for all x on a dense set of points. Then, for a unique $\alpha \in \mathbb{R}$,

$$\phi(x) = x^\alpha, \quad x > 0.$$

Proof. The proof of this central theorem and extensive discussion of it can be found in [4]. □

Functions in this class are called regularly varying.

Definition 1. *A eventually positive Lebesgue measurable function $f : \mathbb{R} \to \mathbb{R}$ is regularly varying (at infinity), if for some $\alpha \in \mathbb{R}$,*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha, \quad x > 0.$$

Notation: $f \in RV_\alpha$. The index α is called the index of regular variation. If $\alpha = 0$, we call f slowly varying.

Regularly varying functions have useful properties, of which we will provide some central ones in the following paragraphs. Since we will focus on applications to probability theory in this thesis, we focus on key results for measurable functions, although regular variation is not limited to real-valued measurable functions. In later chapters we will focus on random variables with regularly varying survival functions and more general measures, where we will extend the approach developed here to a multivariate and set-theoretic framework. Since the class uniqueness of regularly varying functions relies on Lusin's theorem, which states that for any measurable function f there exists a compact set K such that f is continuous on it, it is not trivial to think that non-continuous functions, particularly sequences, can vary regularly too. Nevertheless, Bojanic and Seneta established equivalent existence, uniqueness and representation theorems for regular varying sequences.

Definition 2. *A sequence $(c_n)_{n\in\mathbb{N}}$ of eventually positive numbers is regularly varying if*

$$\lim_{n\to\infty} c_{[tn]}/c_n = t^\alpha, \quad t > 0. \tag{2.2}$$

Further and extended discussion of the mathematical properties and consequences of this can be found in [21], [43] and [4].

2.1.2 Karamata theory

We can recognize from the limit behaviour in (2.1) that a function which is regularly varying must behave like a perturbed monomial of order α. This was recognized by Karamata, who proposed the following characterisation and representation theorems. As preparation, we first show that (2.1) holds uniformly on compact intervals in $\mathbb{R}$. This is an important result, since it implies that we can actually access the limiting behaviour for regularly varying functions on finite intervals (hence, the property of regular variation is not just existent in the limit).

Proposition 1 (Uniform convergence theorem)**.** *If $f \in RV_\alpha$, then relation* (2.1) *holds uniformly for $x \in A \subset \mathbb{R}$ with A compact.*

Proof. There are several ways to prove this. For a broad discussion of this theorem, we refer to [4]. □

With this, one can show the following important property of regularly varying functions.

Theorem 2 (Karamata's Theorem)**.** *Suppose $f \in RV_\alpha$. There exists $t_0 > 0$ such that $f(t)$ is positive and locally bounded for $t \geqslant t_0$. If $\alpha \geqslant -1$, then regular variation of f is equivalent to*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_{t_0}^{t} f(s)ds} = \alpha + 1, \tag{2.3}$$

and, if $\alpha < -1$ or $\alpha = -1$ and $\int_{\mathbb{R}^+} f(s)ds < \infty$,

$$\lim_{t\to\infty} \frac{tf(t)}{\int_{t}^{\infty} f(s)ds} = -(\alpha + 1), \tag{2.4}$$

Proof. This proof can be found in [21] and in [4]. □

Theorem 2 tells us that integrals of regular varying functions behave similarly to integrals of mononomes. In fact, every regular varying function is a mononome perturbed by a slowly varying function.

Corollary 1. *If (and only if) $f \in RV_\alpha$, then there exists a function $l \in RV_0$, such that*

$$f(x) = l(x)x^\alpha, \quad x \to \infty. \tag{2.5}$$

Proof. The result follows by explicitly calculating the regular variation property with Theorem 1. □

This implies that regularly varying functions have a specific representation through a mononome and a multiplicative residual term. This is still a very coarse way to represent those functions, as slowly varying functions might grow asymptotically more slowly than any mononome but could have arbitrary forms. The following theorem shows that this is not the case and that both slowly and regularly varying functions always have a specific form, which can be deducted easily from Karamata's Theorem.

Proposition 2 (Karamata's representation Theorem)**.** *If $f \in RV_\alpha$, there exist measurable functions $a : \mathbb{R}^+ \to \mathbb{R}$ and $c : \mathbb{R} \to \mathbb{R}$ with*

$$\lim_{t\to\infty} c(t) = c_0 \quad \text{with } 0 < c_0 < \infty,$$
$$\lim_{t\to\infty} a(t) = \alpha,$$

and $t_0 > 0$, such that for $t > t_0$,

$$f(t) = c(t) \exp\left(\int_{t_0}^{t} \frac{a(s)}{s} ds\right). \tag{2.6}$$

Furthermore, if equation (2.6) holds with a and c satisfying the properties above, then $f \in RV_\alpha$.

Proof. This follows by applying Karamata's Theorem to equation (2.6) and can be found in [11]. □

2.1.3 Useful properties of regularly varying functions

Karamata's way of representing and describing regularly varying functions paves the way for a vast amount of structure in the functional set. Through the definition of smooth regular variation, i.e., the subset of regularly varying functions that fulfill $\exp(f(\log t)) \in C^\infty$, one can define an algebra of equivalence classes for examples and use this to develop strong solutions for ordinary and partial differential equations. This gives a very convenient environment for mathematical analysis, but we don't need this assumption here. Still, one can develop a list of useful properties related to closedness of this functional set, of which we will provide a few in this section. We will give no proofs, since they are mostly direct consequences of the theorems. More detailed proofs and discussions are found in [43] and [11].

Proposition 3. *The following properties hold for regularly varying functions. If*

1. $f \in RV_\alpha, g \in RV_\beta$, *then* $f + g \in RV_{\max(\alpha,\beta)}$;
2. $f \in RV_\alpha, g \in RV_\beta$ *and* $\lim_{x\to\infty} g(x) = \infty$, *then* $f \circ g \in RV_{\alpha\beta}$;
3. $f \in RV_\alpha$, $\alpha \neq 0$, *then there exists a version* f^* *that is absolutely continuous, strictly monotone and fulfils*

$$f \sim f^*, \ x \to \infty;$$

4. $f \in RV_0$, *then there exists a version* f^* *that is absolutely continuous and fulfils*

$$f \sim f^*, \ x \to \infty.$$

   If f *is eventually monotone, so is* f^*;

5. *(Potter, 1942)* $f \in RV_\alpha$, *then there exists* $t_0 = t_0(\delta_1, \delta_2)$ *for arbitrary* $\delta_1, \delta_2 > 0$, *such that for* $t \geqslant t_0, tx \geqslant t_0$;

$$(1-\delta_1)x^\alpha \min(x^{\delta_2}, x^{-\delta_2}) < \frac{f(tx)}{f(t)} < (1+\delta_1)x^\alpha \max(x^{\delta_2}, x^{-\delta_2}). \quad (2.7)$$

A natural topic to elaborate is the following question: since we have overall perturbed power law behaviour, can we define an inverse of a regularly varying function? Property 5 in Proposition 3 implies that regularly varying functions are asymptotically equivalent to a monotone function, which supports the idea of seeking an inverse. To show this we first define the generalized inverse for a function in form of a category-theoretic retraction. Throughout this thesis, we understand increasing in the sense of

non-decreasing, i.e., f is increasing if $f(x) \leqslant f(y)$ for all $x < y$. The following definition captures the notion of an inverse for such functions. Note that evaluating increasing functions at the symbols $-\infty$ or ∞ is always understood as the corresponding limit (possibly $-\infty$ or ∞).

Definition 3 (Generalized inverse function)**.** *For an increasing real-valued function $f : \mathbb{R} \to \mathbb{R}$ with $f(-\infty) = \liminf_{x\to-\infty} f(x)$ and $f(\infty) = \limsup_{x\to\infty} f(x)$, the generalized inverse $f^{\leftarrow} : R \to \bar{\mathbb{R}} = [-\infty, \infty]$ of T is defined by*

$$T^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : T(x) \geqslant y\}, \ y \in \mathbb{R}, \tag{2.8}$$

with the convention that $\inf \varnothing = \infty$.

If T is continuous and strictly increasing then $T^{\leftarrow}$ coincides with the ordinary inverse of T. Such a function exists for regularly varying function in an asymptotic sense.

Proposition 4. *If $f \in RV_\alpha$ with $\alpha > 0$, there exists $g \in RV_{1/\alpha}$, such that*

$$f(g(x)) \sim g(f(x)) \sim \mathrm{id}_{\mathbb{R}}, \quad x \to \infty.$$

Here g is determined uniquely to within asymptotic equivalence, and one version of g is $f^{\leftarrow}$.

Proof. A proof can be found in [4]. By Propositions 3, 2 and equation (2.5), we know that f is asymptotically equivalent to a function $f_1(x) = x^\alpha l(x)$ with a smooth slowly varying function defined in Proposition 3. The argument then follows by a nested interval argument. □

We conclude that if we have a non-trivially regularly varying function, we will find an asymptotically unique generalized inverse function for it. For the case $\alpha = 0$, this fails. We can nevertheless still define a multiplicative inverse as given by de Bruijn.

Lemma 1. *If l varies slowly, there exists a slowly varying function $l^{\#}$, unique up to asymptotic equivalence, with*

$$l(x)l^{\#}(xl(x)) \to 1,$$
$$l^{\#}(x)l(xl^{\#}(x)) \to 1,$$

then $l^{\#\#} \sim l$.

Proof. Existence and asymptotic uniqueness of $l^{\#}$ follow from Proposition 4 by taking $\alpha = 1, f(x) = xl(x), g(x) = xl^{\#}(x)$ and symmetry. □

2.1.4 Further extensions

The concept of regular variation has various extensions and generalizations with a very broad range of literature covering the topic, the most prominent being the book of Bingham, Goldie and Teugels, published in 1989 (see [4]) where the content of this chapter is discussed very extensively. We will therefore just briefly discuss the class of extended regularly varying functions (denoted ERV_α), which is a generalization of standard regular variation, which formulates this property somewhat differently. After that we will focus on second order (extended) regular variation, which will be very important later.
We begin with a proposition that generalizes Theorem 1.

Proposition 5 (Extended regularly varying functions)**.** *Suppose $f : \mathbb{R}^+ \to \mathbb{R}$ is measurable and a is positive. If the limit*

$$\lim_{t\to\infty} \frac{f(tx) - f(t)}{a(t)} = \phi(x)$$

exists for all $x > 0$, then

$$\phi(x) = c\frac{x^\alpha - 1}{\alpha}, \quad x > 0,$$

for a specific α.

Proof. The proof can be found in [4], [21] and [11]. □

We say f is extended regularly varying, written $f \in ERV_\alpha$ for $\alpha \neq 0$, if α is zero, we consider $\phi(x) = \log(x)$ and say that f belongs to the class Π, written $f \in \Pi$. The function a is called the auxiliary function for f. This can be seen as a generalization of standard regular variation by setting $a(t) = \alpha f(t)$. This formulation concentrates on fixed differences rather than on fixed proportionality and therefore gives a slightly different perspective on regularly varying functions. This can somewhat be seen as differentiability at infinity, which brings an additional motivation for continuous and differentiable versions in earlier propositions. The interrelation between the two function classes is given by the next proposition.

Proposition 6. *Suppose $f \in ERV_\alpha$ and f fulfills the properties in Proposition 5.*

1. *If $\alpha > 0$, then $\lim_{t\to\infty} \frac{f(t)}{a(t)} = \frac{c}{\alpha}$, and hence $f \in RV_\alpha$.*
2. *If $\alpha < 0$, then $f(\infty) = \lim_{t\to} f(t)$ exists finite, $\lim_{t\to\infty} \frac{f(\infty)-f(t)}{a(t)} = \frac{-c}{\alpha}$ and hence $f(\infty) - f(t) \in RV_\alpha$.*

Proof. After adapting a corresponding version of the uniform convergence Theorem for ERV_α, the proof works analogously to the proof of Theorem 1 and can also be found in [4], [21] and in [11]. □

Another interesting question which will be central in the scope of this work is the following — suppose that we have $f \in RV_\alpha$ or $f \in ERV_\alpha$. We know from Theorem 1 and from Proposition 5 that we have a limit relationship, but we have no information about the convergence speed towards the limit, which will turn out to be a large flaw of the probabilistic application in later chapters. It is therefore useful to build a theory in which the speed of convergence is the same for all $x > 0$, i.e., there exists a positive function A that converges to zero eventually such that

$$H(x) = \lim_{t\to\infty} \frac{\frac{f(tx)-f(t)}{a(t)} - \frac{x^\alpha-1}{\alpha}}{A(t)}$$

exists finite for all $x > 0$. We exclude the case $H(x) = c\frac{x^\alpha-1}{\alpha}$, since the limit relation yields no real additional information. We can identify the general form of H with the next theorem.

Theorem 3. *Suppose that for some measurable functions f and positive functions a and A the limit*

$$H(x) = \lim_{t\to\infty} \frac{\frac{f(tx)-f(t)}{a(t)} - \frac{x^\alpha-1}{\alpha}}{A(t)}$$

exists. Then there exist real constants c_1, c_2 and a parameter $\rho \leqslant 0$ such that for $x > 0$,

$$H(x) = \begin{cases} \frac{1}{\rho}\left(\frac{x^{\alpha+\rho}-1}{\alpha+\rho} - \frac{x^\alpha-1}{\alpha}\right), & \rho \neq 0, \\ \frac{1}{\alpha}\left(x^\alpha \log(x) - \frac{x^\alpha-1}{\alpha}\right), & \rho = 0, \alpha \neq 0, \\ \frac{1}{2}\log x, & \rho = \alpha = 0, \end{cases}$$

$$\lim_{t\to\infty} \frac{\frac{a(tx)}{a(t)} - x^\alpha}{A(t)} = c_1 x^\alpha \frac{x^\rho - 1}{\rho},$$

$$\lim_{t\to\infty} \frac{A(tx)}{A(t)} = x^\rho.$$

Proof. Similarly to Theorem 1, the proof evolves around solving a functional equation for H resulting from an assumed identity. The full proof can be found in [11]. □

This yields the following definition.

Definition 4. *A eventually positive Lebesgue measurable function $f : \mathbb{R} \to \mathbb{R}$ is second-order extended regularly varying (at infinity), if for some $\alpha \in \mathbb{R}$*

and $\rho \leqslant 0$*,*

$$\lim_{t\to\infty} \frac{\frac{f(tx)-f(t)}{a(t)} - \frac{x^\alpha - 1}{\alpha}}{A(t)} = \begin{cases} \frac{1}{\rho}\left(\frac{x^{\alpha+\rho}-1}{\alpha+\rho} - \frac{x^\alpha-1}{\alpha}\right), & \rho \neq 0, \\ \frac{1}{\alpha}\left(x^\alpha \log x - \frac{x^\alpha-1}{\alpha}\right), & \rho = 0, \alpha \neq 0, \\ \frac{1}{2}\log x, & \rho = \alpha = 0. \end{cases}$$

Notation: $f \in 2ERV_{\alpha,\rho}$*. The index* ρ *is called the second-order index of regular variation. If* $\alpha = 0$*, we call* f *slowly varying.*

In some cases we can relate 2ERV functions to classes of functions we have already covered in this chapter. The following theorem shows what has to be fulfilled.

Theorem 4. *Suppose that the function* f *satisfies theorem 3. Then*

1. *if* $\alpha = \rho = 0$*,*

$$f(t) = h(t) + \int_0^t \frac{h(s)}{s} ds,$$

with $h \in \Pi$*;*

2. *if* $\rho = 0,\ \alpha \neq 0,\ \pm t^{-\alpha}\tilde{f}(t) \in \Pi$*, where*

$$\tilde{f}(t) = \begin{cases} f(t), & \alpha > 0, \\ f(\infty) - f(t), & \alpha < 0; \end{cases}$$

3. *if* $\rho < 0$*, there exists a nonzero constant* c *such that*

$$\pm\left(f(t) - c\frac{t^\alpha - 1}{\alpha}\right) \in ERV_{\alpha+\rho}.$$

Proof. The proof is a technical expansion of Theorem 3 and Proposition 6; we will skip it here. For more details regarding this theorem, we refer to [4], where a proof can be found. □

One can even show an equivalence for the claims in Theorem 4, which is done in [11]. Together with Theorem 4, we conclude that a first-order limit is not the only limit we can derive from a regularly varying function: we can access the speed of convergence as well by a second-order approximation. This is not the best approximation; there are possibilities to derive third- and higher-order effects as well, but we will stay at second-order convergence in the present work. We could further show that second-order effects imply different kinds of regularly varying and extended regularly varying properties. If one translates this to the multivariate setting, as will be done in later chapters, we will see that second order effects are helpful in providing finer representations for regularly varying functions when the image of a function is largely dominated by the behaviour of first-order limits, thus

losing all higher-order behaviour. For application, this might be too coarse for a representation of the function behaviour. We will now introduce regularly varying functions into a probabilistic setting and derive a couple of fundamental univariate stability properties.

2.2 Extreme value theory in one dimension

2.2.1 A primer on heavy-tailed distributions

In this thesis we will work with distributions that provide certain properties in the (right) tail, since we will look at the so-called Value-at-Risk in later chapters. In this first section we will provide a short overview about heavy-tailed distributions and motivate their use.

Heavy tails are a characteristic of phenomena where the probability of a huge value is relatively big. Record-breaking insurance losses, financial log-returns, sizes of files stored on a server and transmission rates of files are all examples of heavy-tailed phenomena. We call a distribution F on $\mathbb{R}$ heavy-tailed if

$$\int_{-\infty}^{\infty} e^{\lambda x} F(dx) = \infty, \quad \lambda > 0, \tag{2.9}$$

that is, if and only if F has no positive exponential moments. Otherwise F is said to be light-tailed. This implies that the survival function (also called the tail function) $\bar{F}$ is not bounded by any exponentially decreasing function. Another consequence is that not all (if any) moments of a heavy-tailed distribution have to be finite, which gives the following characterizing proposition.

Proposition 7. *Let $y \geqslant 0$ be a random variable with a heavy-tailed distribution. Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a concave function such that $E[e^{f(y)}] = \infty$. Let $g : \mathbb{R}^+ \to \mathbb{R}$ be such that $\lim_{x\to\infty} g(x) = \infty$. Then there exists a concave function $h : \mathbb{R}^+ \to \mathbb{R}$ such that*

$$\begin{aligned} &h \leqslant f, \\ &E[e^{h(y)}] < \infty, \\ &E[e^{h(y)+g(y)}] = \infty. \end{aligned}$$

Proof. See [13]. □

This is of course a very broad definition and, as a deeper analysis of this class of distributions can show, there are several subclasses to consider, like long-tailed distributions and fat-tailed distributions. From a modelling perspective a very successful approach in risk application is to consider the

subclass of regular varying distributions with tail coefficient $-\alpha < 0$, elaborated in the following definition.

Definition 5 (Regularly varying random variables)**.** *A nonnegative random variable X and its distribution are said to be regularly varying with index $\alpha > 0$ if the right distribution tail $\bar{F}$ is regularly varying with index $-\alpha$ that is, $\bar{F} \in RV_{-\alpha}$.*

We will focus for the rest of this thesis on this particular subclass of heavy-tailed distributions. The biggest reason for this will be explained in Chapter 3, where we will show that we can extend the theory of regular variation to random vectors.

2.2.2 Fundamentals of extreme value theory

Extreme value theory is a branch of statistics dealing with extreme deviations from the median of probability distributions. It seeks to assess the probability of events that are more extreme than any previously observed. The central and interesting result of this section is the answer to the following question: suppose one is interested in the distribution of the maxima of a series of random variables. If we have unbounded upper support, i.e., $\bar{F}(x) > 0$ for all $x \in \mathbb{R}$, then there is always a non-zero probability that the maximum surpasses every finite threshold. This is especially of interest if we have probability distributions that possess relatively high probabilities at very high levels, such as heavy-tailed distributions. The question is, can we scale those maxima in a way that we obtain a non-degenerate distribution and can hence access the properties of the extreme values of that sequence? More precisely, consider $X_1, \ldots, X_n$, i.i.d. random variables with common distribution function F. The maximum is defined as

$$M_n = \bigvee_{i=1}^{n} X_i = \max(X_1, X_2, \ldots, X_n).$$

Clearly

$$\begin{aligned} P\left(M_n \leqslant x\right) &= P\left(X_1 \leqslant x, X_2 \leqslant x, \ldots, X_n \leqslant x\right) \\ &= \prod_{i=1}^{n} P\left(X_i \leqslant x\right) \\ &= \prod_{i=1}^{n} F(x) \\ &= F^n(x). \end{aligned}$$

What are the conditions on F so that there exist so-called normalization sequences a_n and b_n such that

$$P\left(\frac{M_n - a_n}{b_n} \leqslant x\right) = F^n(b_n x + a_n) \to G(x), \quad x \in \mathbb{R}, n \to \infty \tag{2.10}$$

for a non-degenerate distribution G? What properties does G possess and how can we characterize the normalization sequences?

The surprising result is that this is indeed possible and, more surprising that the possible limit distributions are limited to just three possible cases. The first half of the answer, existence and classification of a limit distribution, is known as the central extreme value theorem or the Fisher–Tippett–Gnedenko theorem. Credit for it is given to Boris Gnedenko, who could state necessary conditions for a distribution belonging to a limit class, though previous versions were stated by Ronald Fisher and Leonard Henry Caleb Tippett in 1928 and Maurice René Fréchet in 1927. Before we state the central extreme value theorem, we first introduce some background, which brings up parallels with the theory of stable distributions. The proof of the theorem is rather technical and we will focus more on some specific results; we will instead provide the main tools needed to understand the theorem and its mathematical basis. Full proofs can be found in the sources.

Definition 6 (Classification types of distributions)**.** *Let X and Y be random variables with distributions μ and ν respectively. We say that μ and ν are of the same type if there exist $a > 0$ and $b \in R$ such that $aX + b$ has the same distribution as Y.*

An important tool for classifying the possible limit distributions in equation (2.10) is the following theorem due to Khintchine, which allows us to provide sequence-linear versions of limit distributions.

Theorem 5 (Convergence to Types)**.** *Let W_n be a sequence of random variables converging weakly to W, and for some $a_n > 0$ and $b_n \in \mathbb{R}$, let $a_n W_n + b_n \to W'$, where both W and W' are non-degenerate. Then $a_n \to a$ and $b_n \to b$ for some $a > 0$ and $b \in \mathbb{R}$. Equivalently, if F_n, F and F' are distribution functions with F and F' non-degenerate, and there exist $a'_n, a_n > 0$ and $b_n, b'_n \in \mathbb{R}$ such that $F_n(a_n x + b_n) \to F(x)$ and $F_n(a'_n x + b'_n) \to F'(x)$ at all points of continuity of F, respectively F', then $a_n/a'_n \to a$, $(b_n - b'_n)/a'_n \to b \in \mathbb{R}$ and $F'(ax + b) = F(x)$ for all $x \in \mathbb{R}$.*

The next step is now to characterize all non-trivial limiting distributions of $(M_n - a_n)/b_n$. Analogous to the Central Limit Theorem, we need the limit distribution to have essential stability properties.

Definition 7 (Max-stability). *A non-degenerate probability distribution μ on $\mathbb{R}$ is called max-stable if, for a sequence of i.i.d. random variables $(X_i)_{i\in\mathbb{N}}$ with distribution μ, and for each $n \in \mathbb{N}$, there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that $(M_n - a_n)/b_n$ also has distribution μ. Equivalently, μ is max-stable if its distribution function $F(x) := \mu(-\infty, x]$ satisfies: for each $n \in \mathbb{N}$, there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that $F^n n(a_n x + b_n) = F(x)$ for all $x \in \mathbb{R}$.*

This leads to

Theorem 6 (Fisher–Tippett–Gnedenko). *Let $X_1, \ldots, X_n$ be independent and identically-distributed random variables, and let $M_n = \max\{X_1, \ldots, X_n\}$. If a sequence of pairs of real numbers (a_n, b_n) exists such that $a_n > 0$ and $\lim_{n\to\infty} P\left((M_n - b_n)/a_n \leqslant x\right) = \mu(-\infty, x)$, where $G(x) := \mu(-\infty, x)$ is a non-degenerate distribution function, then the limit distribution G belongs to either the Gumbel, the Fréchet or the Weibull family,i.e.,*

- $G(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$, *(Gumbel),*
- $G(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0, \\ 0, & x \leqslant 0, \end{cases}$ *(Fréchet),*
- $G(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x < 0, \\ 1, & x \geqslant 0, \end{cases}$ *(Weibull),*

Proof. Proofs can be found in books about extreme value theory; we refer here to [28], [11] and [44]. The essential strategy of the proof is that the normalizing constants a_n for max-stable distributions can only be of the form $a_n = n^{\phi}$ for some $\phi \in \mathbb{R}$, with $\phi < 0, \phi = 0$ and $\phi > 0$ corresponding to the three types of distributions. □

This is an important result and led to a large amount of applications and ongoing research. We say that a distribution F for which $F^n(x)$ converges towards one of these extreme value distributions for all $x \in \mathbb{R}$ as $n \to \infty$ is in the Domain of Attraction of one of the limit distributions described in Theorem 6.

Definition 8 (Domain of Attraction). *A distribution μ is said to be in the domain of attraction of an extreme value type distribution ν (either Gumbel, Fréchet, or Weibull), denoted by $\mu \in D(\nu)$ for measures or equivalently $F \in D(G)$, if there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that the distribution of $(M_n - a_n)/b_n$ converges weakly to ν.*

What is yet unclear is what conditions F must fulfill for each of those limit distributions. We will mostly focus on the Fréchet case for a reason that is immediately clear from the next theorem.

Theorem 7 (Gnedenko, 1943)**.** *Let G_α be the Fréchet distribution with shape parameter α. Then*

$$F \in D(G_\alpha) \iff \bar{F} \in RV_{-\alpha}$$

with $b_n = 0$ and $a_n = \left(\frac{1}{1-F}\right)^{\leftarrow}(n)$.

Proof. A proof is provided in [44]. The idea is to use an argument similar to Proposition 2 and show that $F \in RV_{-\alpha}$ implies that $F \in D(G_\alpha)$. The other implication follows by calculating that $1/(1-F) \in RV_\alpha$, which implies $1 - F \in RV_{-\alpha}$. □

The significance of this is immediately apparent. There are similar results for the other two classes, which we will not show in detail. More generally, the three cases in Theorem 6 can be collected into one distribution, called the generalized extreme value distribution (GEV). So distributions that lie in the Domain of Attraction of one of those classes lie in the Domain of Attraction of the GEV-distribution, which leads to the following proposition.

Proposition 8. *For $\gamma \in \mathbb{R}$ the following two statements are equivalent:*

1. *there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-(1+\gamma x)_+^{-1/\gamma});$$

2. *there exists a positive function f such that*

$$\lim_{x \to x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1+\gamma x)_+^{-1/\gamma}$$

for all x for which $1 + \gamma x > 0$, where $x^* = \sup_{x\in\mathbb{R}} \{x : \ F(x) < 1\}$.

Proof. The proof will be left out. For reference, see [11]. □

We see that for a distribution to be max-stable is equivalent to being a very generalized form of regularly varying. It's immediately clear that for the Fréchet case, which we will mostly consider in this thesis, one chooses $\alpha = 1/\gamma$, $f(t) = \gamma t$ and $y = 1 + \gamma x$.

This concludes this chapter on univariate extreme value theory. A major expansion of this is described in the next chapter, the move to the multivariate setting, before we proceed to apply the collected theoretical groundwork. We refer to [44] and [11] for a more detailed discussion on this topic and further deeper results.

Chapter 3

Multivariate theory

3.1 Extending to multivariate extremes

3.1.1 Max-stable limit distributions in the multivariate setting

We will now extend the theory from the previous chapters to a multivariate setting and will consider a random vector $\mathbf{X}$ on $\mathbb{R}^d$. It is not obvious how one should define the maximum of a set of vectors. In fact we know from order theory that there are many preorders " $\leqslant$ " one could define, but it turns out that a very naive definition leads to a theory rich enough to tackle applications. We will use a component-wise order relation to bring a partial ordering into $\mathbb{R}^d$, meaning that for $\mathbf{x} = (x_1, \dots x_d)$ and $\mathbf{y} = (y_1, \dots y_d)$, we have

$$\mathbf{x} \leqslant \mathbf{y} \quad \Leftrightarrow \quad x_i \leqslant y_i \quad i \in \{1, \dots d\}.$$

which implies that we can just take the maximum of each coordinate separately and assemble those maxima into a vector to define $M_n = \max_{1 \leqslant i \leqslant n} \mathbf{X}_i$. To keep the notation short we note the d-dimensional rectangle for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ as

$$[\mathbf{x}, \mathbf{y}] = [x_1, y_1] \times \ldots \times [x_d, y_d].$$

We will generalize the extreme value distributions first.

Suppose $(X_1^{(1)}, \dots, X_1^{(d)}), (X_2^{(1)}, \dots, X_2^{(d)}), \dots$ are independent and identically distributed random vectors with joint distribution F. Analogous to the previous chapter, we assume the existence of vector-valued sequences $a_n > 0$ and $b_n \in \mathbb{R}^d$ such that

$$\lim_{n \to \infty} P\left(\frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leqslant x_1, \dots, \frac{M_n^{(d)} - b_n^{(d)}}{a_n^{(d)}} \leqslant x_d \right) = G(x_1, \dots x_d), \quad (3.1)$$

for a distribution function G with nondegenerate marginals on all its continuity points. Any limit distribution G that results from equation (3.1) is called a multivariate extreme value distribution.

Equation (3.1) also implies convergence of the marginals via

$$\lim_{n\to\infty} P\left(\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \leqslant x_i\right) =: G(\ldots, \infty, x_i, \infty, \ldots),$$

meaning that the marginals have to be univariate max-stable distributions covered in the last chapter. We can therefore identify $a_n > 0$ and $b_n \in \mathbb{R}^d$ with the vectorized version of the single marginal scaling functions and set in its general form, as before

$$G(\ldots, \infty, x, \infty, \ldots) = \exp(-(1+\gamma_i x)_+^{-1/\gamma_i}), \quad x \in \mathbb{R}.$$

It is obvious that the random variable

$$Y_i = (1+\gamma_i X_i)^{1/\gamma_i}$$

follows a unit Fréchet distribution, so we can (through the transformation $X_i = \frac{Y_i^{\gamma_i}-1}{\gamma_i}$) transform any nondegenerate multivariate extreme value distribution to a standardized extreme value distribution

$$G_0(x_1, \ldots, x_d) := G\left(\frac{x_1^{\gamma_1}-1}{\gamma_1}, \ldots, \frac{x_d^{\gamma_d}-1}{\gamma_d}\right), \quad x_1, \ldots x_d \in \mathbb{R}. \tag{3.2}$$

We can reach standardization not just by standardizing the limit distribution, but by standardizing the joint distribution function F. Suppose that $U_i(x) = F_i^{\leftarrow}(1-\frac{1}{x})$ is the generalized inverse of the i-th marginal distribution. Then

$$\lim_{n\to\infty} F^n(U_1(nx_1), \ldots U_d(nx_d)) = G_0(x_1, \ldots, x_d). \tag{3.3}$$

If such a limit distribution exists, we can reformulate the convergence in equation (3.3) via the following corollary.

Corollary 2. *For any $(x_1, \ldots, x_d)$ for which $0 < G_0(x_1, \ldots, x_d) < 1$, we have the alternative convergence condition*

$$\lim_{n\to\infty} n\,(1 - F(U_1(nx_1), \ldots U_d(nx_d))) = -G_0(x_1, \ldots, x_d)$$

Proof. We first take logarithms on both sides of equation (3.3) and get

$$\lim_{n\to\infty} -n \log(F(U_1(nx_1), \ldots U_d(nx_d)) = -\log G_0(x_1, \ldots, x_d)$$

Since F is a probability distribution, we immediately get that

$$\frac{-\log(F(U_1(nx_1), \ldots U_d(nx_d))}{1 - F(U_1(nx_1), \ldots U_d(nx_d)} \to 1, \quad n \to \infty,$$

which implies the result. □

A fundamental question in this chapter is the general representation for multivariate extreme value distributions. The Fisher–Tippett–Gnedenko theorem (see Theorem 6), or, more generally, Proposition 8 gives a compact formulation of the possible limit distributions. In the multivariate setting we have to incorporate that there could be any dependency between the single components from complete independence to complete dependence. To do so, we will examine the implications of equation (3.1) and what possible forms result from doing so. We first recognize that $F^n(x)$ must be nondegenerate in the limit, which is not trivial for every distribution. Distributions which satisfy this condition are called max-infinitely divisible.

Definition 9 (Max-infinitely divisible distributions)**.** *The distribution function F on $\mathbb{R}^d$ is max-infinitely divisible (max-id) if for every n there exists a distribution F_n on $\mathbb{R}^d$ such that*

$$F(x) = F_n^n(x), \quad x \in \mathbb{R}_d,$$

i.e., $F^{1/n}$ is a distribution.

By this definition, we will need to analyse the sequence of $F_n = F^{1/n}$, determine under what circumstances a limit exists, and how general measure spaces can be formulated based on the properties of the problem.

3.1.2 Vague topology

The natural way to talk about multivariate distributions and their limits is via probability measures. We will therefore, before we proceed, briefly introduce some basic concepts that eventually lead to a usable basic tool-set for dealing with multivariate extremes.

Definition 10 (Radon measures and Radon measure spaces)**.** *A measure μ defined on the σ-algebra $\mathcal{B}(\mathbb{E})$ of Borel sets of a topological Hausdorff space $\mathbb{E}$ which is locally finite (i.e., for any point $x \in \mathbb{E}$ there is a neighborhood with finite measure) and having the property*

$$\mu(B) = \sup\{\mu(K) : \ K \subset B, \ K \text{ compact}\}$$

is called a Radon measure. If $\mathbb{E}$ is locally compact, every finite Radon measure on $\mathbb{E}$ is also outer regular, i.e.,

$$\mu(B) = \inf\{\mu(K) : \ K \supset B, \ K \text{ open}\}.$$

The first property is called inner regularity or tightness of the measure μ, whereas the second property is called outer regularity. If $\mathbb{E}$ has a countable basis, Radon measures as defined above are necessarily σ-finite. The space

of Radon measures over $\mathbb{E}$ *is defined as*

$$M_+(\mathbb{E}) = \{\mu : \ \mu \text{ is a non-negative Radon measure on } \mathbb{E}\}.$$

We have the mathematical luxury of working with somewhat 'nice' spaces: $\mathbb{E}$ will be a finite-dimensional compact Hausdorff space, which is therefore Polish, so we have a rich arsenal of topological weapons. To work with sequences of measures and their limits, which is what we desire to elaborate with regard to Definition 9, we introduce a suitable space of test functions. We integrate a test function that is bounded and continuous on $\mathbb{E}$, and if the resulting sequence of numbers converges, then we have a form of weak convergence. Since our measures are also Radon, this suggests using functions that vanish on complements of compact sets. So define

$$C_K^+(\mathbb{E}) = \{f : \mathbb{E} \to \mathbb{R}^+ : \ \text{f is continuous with compact support}\}. \tag{3.4}$$

If we take a look at Proposition 3, we see that regularly varying functions have a continuous version, and after a compactification of $[0, \infty)$ to $[0, \infty]$, one might see this space as a useful start to do analysis for extreme value problems; as shown in the previous chapters, these two concepts are closely related. We often consider integrals of functions rather than measures of sets, because proofs are a bit simpler with this formulation and it is easier to capitalize on continuity arguments. One can always formulate parallel definitions and concepts with sets using a variant of Urysohn's lemma.

On a locally compact Hausdorff space, Radon measures correspond to positive linear functionals on the space of continuous functions with compact support, meaning

$$M_+(\mathbb{E}) \subset (C_K^+(\mathbb{E}))^*,$$

where $\mathbb{A}^*$ denotes the dual space of a space $\mathbb{A}$. This property is the main motivation for the definition and usage of Radon measures here.

We wonder how dense the space $M_+(\mathbb{E})$ actually is. As elaborated in [2], by application of Riesz' representation theorem we can deduce that $M_+(\mathbb{E})$ is isometrically isomorphic to $(C_K^+(\mathbb{E}))^*$, implying that it is always sufficient in the analysis of weak kinds of convergence of measure sequences in $M_+(\mathbb{E})$ to contemplate $C_K^+(\mathbb{E})$, which we can now consider. A Radon measure sequence $(\mu_n)_n$ converging on C_K^+ is called vaguely converging.

Definition 11 (Vague convergence)**.** *If* $\mu_n \in M_+(\mathbb{E})$ *for* $n \geqslant 0$*, then* μ_n *converges vaguely to* μ*, written* $\mu_n \xrightarrow{v} \mu$*, if for all* $f \in C_K^+(\mathbb{E})$*, we have*

$$\mu_n(f) = \int_{\mathbb{E}} f(x)\mu_n(dx) \to \mu(f) = \int_{\mathbb{E}} f(x)\mu(dx), \quad n \to \infty.$$

We can go further. Applying the Banach–Alaouglu theorem to $C_K^+(\mathbb{E})$, we conclude that the closed unit ball in $M_+(\mathbb{E})$ is a compact metrizable space, with the weak* topology. By the definition of weak* topology, we therefore see that a sequence μ_n in $M_+(\mathbb{E})$ converges weak* to μ iff the convergence is vague. The resulting topology is called the vague topology. The topology is metrizable as a complete, separable metric space.

Following this we can define a small Portemanteau theorem for vaguely converging functions which gives an essential idea of how vague convergence has to be understood.

Proposition 9 (Portemanteau theorem)**.** *Let $(\mu_n)_n, \mu \in M_+(\mathbb{E})$. Then the following are equivalent:*

1. $\mu_n \xrightarrow{v} \mu$;
2. $\mu_n(B) \to \mu(B)$ *weakly as* $n \to \infty$ *for all relatively compact* B *satisfying* $\mu(\partial B) = 0$.

Proof. The proof is technical and will be omitted here. We refer to [27]. □

A more detailed look into the topological aspects of vague convergence can be found in [48], [19] and [43].

3.1.3 Characterization of max-stable distributions

We now present a discussion which culminates in a characterization of max-id distributions, which was first introduced by [1] and is further expanded in [44]. We will then return to measures and distributions on $\mathbb{R}^d$.

Theorem 8 (Representation of max-stable distributions)**.** *Let F be a distribution function on $\mathbb{R}^d$. The following are equivalent:*

1. *F is max-id;*
2. *for some $\mathbf{l} \in [\mathbf{0}, \infty)$ there exists a Radon measure μ on $[\mathbf{l}, \infty]\backslash\{\mathbf{l}\}$ such that*

$$F(x_1, \ldots, x_d) = F(\mathbf{x}) = \begin{cases} \exp\left(-\mu([\mathbf{0}, \mathbf{x}]^c)\right), & \mathbf{x} \geqslant \mathbf{l}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

Proof. See [44], Proposition 5.8. □

We recognize a potential problem. The second point implies the existence of a Radon measure on $[\mathbf{l}, \infty]\backslash\{\mathbf{l}\}$, meaning that if we want the measure to be finite in the tail regions, we need to work with the compactified versions of the sets $[\mathbf{l}, \infty)\backslash\{\mathbf{l}\}$, so those sets need to be regarded as bounded

in an appropriate topology so that the sequence of distributions defined in Definition 9 can converge nontrivially. This is achieved by focusing on tail probabilities or exceedance probabilities, which naturally consider probabilities of sets in a neighborhood of infinity.
Clearly, if one removes a set A from a compact superset, the resulting complement of A with respect to its superset doesn't need to be compact, which would lead to trouble. To adapt the compact sets with respect to the induced sub topology, we refer to the following Proposition.

Proposition 10. *Let $(\mathbb{E}, \tau)$ be a locally compact Hausdorff space. Consider a subset $\mathbb{D} \subset \mathbb{E}$. Let $\mathcal{K}(\mathbb{E})$ denote the compact subsets of $\mathbb{E}$. Then*

$$\mathcal{K}(\mathbb{E}\backslash\mathbb{D}) = \{K \in \mathcal{K}(\mathbb{E}) : K \cap \mathbb{D} = \varnothing\}$$

are the compact subsets of $\mathbb{E}\backslash\mathbb{D}$.

Proof. We omit the proof here, since it just uses standard topological arguments to show that the compact sets in the new topology also have finite coverings. It can be found in [43]. □

Proposition 10 is useful if we have to remove subsets from a topological space in which a singularity exists. If we reduce this removed set to a single point $\{\mathbf{x}\}$, then we speak of a one-point-uncompactification, see [43]. This however clarifies the idea of Theorem 8: if F is max-id, then there exists a compact neighbourhood around ∞ due to some unique sub topology given by Proposition 10, on which we can find a Radon measure representing F. We now proceed to characterize multivariate extreme value distributions. We can show that the class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with nondegenerate marginals introduced in the beginning of this chapter. With Theorem 8, we can therefore deduce that if we consider standardization to unit Fréchet marginal distributions (in the definition of the vaguely converging sequence μ_n given by Proposition 5.10 in [44]), there exists a finite measure μ on $[\mathbf{0}, \infty]\backslash\{\mathbf{0}\}$, putting no mass on the lines through ∞ and $\inf_{(x_1,\dots,x_d)\in\mathbb{R}^d_+\backslash\{0\}} \max(x_1, \dots, x_d) > 0$, which is a consequence elaborated in the proof of Proposition 5.8 in [44]. This measure is called the exponent measure.

Definition 12 (Exponent measure)**.** *The measure μ defined in Theorem 8 is called the exponent measure of G_0, since*

$$G_0(x_1, \dots, x_d) = \exp\left(-\mu(A_{x_1,\dots,x_d})\right)$$

with

$$A_{x_1,\dots,x_d} = \left\{(y_1, \dots, y_d) \in \mathbb{R}^d_+\backslash\{\mathbf{0}\} : \exists i \in \{1, \dots, d\} : y_i > x_i\right\}.$$

A characterizing property of the exponent measure is the following homogeneity property. Consider F with standardized marginals. From equation (3.3) we can translate this limit with Theorem 8 to

$$\mu([\mathbf{0}, \mathbf{x}]^c) = t\mu([\mathbf{0}, t\mathbf{x}]^c), \quad t > 0, \ \mathbf{x} \geqslant \mathbf{0}.$$

Now consider

$$\begin{aligned}
[\mathbf{0}, t\mathbf{x}]^c &= \bigcup_{i=1}^{d} \left\{ \mathbf{y} \in [\mathbf{0}, \infty] \backslash \{\mathbf{0}\} : y^{(i)} > t x^{(i)} \right\} \\
&= \bigcup_{i=1}^{d} \left\{ t\mathbf{y} \in [\mathbf{0}, \infty] \backslash \{\mathbf{0}\} : y^{(i)} > x^{(i)} \right\} \\
&= t[\mathbf{0}, \mathbf{x}]^c,
\end{aligned}$$

where we write for a Borel set $B \subset [\mathbf{0}, \infty] \backslash \{\mathbf{0}\}$ that

$$tB = \{tb : b \in B\}.$$

This implies that

$$t^{-1}\mu([\mathbf{0}, \mathbf{x}]^c) = \mu(t[\mathbf{0}, \mathbf{x}]^c). \tag{3.6}$$

For a fixed $t > 0$, this equation can readily be extended to hold for all rectangles $C \subset [\mathbf{0}, \infty] \backslash \{\mathbf{0}\}$, so the equation holds on a generating π-system and is therefore true for all Borel subsets of $[\mathbf{0}, \infty] \backslash \{\mathbf{0}\}$.

This is an essential result. Equation (3.6) implies a regular variation property with index $\alpha = -1$, meaning that characterizing and describing multivariate extreme value distribution boils down to characterizing regular varying measures on cones, which we will further discuss in the next section. More and deeper results can be found in [44] and [11].

3.2 Multivariate regular variation

3.2.1 Regularly varying measures and a spectral representation

We concluded at the end of Section 3.1.1 that characterizing the limit distributions of multivariate extreme value distributions comes down to a problem of characterizing multivariate regularly varying measures on cones. To formally reintroduce the term, a cone (sometimes called a linear cone) is a subset C of a vector space $\mathbb{E}$, if for each $\mathbf{x}$ in C and positive scalars λ, the product $\lambda\mathbf{x}$ is in C.

The theory developed in this thesis could be applied to any cone in a locally finite Hausdorff space on which we can formulate vague convergence but with focus on the application, we will concentrate on the positive compactified orthant in $\mathbb{R}^d$ with excluded origin, namely

$$\mathbb{E} := [\mathbf{0}, \infty] \backslash \{\mathbf{0}\}. \tag{3.7}$$

Another cone that will become important later is the positive inner orthant, defined as

$$\mathbb{E}_1 := \mathbb{E} \backslash \bigcup_{i=1}^{d} \{te_i, t > 0\} = \left\{ s \in \mathbb{E} : \text{ for some } 1 \leqslant i < j \leqslant d : s^{(i)} \wedge s^{(j)} > 0 \right\}, \tag{3.8}$$

where e_i is the i-th base vector in $\mathbb{R}^d$. For more general concepts, we refer to more recent developments in this area, given in [24], where multivariate regular variation is reformulated without using vague convergence, but for the framework discussed it is sufficient to limit ourselves to cones and subcones similar to $\mathbb{E}$.

Regularly varying measures can be characterized in the following way.

Definition 13 (Regularly varying measures)**.** *A sequence $(\mu_n)_n$ in $M_+(\mathbb{E})$ is regularly varying if there exists an increasing sequence $(c_n)_n$ of positive numbers which is regularly varying and a nonzero $\mu \in M_+(\mathbb{E})$ such that*

$$c_n \mu_n \xrightarrow{v} \mu, \quad n \to \infty.$$

We call a single measure $\nu \in M_+(\mathbb{E})$ regularly varying if the sequence $(\nu(n\cdot))_n$ is regularly varying in $M_+(\mathbb{E})$.

Regular varying measures have a lot of equivalent formulations, some of which are more practical in statistical inference, some of which are more common in theoretical approaches. For completeness, we will list them here.

Proposition 11. *Let $\alpha \geqslant 0$. The following statements are equivalent:*

1. *there exist a nonzero $\mu \in M_+(\mathbb{E})$ and a regularly varying sequence $(c_n)_n \in RV_\alpha$ of positive numbers such that*

$$c_n \nu(n\cdot) \xrightarrow{v} \mu(\cdot), \quad n \to \infty;$$

2. *there exist a nonzero $\mu \in M_+(\mathbb{E})$ and a regularly varying function $c \in RV_\alpha$ such that*

$$c(t)\nu(t\cdot) \xrightarrow{v} \mu(\cdot), \quad t \to \infty;$$

3. *there exist a nonzero* $\mu \in M_+(\mathbb{E})$ *and a set* $E \in \mathcal{B}(\mathbb{E})$ *bounded away from* C *such that*
$$\frac{\nu(t\cdot)}{\mu(tE)} \xrightarrow{v} \mu(\cdot), \quad t \to \infty;$$

4. *there exist a nonzero* $\mu \in M_+(\mathbb{E})$ *and an increasing sequence* $(b_n)_n \in RV_{1/\alpha}$ *of positive numbers such that*
$$n\nu(b_n\cdot) \xrightarrow{v} \mu(\cdot), \quad n \to \infty;$$

5. *there exist a nonzero* $\mu \in M_+(\mathbb{E})$ *and an increasing positive function* $b \in RV_{1/\alpha}$ *such that*
$$t\nu(b(t)\cdot) \xrightarrow{v} \mu(\cdot), \quad t \to \infty.$$

They furthermore all imply that the limit measure fulfils the homogeneity property
$$\mu(\lambda A) = \lambda^{-\alpha}\mu(A) \tag{3.9}$$
for $\lambda > 0$ *and an arbitrary set* $A \in \mathcal{B}(\mathbb{E})$.

Proof. The proof will be omitted here. A version can be found in [43]. □

Proposition 11 especially applies for probability measures, since they are a closed subset in $M_+(\mathbb{E})$. We can therefore give a formulation for a regularly varying random variable by defining its cumulative distribution function via a regular varying measure.

It is common to say that a random variable (or a measure) is regularly varying with index α if it fulfils equation (3.9), meaning that the affiliated distribution function is regularly varying with index $-\alpha$. We will align with this convention whenever we speak about regularly varying measures or random variables. Since we are, as assumed, in a Polish space (we will assume $\mathbb{E}$ as defined in (3.7) from now on, if not specified otherwise), we can define a coordinate system as we did in the proof of Theorem 8. This allows us to formulate a regular variation notation for random vectors.

Definition 14 (Multivariate regular varying random vectors)**.** *A random vector* $\boldsymbol{X} = (X_1, \dots X_d)$ *is multivariate regularly varying on a space* $\mathbb{D}$*, written* $\boldsymbol{X} \in MRV_\alpha(\mathbb{D})$*, if there exists a function* $b \in RV_{1/\alpha}$*, so that*
$$tP\left(\frac{\boldsymbol{X}}{b(t)} \in A\right) \xrightarrow{v} \mu(A), \quad t \to \infty,$$
for a measure μ *on* $\mathbb{D}$.

We get a lot of equivalent formulations for this from Proposition 11 and will them use them interchangeably.

It is important to know b, or give at least a somewhat explicit expression for it. A small assumption is that b is the same for all marginal random variables of $\mathbf{X}$, in which case we can without loss of generality consider an arbitrary marginal X_i and $A = [\mathbf{0}, \mathbf{x}]^c$ and see which function b must resemble. We calculate

$$\begin{aligned} tP\left(\frac{X_i}{b(t)} > x_i\right) &= tl(b(t)x_i)(b(t)x_i)^{-\alpha} \\ &= tl(g(t)t^{1/\alpha}x_i)(g(t)t^{1/\alpha}x_i)^{-\alpha} \\ &= tl(g(t)t^{1/\alpha}x_i)g(t)^{-\alpha}t^{-1}x_i^{-\alpha} \\ &= l(g(t)t^{1/\alpha}x_i)g(t)^{-\alpha}x_i^{-\alpha}, \end{aligned}$$

where $g \in RV_0$. With Lemma 1, we can show that this converges to $x_i^{-\alpha}$ if and only if

$$b(t) = l^{\#}(t^{1/\alpha})t^{1/\alpha}. \tag{3.10}$$

We will call this function the scale function of $\mathbf{X}$. The equivalent versions of scaling functions given in Proposition 11 can be calculated in a similar way.

One way equation (3.6) can be understood is as a homogeneity condition, which motivates that one can define a unit set and represent values in the tail regions by stretching it out. It turns out such a representation exists and is unique; we will construct it now. It is important to ensure mesurability and to preserve continuity under those transformations, otherwise we would not be able to build consistent estimators in statistical inference. We therefore consider the following mapping result which is sometimes called the continuous mapping theorem.

Proposition 12. *Suppose that $\mathbb{E}_1, \mathbb{E}_2$ are two locally compact Hausdorff spaces and $h : \mathbb{E}_1 \to \mathbb{E}_2$ is continuous. Then*

$$\mu_n \xrightarrow{v} \mu \text{ implies } \mu_n \circ h^{-1} \xrightarrow{v} \mu \circ h^{-1}.$$

Proof. Suppose $\mu_n \xrightarrow{v} \mu$. Let $f \in C_K^+(\mathbb{E}_2)$ be arbitrary. We must show that

$$\mu_n \circ h^{-1}(f) \xrightarrow{v} \mu \circ h^{-1}(f).$$

Representing the sequence as an integral, we get with the integral transformation theorem

$$\begin{aligned}\mu_n \circ h^{-1}(f) &= \int_{\mathbb{E}_2} f(y)\mu_n \circ h^{-1}(dy) \\ &= \int_{\mathbb{E}_1} f(h(x))\mu_n(dx).\end{aligned}$$

It remains to show that $f \circ h \in C_K^+(\mathbb{E}_1)$. Since both f and h are continuous, so is $f \circ h$. Since $f \in C_K^+(\mathbb{E}_2)$, there exists a set $K_2 \in \mathcal{K}(\mathbb{E}_2)$ in the set of compact sets in $\mathbb{E}_2$, such that $f(y) = 0,\ \forall y \neq K_2$. Therefore we have $(f \circ h)(x) = 0\ \ \forall h(x) \neq K_2$. Since both spaces are locally compact, we know that

$$h^{-1}(K_2) \in \mathcal{K}(\mathbb{E}_1),$$

implying that $f \circ h$ is not null on some compact set and null everywhere else, thus $f \circ h \in C_K^+(\mathbb{E}_1)$, implying the claim. □

This allows us to define continuous maps and still preserve vague convergence. Motivated by the homogeneity property, we can transform the random vector from a Cartesian to a polar coordinate system and in that system we define the following map,

$$h : \mathbf{x} \mapsto \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right), \tag{3.11}$$

with an arbitrary norm $\|\cdot\|$ on $[0, \mathbf{x}]^c$ on $\bar{\mathbb{R}}^d$.

This map creates problems on the full positive compactified cone, since h is obviously not defined at $\mathbf{x} = \mathbf{0}$ due to a singularity at this point. However, by taking out the origin and applying the one-point-uncompactification on $[\mathbf{0}, \mathbf{x}]^c$ as described in Proposition 10, we have a sub-topology with a corresponding set of compact subsets of $[\mathbf{0}, \mathbf{x}]^c \backslash \{\mathbf{0}\}$, under which the polar coordinate transformation (3.11) is continuous.

We define the unit ball on $\mathbb{E}$ as

$$\mathcal{N} := \{\mathbf{x} \in \mathbb{E} :\ \|\mathbf{x}\| = 1\}, \tag{3.12}$$

recognize by inspecting the map (3.11) that

$$h : \mathbb{E} \to (0, \infty] \times \mathcal{N},$$

and deduce with Proposition 12 that we can decompose a regularly varying random vector into a spectral and a radial part,

$$h(\mathbf{X}) = (R, \Theta),$$

with inverse map $\mathbf{X} = h^{-1}(R, \Theta) = R\Theta$, which preserves vague convergence by the continuous mapping theorem.

Consider now the measure $\mu \circ h^{-1}$. We can deduce from equation (3.6) that decomposition is valid for arbitrary rectangles (and finite intersections of them) bounded away from 0. By Dynkin's $\pi - \lambda$-theorem, we know that this is therefore valid for generally measurable sets A, so

$$t^{-1}(\mu \circ h^{-1})(A) = (\mu \circ h^{-1})(tA).$$

This implies, for any $t > 0$ and any measurable set $S \subset \mathcal{N}$, that after transformation with h,

$$\begin{aligned}
\mu\left\{\mathbf{x} \in (\mathbf{0}, \infty] : \|\mathbf{x}\| > t, \tfrac{\mathbf{x}}{\|\mathbf{x}\|} \in S\right\} &= \mu\left\{\mathbf{x} \in (\mathbf{0}, \infty] : \|t^{-1}\mathbf{x}\| > 1, \frac{t^{-1}\mathbf{x}}{\|t^{-1}\mathbf{x}\|} \in S\right\} \\
&= \mu\left\{t\mathbf{y} \in (\mathbf{0}, \infty] : \|\mathbf{y}\| > 1, \frac{\mathbf{y}}{\|\mathbf{y}\|} \in S\right\} \\
&= t^{-1}\mu\left\{\mathbf{y} \in (\mathbf{0}, \infty] : \|\mathbf{y}\| > 1, \frac{\mathbf{y}}{\|\mathbf{y}\|} \in S\right\} \\
&= t^{-1}cH(S),
\end{aligned}$$

for a measure H on $\mathcal{N}$ and a constant $c > 0$. Thus, on $(0, \infty] \times \mathcal{N}$,

$$\mu \circ h^{-1} = c\nu_1 \times H. \tag{3.13}$$

The measure ν_1 is a unit Pareto measure and the measure H is called the spectral measure of μ. Since $\mathcal{N}$ is a compact set and μ is Radon, H must be Radon as well and is therefore finite, so we can assume that H is a probability measure, chosen so that $\Theta \sim H$. To summarize we can show that

$$\begin{aligned}
tP\left(\frac{\mathbf{X}}{b(t)} \in A\right) = tP\left(\left(\frac{R}{b(t)}, \Theta\right) \in h(A)\right) &\xrightarrow{v} (\nu \circ h^{-1})(A) \\
&= (c\nu_1 \times H)((0, x] \times S)
\end{aligned}$$

on $M_+((0, \infty] \times \mathcal{N})$. Without loss of generality, we can choose a spectral decomposition induced by the L_1-norm. Then the spectral decomposition allows us to represent the exponent measure in a very convenient way.

Theorem 9 (Integral representation theorem)**.** *Let h be the polar coordinate transformation described in equation* (3.11)*. Assume that*

$$\mu \circ h^{-1} = c\nu_1 \times H,$$

where $\nu_1((x,\infty]) = x^{-1},\ x > 0$ and H is a probability measure on $\mathcal{N}$. Then

$$\mu([\mathbf{0},\mathbf{x}]^c) = d\int_{\mathcal{N}} \bigvee_{i=1}^{d} \frac{\omega^{(i)}}{x^{(i)}} H(d\omega)$$

with the condition that

$$\int_{\mathcal{N}} \omega_i H(d\omega) = \frac{1}{d}.$$

Proof. The proof can be found in [43]. We have

$$\mu([\mathbf{0},\mathbf{x}]^c) = \mu \circ h^{-1}(h([\mathbf{0},\mathbf{x}]^c)).$$

Now

$$\begin{aligned}
h([\mathbf{0},\mathbf{x}]^c) &= h\left(\left\{\mathbf{y} \in (\mathbf{0},\infty] : y^{(i)} > x^{(i)} \text{ for some } i\right\}\right) \\
&= \left\{(r,d) \in (0,\infty] \times \mathcal{N} : (ra)^{(i)} > x^{(i)} \text{ for some } i\right\} \\
&= \left\{(r,d) \in (0,\infty] \times \mathcal{N} : r > \frac{x^{(i)}}{\omega^{(i)}} \text{ for some } i\right\} \\
&= \left\{(r,d) \in (0,\infty] \times \mathcal{N} : r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{\omega^{(i)}}\right\}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mu([\mathbf{0},\mathbf{x}]^c) &= \mu \circ h^{-1}\left(\left\{(r,d) \in (0,\infty] \times \mathcal{N} : r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{\omega^{(i)}}\right\}\right) \\
&= c\nu_1 \times H\left(\left\{(r,d) \in (0,\infty] \times \mathcal{N} : r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{\omega^{(i)}}\right\}\right) \\
&= c\int_{\mathcal{N}} \int_{r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{\omega^{(i)}}} \nu_1(dr) H(d\omega) \\
&= c\int_{\mathcal{N}} \left(\bigwedge_{i=1}^{d} \frac{x^{(i)}}{\omega^{(i)}}\right)^{-1} H(d\omega) \\
&= c\int_{\mathcal{N}} \bigvee_{i=1}^{d} \frac{\omega^{(i)}}{x^{(i)}} H(d\omega).
\end{aligned}$$

We now seek to fulfil the moment condition, yet this is simply equivalent to the assumption that μ was derived from a standardized extreme value distribution G_0 from Theorem 8 and the condition simply ensures that the

marginals are unit-Frechét distributed. □

The representation described in Theorem 9 resembles that the limit distribution in Definition 12 is characterized by the spectral measure and the tail coefficient of the marginal distributions. Of course we don't have to take the measure H on the unit simplex described here; there exist many other equivalent versions, depending on the norm we choose for the polar decomposition and on the resulting unit ball we use, but we will focus on measures induced by the L_1-norm, since they are the most common. We can also assume a measure H on a unit simplex with the moment condition given in the Theorem 9 and then show that this measure defines a unique extreme value distribution. Further discussion about this, different representations and their properties can be found in [11] and [43].

3.2.2 Tail dependence structures

The essential result of the last section is that we can split up limiting multivariate extreme value behaviour into a radial component ν_1 and a spectral component H, which describes the tail dependence structure. To characterize dependence, we need a further understanding of these measures. We will thus derive the expression

$$\mu([\mathbf{0},\mathbf{x}]^c) = \lim_{t\to\infty} tP\left(\frac{\mathbf{X}}{b(t)} \in [\mathbf{0},\mathbf{x}]^c\right)$$

under various assumptions. While we know from Theorem 9 that the class of admissible spectral measures H is very broad, we will mainly focus on two extremal cases of coordinate dependence, namely perfect dependence and perfect independence.

Before we start, we have to make some assumptions. We assume that all marginal distribution are regularly varying with the same tail index $\alpha = 1$ and that they have the same scale function $b(t)$. The first assumption is trivial, since, as long as all marginals have the same marginal tail index $\alpha > 0$, we can always transform μ to have unit tail index by using

$$\mu^*([\mathbf{0},\mathbf{x}]^c) = \mu([\mathbf{0},\mathbf{x}^{1/\alpha}]^c).$$

Thus, without loss of generality, we will use the transformed μ^* in this chapter if not explicitly said otherwise. The second assumption is easily generalizable through the concept of tail equivalence, giving us more freedom in modelling. This concept was introduced in [41] to which we refer for a more detailed discussion of the topic. Two random variables X_i, X_j with probability distributions F_i, F_j on $\mathbb{R}_+$ are tail equivalent if they have the

same right endpoint x_0 and, for some $c \in (0, \infty)$,

$$\lim_{x \to x_0} \frac{P(X_i > x)}{P(X_j > x)} = c. \tag{3.14}$$

Then, upper extremes of i.i.d. samples from F_i behave asymptotically in the same way as upper extremes of i.i.d. samples from F_j. In the heavy-tailed case, this definition is equivalent to supposing that there exists a function $b(t) \to \infty$ such that for some $\alpha > 0$, as $t \to \infty$

$$\lim_{t \to \infty} tP\left(\frac{X_i}{b(t)} > x\right) = cx^{-\alpha},$$
$$\lim_{t \to \infty} tP\left(\frac{X_j}{b(t)} > x\right) = x^{-\alpha}.$$

Now let's suppose that $\mathbf{Y}$ and $\mathbf{Z}$ are $[\mathbf{0}, \infty)$-valued random vectors with distributions F and G. In the multivariate regular variation context, we say that F and G (or $\mathbf{Y}$ and $\mathbf{Z}$) are tail-equivalent on the cone $\mathcal{C} \subset \mathbb{E}$ if there exists a scaling function $b(t) \to \infty$ such that

$$tP\left(\frac{\mathbf{Y}}{b(t)} \in A\right) \xrightarrow{v} \nu(A),$$
$$tP\left(\frac{\mathbf{Z}}{b(t)} \in A\right) \xrightarrow{v} c\nu(A)$$

in $M_+(\mathcal{C})$ for an arbitrary relatively compact Borel set A, for some constant $c > 0$ and non-zero Radon measure ν on $\mathcal{C}$. We then write

$$\mathbf{Y} \overset{\mathrm{te}(\mathcal{C})}{\sim} \mathbf{Z}.$$

We can therefore relax the assumption that all marginals must have the same scale function to the assumption that the random variable is tail equivalent to such a random variable.

Example 1. We start with a perfectly dependent case and assume the perfectly dependent d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_1)$ with regular varying random variable X_1. Then Theorem 9 and Corollary 2 give

$$\begin{aligned}
\mu([\mathbf{0}, \mathbf{x}]^c) &= \lim_{t \to \infty} tP\left(\frac{\mathbf{X}}{b(t)} \in [\mathbf{0}, \mathbf{x}]^c\right) \\
&= \lim_{t \to \infty} tP\left(\bigcup_{i=1}^{d} \left\{\frac{X_1}{b(t)} > x_i\right\}\right) \\
&= \lim_{t \to \infty} tP\left(\frac{X_1}{b(t)} > \min_{1 \leqslant i \leqslant d} x_i\right) \\
&= \max_{1 \leqslant i \leqslant d} x_i^{-1}.
\end{aligned}$$

The spectral measure concentrates here on a single point, i.e.,

$$H(\omega) = \delta_{\frac{1}{d}\sum_{i=1}^{d} e_i}(\omega), \tag{3.15}$$

where δ is the Dirac measure on the set $\left\{\frac{1}{d}\sum_{i=1}^{d} e_i\right\}$, which fulfils the conditions in Theorem 9.

Example 2. We turn to perfect independence now. Assume $\mathbf{X} = (X_1, \dots X_d)$ to be a perfectly independent random vector and with Proposition 11 that $b(t) = t$. Then Proposition 9 and Corollary 2 imply that

$$\begin{aligned}
\mu([\mathbf{0}, \mathbf{x}]^c) &= \lim_{t\to\infty} tP\left(\frac{\mathbf{X}}{b(t)} \in [\mathbf{0}, \mathbf{x}]^c\right) \\
&= \lim_{t\to\infty} tP\left(\bigcup_{i=1}^{d}\left\{\frac{X_i}{b(t)} > x_i\right\}\right) \\
&= \lim_{t\to\infty}\left(\sum_{i=1}^{d} tP\left(\frac{X_i}{b(t)} > x_i\right) - \sum_{1\leqslant i<j\leqslant d} tP\left(\frac{X_i}{b(t)} > x_i, \frac{X_j}{b(t)} > x_j\right) + \dots\right) \\
&= \lim_{t\to\infty}\left(\sum_{i=1}^{d} t\frac{1}{b(t)x_i} - \sum_{1\leqslant i<j\leqslant d} tP\left(\frac{X_i}{b(t)} > x_i\right) P\left(\frac{X_j}{b(t)} > x_j\right) + \dots\right) \\
&= \sum_{i=1}^{d} x_i^{-1} - \lim_{t\to\infty}\left(\sum_{1\leqslant i<j\leqslant d} x_i^{-1} \underbrace{P\left(\frac{X_j}{b(t)} > x_j\right)}_{\to 0} + \dots\right) \\
&= \sum_{i=1}^{d} x_i^{-1}
\end{aligned}$$

The spectral measure H therefore concentrates on the axes and is

$$H(\omega) = \frac{1}{d}\sum_{i=1}^{d} \delta_{e_i}(\omega) \tag{3.16}$$

which also fulfils the moment conditions in Theorem 9.

Those two examples mark the extreme dependence structures, concentrating all mass at the axes and on the inside of the cone on a single point. The corresponding extreme value distributions are consequently

$$G_0(x_1, \dots, x_n) = \exp\left(-\max_{1\leqslant i\leqslant d} x_i^{-1}\right), \quad x_1, \dots, x_n > 0,$$

for the perfectly dependent case and

$$G_0(x_1, \dots, x_n) = \exp\left(-\sum_{i=1}^{d} x_i^{-1}\right), \quad x_1, \dots, x_n > 0.$$

for the perfectly independent case; in particular this means that the extreme value distribution is a product of its single coordinates.

While independence of components implies the concentration of extreme values on the axes, the converse is false. A well-known counterexample comes from Sibuya, who showed in 1959 in [46] that a perfectly independent limit dependence structure also holds for the bivariate normal distribution, if its components are not perfectly correlated.

The problematic implication of this is obvious. If we want to assess sub-asymptotic tail regions in practical risk applications, it is not sufficient to simply use the extreme value distribution of the model, since, while still converging towards the limit, the convergence rate can be arbitrarily slow. Since the random vector still eventually behaves like an independent model, we call it asymptotically independent. To discuss a suitable framework which deals with this we will briefly summarize the work done on the bivariate case, so we assume dimension $d = 2$ in this section unless otherwise specified, and extend to higher dimensions later. Formally, a bivariate random vector $\mathbf{X} = (X_1, X_2)$ possesses asymptotic independence that if and only if

$$\nu(\mathbb{E}_1) = 0, \tag{3.17}$$

with $\mathbb{E}_1$ defined in equation (3.8), so that ν concentrates on the axes. This means the probability of simultaneous occurrence of large values on both coordinates is estimated to be zero. As a side note, one can discuss asymptotic dependent models in a similar manner.

In theory, the dependence measure of an asymptotic or perfectly independent random vector is singular, but this singularity cannot be identified from finite samples. There have been several ways to build extensions towards solving this problem, most of them for the bivariate case, but they can be extended to a general multivariate setting.The first more general approach to develop a way to identify and treat asymptotic independence stems from the papers of Ledford and Tawn([29] and [30]) who proposed a model based on the following observation for the bivariate case. Assume that $\mathbf{X}$ is a perfectly independent standard Pareto distributed random vector. We derive

$$P(X_1 > t, X_2 > t) = P(X_1 > t)\, P(X_2 > t) = t^{-2}, \quad t \geqslant 1,$$

whereas for the perfect dependent case, we get

$$P(X_1 > t, X_2 > t) = P(X_1 > t, X_1 > t) = P(X_1 > t) = t^{-1}, \quad t \geqslant 1,$$

Ledford and Tawn showed that in the asymptotic independent case,

$$P(X_1 > t, X_2 > t) = P(X_1 \wedge X_2 > t) = l(t)t^{-1/\eta}, \quad t \geqslant 0, \tag{3.18}$$

with $l \in RV_0$ and $\frac{1}{2} < \eta < 1$, where η is called the coefficient of tail dependence, and investigated several special cases, depending on both η and the limit behaviour of l. They noted that $\eta < \frac{1}{2}$ is also achievable and leads to negative dependence between the variables. A year later, this was extended towards what is now known as the bivariate Ledford–Tawn model, presented in [30]. They noted that a max-id survivor function given through

$$P(X_1 > x_1, X_2 > x_2) = P(X_1 > x_1) + P(X_2 > x_2) - P(X_1 > x_1 \text{ or } X_2 > x_2)$$

can be expanded into a flexible joint asymptotic expansion that contains and extends model (3.18) and gave conditions under which the components of this expansion are uniquely determined. They considered

$$P(X_1 > x_1, X_2 > x_2) = l_1(x_1, x_2)x_1^{-c_1}x_2^{-c_2} + l_2(x_1, x_2)x_1^{-d_1}x_2^{-d_2} \tag{3.19}$$

where $c_1 + c_2 = 1/\eta$, $d_i \geqslant 0$, and $0 \not\equiv l_i(x_1, x_2) \in RV_0$ bivariate slowly varying functions. This has been a very successful approach in modelling asymptotically independent models and a great starting point for other extensions, but is not without shortcomings.

Heffernan and Tawn (2004) recognized that those techniques are based on limiting arguments in which all components of the variable become large at the same rate. This is inappropriate when the extreme values of all the variables are unlikely to occur together or when interest is in regions of the support of the joint distribution where only a subset of components is extreme. Their approach was to model extremal sets separately and investigate how the limit distribution behaves conditioned on the event that one of the coordinates exceeds a very high threshold, calling this the conditional extreme value model (CEV model). One can find further details and examples in [23].

A fundamental step into another, very important, model class direction was given through the discovery that assuming second-order regular variation effects on the joint tail distribution induces an implicit way to represent η via an assumed second order condition on the joint tail. Resnick picked up this thought in 2002 and developed his theory of hidden regular variation.

3.3 Hidden Regular Variation

Hidden regular variation is a young concept which has found very much traction lately, since it brings the Ledford–Tawn formulation in equation (3.19) back into a regularly varying context and delivers a more powerful limit theory. Classically formulated for the bivariate case, we will switch back to an arbitrary finite multivariate setting.

We assume now that $\mathbf{X} \in MRV_\alpha(\mathbb{E})$ also fulfills a non-trivial second-order regular variation condition, i.e., $\mathbf{X} \in 2RV_{\alpha,\rho}$ as introduced in Definition 4. Therefore assume that there exist functions $b(t) \in RV_{1/\alpha}$ and $a(t) \to 0$, which is eventually positive, such that $|a| \in RV_\rho$ and

$$\frac{tP\left(\frac{X}{b(t)} \in [\mathbf{0}, \mathbf{x}]^c\right) - \nu([0, \mathbf{x}]^c)}{a(b(t))} \to H(\mathbf{x}), \tag{3.20}$$

locally uniformly in $x \in \mathbb{E}_1$, with H finite and not identically zero. Assuming asymptotic dependence now, we get for any relatively compact Borel set $A \in \mathbb{E}_1$ that the exponent measure puts zero mass on it, since it concentrates onto the axes, as we saw in Example 2. These sets can be identified using Proposition 10. This also implies that

$$\begin{aligned}\frac{tP\left(\frac{\mathbf{X}}{b(t)} \in A\right) - \nu(A)}{a(b(t))} &= \frac{tP\left(\frac{\mathbf{X}}{b(t)} \in A\right)}{a(b(t))} \\ &\to \nu_1(A),\end{aligned}$$

with $\nu_1([\mathbf{0}, \mathbf{x}]^c) = H(\mathbf{x})$. Since $b(t) \to \infty$ we can assume A to be positive. Since relatively compact sets imply vague convergence in $M_+(\mathbb{E}_1)$ due to Proposition 9, we get

$$\frac{t}{a(b(t))} P\left(\frac{\mathbf{X}}{b(t)} \in A\right) \xrightarrow{v} \nu_1(A) \tag{3.21}$$

for a relatively compact Borel set A on $M^+(\mathbb{E}_1)$. We will now use a trick that is often used in proofs and interpret the convergence described in (3.21) as convergence of a scaled random vector with a different rate. Note that $\mathbf{X}$ is regularly varying and the function

$$\frac{t}{a(b(t))} =: U(t) \in RV_{1-\frac{\rho}{\alpha}}$$

is asymptotically equivalent to a strictly increasing function $\tilde{U}$ with an generalized inverse $\tilde{U}^{\leftarrow} \in RV_{\frac{1}{1-\frac{\rho}{\alpha}}}$. With Proposition 11 we can now find a scale

function asymptotically equivalent to $b \circ \tilde{U}^{\leftarrow}(s)$ such that

$$sP\left(\frac{\mathbf{X}}{b \circ \tilde{U}^{\leftarrow}(s)} \in A\right) \xrightarrow{v} \nu_1(A),$$

which is the familiar form of a regularly varying random vector. This means that by considering second-order regular variation, asymptotic independence and the assumption of seeking regular variation on $\mathbb{E}_1$, we can find another limit measure ν_1, such that we observe limit effects of higher order, which would vanish under the standard exponential measure and are therefore "hidden" under the standard regular variation. The order is determined by the regular variation coeficient of $b \circ \tilde{U}^{\leftarrow}$. We call this "hidden regular variation". A distribution possessing hidden regular variation is also regularly varying, which can be shown using Theorem 4. The argument can be found more in detail in [42].

Definition 15. *A random vector* $\mathbf{X}$ *has a distribution possessing hidden regular variation of order* $\alpha_1 > \alpha > 0$*, if there is a subcone* $\mathbb{D} \subset \mathbb{E}$ *and the distribution has regular variation with index* α *on* $\mathbb{E}$ *and regular variation of order* α_1 *on* $\mathbb{D}$*, meaning that*

$$\boldsymbol{X} \in MRV_{\alpha}(\mathbb{E}) \cap MRV_{\alpha_1}(\mathbb{D}).$$

Equivalently, there exists a non-decreasing function $b_1(t) \to \infty$*,* $b_1 \in RV_{1/\alpha_1}$ *for an* $\alpha_1 > \alpha$ *such that* $b(t)/b_1(t) \to \infty$ *and there exists a measure* ν_1 *which is Radon on* $\mathbb{D}$ *and such that*

$$tP\left(\frac{\boldsymbol{X}}{b_1(t)} \in A\right) \xrightarrow{v} \nu_1(A) \tag{3.22}$$

for relatively compact sets A *on* $M_+(\mathbb{D})$*. The measure* ν_1 *satisfies a homogeneity condition analogously to equation* (3.9)*. The measure* ν_1 *is called the hidden measure of* $\mathbf{X}$*.*

We can again assume that b_1 is strictly increasing, continuous and further assume that $b_1(1) = 1$ and hence $b_1^{\leftarrow}(1) = 1$. For now we will focus on the exploration of hidden regular variation for $\mathbb{D} = \mathbb{E}_1$; we will see later that this is not the only possible choice for a subcone.

Since hidden regular variation implies regular variation on the subcone $\mathbb{E}_1$, we can apply the representation theory from Section 3.2.1 and decompose the hidden measure into a spectral representation with a spectral measure H_1 on sets of the form $(0, x] \times \mathcal{N}_1$, where $\mathcal{N}_1 = \mathcal{N} \cap \mathbb{E}_1$. We will call this measure the hidden spectral measure of $\mathbf{X}$. Note that $\mathcal{N}_1$ is not compact and therefore H_1 is not necessarily finite, so ν_1 might not be a finite measure on

$\{\mathbf{x} \in \mathbb{E}_1 : \|\mathbf{x}\| > 1\}$. We can only guarantee that it is Radon. We demonstrate this with two examples.

Example 3. We will start by assuming that $\mathbf{X} = (X_1, \ldots, X_d)$ has i.i.d. Pareto-distributed marginal distributions with tail index $\alpha > 0$ on $\mathbb{E}$, meaning that

$$P(X_i > x) = x^{-\alpha}, \quad x > 0,$$

and we will show that $\mathbf{X}$ has hidden regular variation on $\mathbb{E}_1$. Set $b(t) = t^{1/\alpha}$. As we have calculated in Example 2, we get (after transformation to a unit exponent measure) that

$$tP\left(\frac{\mathbf{X}}{b(t)} \in [\mathbf{0}, \mathbf{x}]^c\right) \xrightarrow{v} \sum_{i=1}^{d} x_i^{-\alpha}.$$

Now choose $\alpha_1 > \alpha$ and take an arbitrary rectangle $[\mathbf{x}, \infty]$ bounded away from the axes. Then

$$\begin{aligned} tP\left(\frac{\mathbf{X}}{b_1(t)} \in [\mathbf{0}, \infty]\right) &= t \sum_{1 \leqslant i < j \leqslant d} P(X_i > b_1(t)x_i, X_j > b_1(t)x_j) \\ &\quad + t \sum_{1 \leqslant i < j < k \leqslant d} P(X_i > b_1(t)x_i, X_j > b_1(t)x_j, X_k > b_1(t)x_k) + \cdots \\ &= t \sum_{1 \leqslant i < j \leqslant d} P(X_i > b_1(t)x_i) P(X_j > b_1(t)x_j) \\ &\quad + t \sum_{1 \leqslant i < j < k \leqslant d} P(X_i > b_1(t)x_i) P(X_j > b_1(t)x_j) P(X_k > b_1(t)x_k) + \cdots \\ &= \sum_{1 \leqslant i < j \leqslant d} \frac{t}{b_1(t)^{2\alpha}} (x_i x_j)^{-\alpha} + \sum_{1 \leqslant i < j < k \leqslant d} \frac{t}{b_1(t)^{3\alpha}} (x_i x_j x_k)^{-\alpha} + \cdots, \end{aligned}$$

which is finite if and only if the leading term converges, therefore $b_1 \in \Theta(t^{\frac{1}{2\alpha}})$ regularly varying. We therefore have hidden regular variation of order $\alpha_1 = 2\alpha$ with $\nu_1([0, \mathbf{x}]^c) = \sum_{1 \leqslant i < j \leqslant d} (x_i x_j)^{-\alpha}$. In this case, ν_1 is infinite on $\{x \in \mathbb{E}_1 : \|x\| > 1\}$, since for any $\epsilon > 0$,

$$\nu_1(\{\mathbf{x} \in \mathbb{E}_1 : \|\mathbf{x}\| > 1\}) \geqslant \nu_1([2, \infty] \times [\epsilon, \infty] \times [1, \infty]^{d-2}) = (2\epsilon)^{-\alpha} \to \infty$$

as $\epsilon \to 0$.

From this example, we can draw a connection to the Ledford–Tawn model in two dimensions. Assume $d = 2$ and a bivariate regular varying random vector $\mathbf{X}$ possessing asymptotic independence and having tail index $\alpha > 0$. Resnick characterizes hidden regular variation via max- and min-linear combinations in his paper (see [42], Theorem 1) and shows that $\mathbf{X}$ has hidden regular variation with tail index $\alpha_1 > \alpha$, if and only if $X_1 \vee X_2 \in RV_\alpha$ and $X_1 \wedge X_2 \in RV_{\alpha_1}$. We can reformulate the Ledford–Tawn model in

equation (3.18) for a general $\alpha > 0$ to

$$P\left(X_1 \wedge X_2 > t\right) = l(t)P\left(X_1 > t\right)^{-1/\eta}, \quad t > 0, \tag{3.23}$$

with $l \in RV_0$. Suppose, as usual, that $b(t)$ satisfies

$$\lim_{t\to\infty} tP\left(\frac{X_1}{b(t)} > 1\right) = 1.$$

Then (3.23) becomes

$$P\left(X_1 \wedge X_2 > b(t)\right) \sim l(b(t))t^{1/\eta} = l_*(t)t^{-1/\eta}, \quad t > 0.$$

Since we assume hidden regular variation, we also get

$$P\left(X_1 \wedge X_2 > b(t)\right) \sim l(b(t))b(t)^{-\alpha_1}, \quad t > 0.$$

Comparing those two expressions gives with Karamata's representation theorem that

$$P\left(X_1 \wedge X_2 > b(t)\right) \sim l_{**}(t)t^{-\alpha_1/\alpha}, \quad t > 0, \tag{3.24}$$

for a slowly varying function l_{**}, from which we conclude by comparing with the initial Ledford–Tawn model that

$$\eta = \frac{\alpha}{\alpha_1}. \tag{3.25}$$

Thus hidden regular variation successfully generalizes this concept.

Example 4. For simplicity, assume $d = 2$ and consider three independent random quantities $B, \mathbf{Y}, \mathbf{U}$. Suppose B is a Bernoulli random variable with

$$P(B = 0) = P(B = 1) = \frac{1}{2}$$

and $\mathbf{Y}$ has, as in the last example, i.i.d. Pareto-distributed marginal distributions with tail index $\alpha > 0$ on $\mathbb{E}$. Set $b(t) = t$. Suppose $\mathbf{U}$ has a multivariate regularly varying distribution on $\mathbb{E}$ (not just on $\mathbb{E}_1$) and that there exist $\alpha_1 > \alpha$, $b_1 \in RV_{1/\alpha_1}$ and a nonzero measure ν_1 that is Radon on $\mathbb{E}$, such that

$$tP\left(\frac{\mathbf{X}}{b_1(t)} \in A\right) \xrightarrow{v} \nu_1(A), \quad t \to \infty.$$

Define

$$\mathbf{X} = B\mathbf{Y} + (1 - B)\mathbf{U}.$$

We easily check with Proposition 3 that $\mathbf{X}$ is regularly varying with tail index α on $\mathbb{E}$. But $\mathbf{X}$ also possesses hidden regular variation, as we can

show. So on $\mathbb{E}_1$, we have

$$\begin{aligned} tP\left(\frac{\mathbf{X}}{b_1(t)} \in [\mathbf{x}, \infty]\right) &= tP\left(\frac{B\mathbf{Y} + (1-B)\mathbf{U}}{b_1(t)} \in [\mathbf{x}, \infty]\right) \\ &= \frac{t}{2}P\left(\frac{\mathbf{Y}}{b_1(t)} \in [\mathbf{x}, \infty]\right) + \frac{t}{2}P\left(\frac{\mathbf{U}}{b_1(t)} \in [\mathbf{x}, \infty]\right) \\ &= I + II, \end{aligned}$$

So, as assumed, we get

$$II \to \frac{1}{2}\nu_1([\mathbf{x}, \infty]).$$

For term I, we know from Example 3 that

$$I = \frac{1}{2}\frac{t}{b_1(t)^{2\alpha}}(x_1 x_2)^{-\alpha},$$

which converges to zero if and only if $\alpha < \alpha_1 < 2\alpha$. So ,together with this assumption, we have that $\mathbf{X}$ is regularly varying on $\mathbb{E}$ with tail index α and possesses hidden regular variation on $\mathbb{E}_1$ with hidden measure defined by $\nu_1 := \nu|_{\mathbb{E}_1}$. Since ν is a Radon measure, ν_1 is finite. This is known as a mixture model.

The second example shows how fundamentally important hidden regular variation can be. We get, together with the assumption that $\alpha < \alpha_1 < 2\alpha$ that this model always possesses effects that can make a huge difference for assessing risks on all subasymptotic levels.

From an applied perspective, the second example yields a great way to generate models that have desired asymptotic characteristics, which is why we will take a closer look into models constructed in such a way. It turns out that they have a couple of very useful properties.

Theorem 10. *Suppose $\boldsymbol{X}$ is a regularly varying random vector on $\mathbb{E}$ with hidden regular variation on $\mathbb{E}_1$, finite hidden spectral measure, and scaling functions $b(t)$, $b_1(t)$ with $\frac{b(t)}{b_1(t)} \to \infty$. Then $\boldsymbol{X}$ is tail equivalent on both the cones $\mathbb{E}$ and $\mathbb{E}_1$ to a random vector which is a mixture of*

1. *a completely asymptotically independent random vector on $\mathbb{E}$ (with no hidden regular variation) whose marginal distributions have scaling function $b(t)$, and*
2. *an extremally dependent random vector on $\mathbb{E}$ with scaling function $b_1(t)$.*

Conversely, if $\boldsymbol{X}$ is tail equivalent to a mixture as above with $\frac{b(t)}{b_1(t)} \to \infty$, then $\boldsymbol{X}$ is multivariate regularly varying on $\mathbb{E}$ and has hidden regular variation on $\mathbb{E}_1$ with finite hidden spectral measure and with scaling functions b and b_1.

Proof. The proof is adapted from [36]. Let $\mathbf{X}$ be a d-dimensional random vector distributed as F, which is multivariate regularly varying on $\mathbb{E}$ with hidden regular variation on $\mathbb{E}_1$. Suppose $\mathbf{X}$ has finite hidden spectral measure which, without loss of generality, we take to be a probability measure. Then $\mathbf{X}$ has a polar decomposition on both $\mathbb{E}$ and $\mathbb{E}_1$, represented as in (3.13) and can be split into a radial and a spectral random variable R and Θ.

Extend the hidden spectral measure H_1 to $\mathcal{N}$ by defining it to be 0 on $\mathcal{N}\backslash\mathcal{N}_1$. Recall that the one-dimensional marginal distributions of $\mathbf{X}$ are identical. Take i.i.d. random variables $X_1, \ldots, X_d$ having the same distribution as Y_1 and independent of the independent pair R_1 and Θ_1 (all defined on the same probability space), where Θ_1 is distributed as H_1 and R_1 is distributed as F_1 given by

$$P(R_1 > x) = \begin{cases} 1, & x \leqslant 1, \\ \frac{1}{b_1^{\leftarrow}(x)}, & x \geqslant 1. \end{cases}$$

We further take an integer-valued random variable I independent all of the rest (and defined on the same probability space as before), which takes value 0 with probability $\frac{1}{2}$ and values $1, \ldots, d$ with probability $\frac{1}{2d}$ each. Define V to be the random vector obtained by transforming the pair (R_1, Θ_1) to Cartesian coordinates, that is,

$$\mathbf{V} = R_1\Theta_1.$$

Finally define

$$\mathbf{Z} = 2^{1/\alpha_1}\mathbb{I}_{I=0}\mathbf{V} + 2^{1/\alpha}\sum_{i=1}^{d}\mathbb{I}_{I=i}X_i e_i \tag{3.26}$$

where e_i is the i-th unit vector in $\mathbb{R}^d$.

We claim that the distribution of $\mathbf{Z}$ is tail equivalent to F on both $\mathbb{E}$ and $\mathbb{E}_1$. First we check that $\mathbf{V}$ is multivariate regularly varying on $\mathbb{E}$ with scaling function $b_1(t)$, and hence also on $\mathbb{E}_1$. We check this using polar coordinates. For any Borel set $S \subseteq \mathcal{N}$ and $x > 0$,

$$tP\left(\frac{R_1}{b_1(t)} > x, \Theta_1 \in S\right) = tP\left(\frac{R_1}{b_1(t)}\right)H_1(S) = t\frac{1}{b_1^{\leftarrow}(b_1(t)x)}H_1(S). \tag{3.27}$$

Also since $H_1 \not\equiv 0$, $H_1(\mathcal{N}\backslash\mathcal{N}_1) = 0$, $\mathbf{V}$ is extremally dependent. So $\mathbf{Z}$ is a mixture of a multivariate regularly varying random vector $\mathbf{V}$ which is extremally dependent and a random vector whose distribution is multivariate regularly varying and completely asymptotically independent. Next observe

that

$$\Theta_{\mathbf{Z}} = \mathbb{I}_{I=0}\Theta_1 + \sum_{i=1}^{d} \mathbb{I}_{I=i} e_i$$

and

$$R_{\mathbf{Z}} = 2^{1/\alpha_1}\mathbb{I}_{I=0}R_1 + 2^{1/\alpha}\sum_{i=1}^{d} \mathbb{I}_{I=i} e_i.$$

Then for S_1, a compact subset of $\mathcal{N}_1$ (and hence $e_i \neq S_1, i = 1, \ldots, d$), we have

$$\begin{aligned} tP\left(\frac{R_{\mathbf{Z}}}{b_1(t)} > x, \Theta_{\mathbf{Z}} \in S_1\right) =& tP\left(I = 0,\ 2^{1/\alpha_1}\frac{R_1}{b_1(t)} > x, \Theta_1 \in S_1\right) \\ &+ \sum_{i=1}^{d} tP\left(I = i, 2^{1/\alpha}\frac{X_i}{b_1(t)} > x, e_i \in S_1\right) \\ =& \frac{t}{2}P\left(2^{1/\alpha_1}R_1 > b_1(t)x\right)H_1(S_1) + 0 \\ \sim& x^{-\alpha_1}H_1(S_1). \end{aligned}$$

Therefore

$$tP\left(\left(\frac{R_{\mathbf{Z}}}{b_1(t)}, \Theta_{\mathbf{Z}}\right) \in A\right) \xrightarrow{v} \nu_{\alpha_1} \times H_1$$

on $(0, \infty] \times \mathcal{N}_1$. Again, for S, a subset of $\mathcal{N}$, we have that

$$\begin{aligned} tP\left(\frac{R_{\mathbf{Z}}}{b(t)} > x, \Theta_{\mathbf{Z}} \in S\right) =& tP\left(I = 0,\ 2^{1/\alpha_1}\frac{R_1}{b(t)} > x, \Theta_1 \in S\right) \\ &+ \sum_{i=1}^{d} tP\left(I = i,\ 2^{1/\alpha}\frac{X_i}{b(t)} > x, e_i \in S\right) \\ =& \frac{t}{2}P\left(2^{1/\alpha_1}R_1 > b(t)x\right)H_1(S) \\ &+ \frac{t}{2d}P\left(2^{1/\alpha}X_1 > b(t)x\right)\sum_{i=1}^{d}\mathbb{I}_{e_i}(S) \\ \to& 0 + c^* x^{-\alpha}\frac{1}{d}\sum_{i=1}^{d}\mathbb{I}_{e_i}(S), \end{aligned}$$

where $c^* = \nu((1, \infty] \times [0, \infty]^{d-1})$. The 0 in the last line of the previous display results from $b_1^{\leftarrow}(b(t)) \to \infty$, since $\frac{b(t)}{b_1(t)} \to \infty$ implies $\frac{b^{\leftarrow}(t)}{b_1^{\leftarrow}(t)} \to 0$, which in turn implies $\frac{1}{b_1^{\leftarrow}(b(t))} \to 0$ with $t \to \infty$. Thus

$$tP\left(\left(\frac{R_{\mathbf{Z}}}{b(t)}, \Theta_{\mathbf{Z}}\right) \in A\right) \xrightarrow{v} \nu_\alpha \times H \tag{3.28}$$

for $A \in (0, \infty] \times \mathcal{N}$, where H is the uniform distribution on $\{e_1, \ldots, e_d\}$. Therefore, $\mathbf{X}$ and $\mathbf{Z}$ are asymptotically equivalent on both the cones $\mathbb{E}$ and $\mathbb{E}_1$.

Conversely, suppose $\mathbf{X}$ has hidden regular variation with $\frac{b(t)}{b_1(t)} \to \infty$ and is tail-equivalent to $\mathbf{Z}$. Since a completely asymptotically independent distribution cannot have a non-zero hidden spectral measure, and the scaling function $b_1(t)$ of the extremally dependent part is of smaller order than that of the completely asymptotically independent part, the hidden spectral measure of F is a multiple of the hidden spectral measure of the extremally dependent part, which is finite. This completes the proof. □

This is a great result. Mixture models have great analytical tractability through the independence of the components, and we can therefore separate them quite easily. We will give an application of this in Chapter 4 and 5, which will focus on a financial application, where we will use a mixture model to refine a quantile estimate for a given portfolio. Theorem 10 shows that a finite spectral measure implies that we can separate the hidden effects from the asymptotic model, so a first intuition in understanding the significance of having a model with a finite hidden spectral measure would be that the hidden effects and the asymptotic effects do not interact with each other in a certain sense. There are further properties that we can benefit from, as we will show now.

Proposition 13. *Suppose* $\mathbf{V}$ *is regularly varying on a subcone* $\mathbb{E}_1$ *of* $\mathbb{E}$*, with index* α_1*, scaling function* $b_1(t)$*, limit measure* ν_1 *and spectral measure* H_1 *on* $\mathcal{N}_1 := \mathcal{N} \cap \mathbb{E}_1$*. The following are equivalent:*

1. H_1 *is finite on* $\mathcal{N}_1$*;*
2. *there exists a random vector* $\mathbf{V}_1$ *tail equivalent to* $\mathbf{V}$ *defined on* $\mathbb{E}_1$*, so that*
$$tP\left(\mathbf{V}_1^{(i)} > b_1(t)x\right) \to cx^{-\alpha_1}, \quad i = 1, \ldots, d,$$
for some $c > 0$*, so that each component of* $\mathbf{V}_1$ *has regularly varying tail probabilities with index* α_1*;*
3. *there exists a random vector* $\mathbf{V}_1$ *tail equivalent to* $\mathbf{V}$*, defined on* $\mathbb{E}_1$ *and regularly varying on the full cone* $\mathbb{E}$ *with scaling function* $b_1(t)$ *and limit measure* ν *and*
$$\nu|_{\mathbb{E}_1} = \nu_1;$$
that is, the restriction of ν *to* $\mathbb{E}_1$ *is* ν_1*.*

Proof. We will prove this statement as well; it is partially adapted from [36]. 1. $\Rightarrow$ 3. : H_1 is defined on $\mathbb{E}_1$ but can be extended to all of $\mathbb{E}$ by setting $H_1(e_i) = 0$. Now repeat the construction of R_1, Θ_1 given in the proof of Theorem 10 and this time define $\mathbf{V}_1 = R_1\Theta_1$ on $\mathbb{E}$, which will be regularly varying on $\mathbb{E}$ with scaling function $b_1(t)$, limit measure ν with $\nu|_{\mathbb{E}_1} = \nu_1$.

3. $\Rightarrow$ 1. : If $\mathbf{V}_1$ is regularly varying on $\mathbb{E}$, then

$$tP\left(\|\mathbf{V}_1\| > b(t)\right) \to \nu\left(\{\mathbf{x} \in \mathbb{E}_1 : \|\mathbf{x}\| > 1\}\right)$$

Therefore

$$\nu\left(\{\mathbf{x} \in \mathbb{E}_1 : \|\mathbf{x}\| > 1\}\right) = \nu_1\left(\{\mathbf{x} \in \mathbb{E}_1 : \|\mathbf{x}\| > 1\}\right) = H_1(\mathcal{N}_1) < \infty.$$

3. $\Rightarrow$ 2. : Since

$$tP\left(\frac{\mathbf{V}_1}{b_1(t)} \in A\right) \to \nu(A)$$

on $\mathbb{E}$, we have that

$$\begin{aligned} tP\left(\frac{\mathbf{V}_1^{(i)}}{b_1(t)} > x\right) &= tP\left(\frac{\mathbf{V}_1}{b_1(t)} \in (x, \infty] \times [0, \infty]^{d-1}\right) \\ &\to \nu\left((x, \infty] \times [0, \infty]^{d-1}\right). \end{aligned}$$

This implies marginal regular variation.
2. $\Rightarrow$ 3. : Suppose $\mathbf{V}_1$ is regularly varying on $\mathbb{E}_1$ and the one-dimensional marginal tails are also regularly varying. For $x \in \mathbb{E}$ such that at least two components are strictly positive, we have by inclusion-exclusion that

$$\begin{aligned} tP\left(\frac{\mathbf{V}_1}{b_1(t)} \in [\mathbf{0}, \mathbf{x}]^c\right) &= tP\left(\bigcup_{i=1}^{d}\left\{\frac{\mathbf{V}_1^{(i)}}{b_1(t)} > x_i\right\}\right) \\ &= t\sum_{i=1}^{d} P\left(\frac{\mathbf{V}_1^{(i)}}{b_1(t)} > x_i\right) - t\sum_{1 \leqslant < j \leqslant d} P\left(\frac{\mathbf{V}_1^{(i)}}{b_1(t)} > x_i, \frac{\mathbf{V}_1^{(j)}}{b_1(t)} > x_j\right) \\ &\quad + \cdots + (-1)^{d+1} tP\left(\frac{\mathbf{V}_1}{b_1(t)} \in [\mathbf{0}, \mathbf{x}]^c\right). \end{aligned}$$

Convergence of the first sum results from one-dimensional marginal convergence. Convergence of the other terms results from vague convergence on $\mathbb{E}_1$. This concludes the proof. □

A finite hidden spectral measure therefore implies that if we have a separated hidden component $\mathbf{V}$ that is regularly varying with tail index α_1, then an arbitrary scalar product $\langle \mathbf{c}, \mathbf{V} \rangle$ is also regularly varying with the same tail index. We can also investigate $\mathbf{V}$ on the full cone, since we can expand it from $\mathbb{E}_1$ to the full cone $\mathbb{E}$.

Careful re-inspection of Example 2 raises the following question: is $\mathbb{E}_1$ the only subcone that can be taken into consideration when $d > 2$? When assuming the leading term in the calculation to be finite, we require the terms of smaller order to converge to zero. Since $\mathbb{E}_1$ is chosen to have at

least two positive components, we could thus define a subcone of $\mathbb{E}_1$, such that at least three components are positive and normalize with even shorter tails to explore further hidden effects. We can push this idea even further and can define a whole sequence of cones $\mathbb{E} = \mathbb{E}_1 \supset \mathbb{E}_1 \supset \cdots \supset \mathbb{E}_{d-1}$, such that

$$E_l := [\mathbf{0}, \infty] \backslash \bigcup_{1 \leqslant j_1 \leqslant \ldots \leqslant j_{d-l} \leqslant d} \left\{ x^{j_1} = 0, \ldots, x^{j_{d-l}} = 0 \right\} \tag{3.29}$$

is the cone with more than l positive components. This also allows us to formulate hidden regular variation without the presence of asymptotic independence. The definition of hidden regular variation on these nested cones is analogous to the standard case. Suppose we have regular variation of a d-dimensional random vector $\mathbf{X}$ on $\mathbb{E}_j$ with scaling function $b_j(t) \in RV_{1/\alpha_j}$ and limiting Radon measure $\nu_j \not\equiv 0$. Assume further that for some $j < l \leqslant d$, $\nu_j(\mathbb{E}_{l-1}) > 0$ and $\nu_j(\mathbb{E}_l) = 0$. The distribution of $\mathbf{X}$ has hidden regular variation on $\mathbb{E}_l$, if, in addition to regular variation on $\mathbb{E}_j$, there is a non-decreasing function $b_l(t) \in RV_{1/\alpha_l}$ such that $b_j/b_l \to \infty$ and a non-negative Radon measure ν_l on $\mathbb{E}_l$, such that

$$tP\left(\frac{\mathbf{X}}{b_l(t)} \in A\right) \xrightarrow{v} \nu_l(A) \tag{3.30}$$

in $M_+(\mathbb{E}_l)$ for an arbitrary Borel set $A \in \mathbb{E}_l$. For example, for the case $d = 3$, we have $\mathbb{E}$ as the full cone, $\mathbb{E}_1$ the cone with the axes removed and $\mathbb{E}_2$ removes the axes, and the faces from $\mathbb{E}$ and fully concentrates on the inner subcone. Formulating hidden regular variation on higher dimensions with these cones is not a trivial task and can be demonstrated with the following example.

Example 5. Suppose that X and Y are i.i.d. standard-Pareto distributed random variables and that

$$\mathbf{Z} = (X, 2XY, Y).$$

It is easy to see that

$$tP\left(\frac{\mathbf{Z}}{2t} \in A\right) \xrightarrow{v} \nu(A), \quad t \to \infty,$$

in $M_+(\mathbb{E})$ and ν has all non-zero marginals. However, since X and $2X$ are not asymptotically independent, we have no asymptotic independence of $\mathbf{Z}$, and since we have positive weight on $\mathbb{E}_1$, we have no hidden regular variation

on $\mathbb{E}_1$. However, we find

$$\begin{aligned}
\nu(\mathbb{E}_2) &= \lim_{y\to 0} \nu\left(\{x_1 \wedge x_2 \wedge x_3 > y\}\right) \\
&= \lim_{y\to 0} \lim_{t\to\infty} tP\left(X > ty, 2X > ty, Y > ty\right) \\
&= \lim_{y\to 0} \lim_{t\to\infty} tP\left(X > ty, Y > ty\right) \\
&= \lim_{y\to 0} \frac{1}{ty^2} \\
&= 0.
\end{aligned}$$

This suggests seeking hidden regular variation on $\mathbb{E}_2$, and indeed this holds with $b_2(t) = \sqrt{t}$, since for $y_1, y_2, y_3 > 0$ we have

$$\begin{aligned}
&\lim_{t\to\infty} P\left(X > \sqrt{t}y_1, 2X > \sqrt{t}y_2, Y > \sqrt{t}y_3\right) \\
&= \lim_{t\to\infty} tP\left(X > \sqrt{t}\left(y_1 \vee \frac{y_2}{2}\right), Y > \sqrt{t}y_3\right) \\
&= \lim_{t\to\infty} t\left(\sqrt{t}\left(y_1 \vee \frac{y_2}{2}\right)\right)^{-1}\left(\sqrt{t}y_3\right)^{-1} \\
&= \frac{1}{\left(y_1 \vee \frac{y_2}{2}\right) y_3}.
\end{aligned}$$

We can summarize:

1. Regular variation holds on $\mathbb{E}$ and $\mathbb{E}_1$, and $\nu(\mathbb{E}) = \nu(\mathbb{E}_1) = \infty$ and $\nu(\mathbb{E}_2) = 0$.

2. Asymptotic independence is absent, but hidden regular variation holds on $\mathbb{E}_2$.

A very detailed discussion of hidden regular variation on sequential cones is given in [39].

This concludes the chapter on hidden regular variation. We will now turn to a financial application and demonstrate how hidden regular variation can help to refine quantile estimates in portfolio risk problems.

Chapter 4

Financial assets with heavy-tailed returns

4.1 Extreme Value Dependence in Financial Markets

Recent studies in finance have highlighted the importance of rare events in asset pricing and portfolio choice. These rare events might be a large change in investment returns, a stock market crash, major defaults, or the collapse of risky asset prices.

The mathematical basis of portfolio diversification was given by Markowitz in 1952 (see [35]) for multivariate Gaussian models. The mean-variance approach of Markowitz has since been extended to many other models and model classes and is one of the central elements in mathematical portfolio theory. Under these assumptions, diversification is always good in the sense that the portfolio risk is always minimized by a mixed portfolio. However, this is not true in general. Negative diversification effects in heavy-tailed models with independent asset returns are known at least since Fama and Miller (see [16]). Even before, other authors noted that, although very practical, Markowitz' assumptions were much too simple. Mandelbrot discussed in [34] that the assumption of Gaussian returns might be very practical, but he suggested to rather use general Paretian laws to model returns, since extreme events occured much more often than they were supposed to under Gaussian assumptions, which has a big impact on how an investor has to evaluate his portfolio and the risk to which he is exposed.

Starting from the early 1960s, the safety-first criterion in constructing portfolios has received more and more attention in risk management. The Value-at-Risk (VaR), which considers only the downside risk, has become a well-accepted risk measure, since it does not rely on specific distribution assumptions. For example, within the Gaussian framework, Gouriéroux investigated in [22] the sensitivity of the VaR. In fact, when fixing the mean

as zero, in a Gaussian framework, the VaR is directly connected with the variance, so it can be seen as a meaningful extension of the approach of Markowitz. Variance analysis is in fact a Value-at-Risk approach. Similarly, in Fama and Miller's symmetric stable framework, when the characteristic component α is fixed, the Value-at-Risk is directly associated to the scale parameter. Hence, a study of the scale parameter in a symmetric stable framework is also a Value-at-Risk approach. We can interpret the scale parameter as an (eventually) constant slowly varying function and therefore also consider the findings of stable distributions in the framework of regularly varying distributions.

In the recent years, there have been numerous applications of extreme value theory and multivariate regular variation in the financial industry, especially studies of which limit characteristics certain markets have and how these can be assessed with risk measures like the Value-at-Risk, for, as we have seen in previous chapters, the theory of regular variation gives a very accessible framework for markets with heavy-tailed return distributions. The assessment of portfolio risk is inherently different from evaluating the risk potential of a single asset. We don't just have to consider the marginal distributions, it's also fundamentally important how the assets of the portfolio are related, especially for large losses. With the theory developed in Chapter 3, we have a framework which can be used to evaluate large losses in markets, based on the tail indices, individual scaling functions and the dependence structure among those assets. As one can imagine, problems arise when the dependence structure given by extreme value theory is too coarse, as in particular with asymptotically independent assets; if we think back to the observation of Sibuya that multivariate normal-distributed random vectors possess asymptotic independence if they are not perfectly dependent, then we see that simply using the asymptotic behaviour as an approximation to describe the portfolio risk is inadequate. Overlooking the potential dependence, especially for large losses, can lead to severe problems in evaluating the aggregated risk. One of our key findings is how the assumed absence of systemic risk leads to a potentially strong underestimation of the aggregated risk in a portfolio and how hidden, asymptotically weaker, effects can still lead to a substantially higher worst-case loss. Based on the theory developed in this thesis, we will investigate how we can improve cases when we handle a market with asymptotic independent assets with searching for weaker systemic risk factors with the findings from hidden regular variation. We first give a short introduction to the portfolio context.

4.1.1 Portfolio losses and risks of heavy-tailed assets

Consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with heavy-tailed marginals in the positive orthant of $\mathbb{R}^d$ representing losses of asset returns. We make the additional assumption that all tail indices are equal to $\alpha > 0$ and that all marginals are tail-equivalent, so $\mathbf{X}$ is multivariate regularly varying as defined in Chapter 3. This assumption is asymptotic and quite restrictive. Multivariate regularly varying models are often criticized for excluding even slightly different marginal tail indices α_i for the components X_i. However, this criticism also affects the multivariate t and multivariate α-stable models, which are widely accepted in practice despite the restriction to a common index α.

If we estimate the tail index α_i for each component X_i separately, one would hardly ever obtain identical values for different i, though the confidence intervals for α_i often overlap; a multivariate regularly varying model with a common tail index α may be close enough to reality and provide a useful result. A common practice is to group assets into several sub-portfolios, each with a common index, which is discussed in [33]. We restrict our asset choice to the sub-portfolio with the smallest tail index α. Zhou shows in [50] that this is a choice worth considering, since the largest portfolio losses will be asymptotically caused by those assets alone, so we can focus ourselves on diversifying away risk on this sub-portfolio. The loss of a portfolio is given by

$$Y = \sum_{i=1}^{d} c_i X_i \tag{4.1}$$

where $\mathbf{c} = (c_i, \ldots, c_d)$ is a vector of portfolio weights, and is also regularly varying with the same tail index α. Since the risk of loss on a short sale is theoretically infinite, we restrict ourselves to long-only portfolios

$$\mathbf{c} \in \left\{ \mathbf{y} \in \mathbb{R}_+^d : \|\mathbf{y}\|_1 = 1 \right\},$$

meaning that $\mathbf{c}$ will be chosen from the unit simplex in $\mathbb{R}_+^d$. We now turn to the Value-at-Risk. Given a confidence level $\delta \in (0, 1)$, the Value-at-Risk of the portfolio at the confidence level δ is given by the smallest number x such that the probability that the loss L exceeds x is at most δ. We denote this by

$$\mathrm{VaR}_\delta(L) = \inf \{ x \in \mathbb{R} : P(L > x) \leqslant \delta \}. \tag{4.2}$$

Since we assume the marginal distributions of $\mathbf{X}$ to be regularly varying, we know by Proposition 3 that the portfolio Y is also regularly varying with tail index α. This gives the following implicit expression for the Value-at-Risk

for a given confidence level δ:

$$\begin{aligned}\delta &= P(Y > \mathrm{VaR}_\delta(Y)) \\ &= l(\mathrm{VaR}_\delta(Y))\mathrm{VaR}_\delta(Y)^{-\alpha},\end{aligned}$$

with $l \in RV_0$ is given by the distribution of P. This implies the implicit formulation

$$\mathrm{VaR}_\delta(Y) = \left(\frac{l(\mathrm{VaR}_\delta(Y))}{\delta}\right)^{1/\alpha} = \left(\frac{a(\delta)}{\delta}\right)^{1/\alpha}. \tag{4.3}$$

Since we ignore any possible hidden regular variation in this approach, we refer to this as the first-order approximation of the Value-at-Risk. We call the function $a(\delta) := l(\mathrm{VaR}_\delta(Y))$ the tail scale function of the random variable P. As pointed out in [50], if the marginals of the single assets are tail equivalent to a common distribution, the distribution of the portfolio is also tail equivalent to that distribution and we can calculate the VaR for a given portfolio as the VaR of the tail-equivalent random variable, adjusted with an integral term describing the tail dependence structure given in the limit. This gives a simple formula to approximate the Value-at-Risk of a portfolio, which can be used in an optimization problem to minimize $\mathrm{VaR}_\delta(Y)$ as a function of the portfolio weights. We ignore possible subasymptotic dependence with this, which means that we completely ignore systemic risk. As we have discussed in Chapter 3, this can be a very bad decision, especially in the case of asymptotic independent assets, where we can have almost arbitrarily strong sub-asymptotic dependence, but assume complete independence as an approximation at subasymptotic levels. Another theoretical example for a heavy-tailed scenario would be an asymptotic independent portfolio with tail index α, which possesses hidden regular variation with tail index $\alpha_1 = \alpha + \epsilon$ for an arbitrary $\epsilon > 0$. Although this dependence structure converges to an independent vector, we might have very heavy-tailed second-order effects which will strongly affect every investment decision.

4.1.2 Value-at-Risk under hidden regular variation

We will develop an implicit analytical expression that approximates the Value-at-Risk for a portfolio when we have investment choices in a market with asymptotic independent marginals, but which still possess systemic risks that have to be incorporated in an optimal portfolio. For this we will work with a surrogate mixture model, which has favourable properties in the sense that we can develop an analytical expression that can be used to calculate the Value-at-Risk of a portfolio with given asset allocation. Since we don't want to introduce systematic errors, we require our surrogate model to be tail equivalent to the true model on all cones on which we seek effects. We

consider the Value-at-Risk computed with that model as an approximation to the true model and will show in Chapter 5 how good this approximation is. If we make the assumption of a finite hidden measure on the sub-cone where we look for hidden effects, we know from Theorem 10 that there exists a mixture model that we can choose to be tail equivalent on all considered cones. Choosing a finite hidden measure may seem restrictive, but from a statistical point of view, any estimated empirical measure is always finite and one can argue that for finite levels in relevant applications we are not far out enough in the tail to really be restricted by this assumption. We can assume without loss of generality that $\mathbf{X}$ is tail-equivalent to a random vector $\mathbf{W}$ with equal marginal scale functions in a way that

$$X_i = \frac{W_i}{m_i^{1/\alpha}}, \quad i = 1, \dots, d,$$

on every considered sub-cone. This gives more freedom in the market model that we will consider in the coming theorem, which we will formulate for the two cones $\mathbb{E}$ and $\mathbb{E}_1$; this can be generalized to more cones in an analogous way.

Theorem 11. *Suppose the losses $\boldsymbol{X} = (X_1, \dots, X_d)$ of a market with d assets are multivariate regularly varying with tail index α on the cone $\mathbb{E}$ and possess hidden regular variation on $\mathbb{E}_1$ with tail index $\alpha < \alpha_1 < 2\alpha$ and a finite measure H_1. Under the assumption that $\boldsymbol{X}$ has comparable marginals on both cones, the Value-at-Risk for a portfolio $Y = \sum_{i=1}^d c_i X_i$ can be approximated by solving the implicit equation*

$$\begin{aligned}\delta = {} & x^{-\alpha} \frac{1}{d} \sum_{i=1}^{d} l_1\left(g_1(x, c_i, m_i)\right) m_i c_i^{\alpha} \\ & + d\; x^{-\alpha_1} \int_{\mathcal{N}} l_2\left(g_2(x, \boldsymbol{c}, \boldsymbol{m}^*, \boldsymbol{\omega})\right) \left(\sum_{i=1}^{d} c_i m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1} \right)^{\alpha_1} H_1(d\omega),\end{aligned}$$

where $l_1, l_2 \in RV_0$.

Proof. We will show that we can find a solution via a tail-equivalent mixture model. Since H_1 is finite, we know by Theorem 10 that there exists a random vector $\mathbf{Z}$ that is tail equivalent on both $\mathbb{E}$ and $\mathbb{E}_1$ and that is representable as a mixture. To ensure tail equivalence to $\mathbf{X}$ we need to make sure that it is also tail equivalent to the scaled version, resulting in a mixture of the form

$$\mathbf{Z} = 2^{1/\alpha_1} \mathbb{I}_{I=0} M \mathbf{V} + 2^{1/\alpha} \sum_{i=1}^{d} \mathbb{I}_{I=i} \tilde{X}_i m_i^{1/\alpha} e_i, \tag{4.4}$$

with $I, \mathbf{V}, \tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_d)$ independent random variables defined as in the proof of Theorem 10 and

$$M = \operatorname{diag}\left(\left\{\frac{1}{m_1^{*1/\alpha_1}}, \ldots, \frac{1}{m_d^{*1/\alpha_1}}\right\}\right).$$

Now define $\tilde{P} = \sum_{i=1}^{d} c_i Z_i$. We will calculate $x = \mathrm{VaR}_\delta(\tilde{P})$ to approximate $\mathrm{VaR}_\delta(Y)$, since, as we will now see, we can analytically separate first- and second-order effects with that mixture. From Proposition 13 we know that we can choose V to have regularly varying marginals and that the scalar product is therefore regularly varying with the same tail index. We begin with

$$\begin{aligned}
\delta &= P\left(\sum_{i=1}^{d} c_i Z_i > x\right) \\
&= P\left(\sum_{i=1}^{d} c_i (2^{1/\alpha_1} \mathbb{I}_{I=0} m_i^{*1/\alpha_1} V_i + 2^{1/\alpha} \mathbb{I}_{I=i} m_i^{1/\alpha} \tilde{X}_i e_i) > x\right) \\
&= P\left(I = 0,\ \sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} V_i > x\right) + \sum_{i=1}^{d} P\left(I = i,\ c_i 2^{1/\alpha} m_i^{1/\alpha} \tilde{X}_i > x\right) \\
&= \frac{1}{2} P\left(\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} V_i > x\right) + \frac{1}{2d} \sum_{i=1}^{d} P\left(\tilde{X}_i > m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha} x\right) \\
&= \frac{1}{2}(I + II),
\end{aligned}$$

We will treat these terms separately and will first develop an expression for the first term. Since we have the distribution of the spectral decomposition, we will base ourselves on the proof of Theorem 9 and calculate a representation with respect to H_1. We first write the scalar product in set form

$$I = P\left(\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} V_i > x\right) = P\left(\mathbf{V} \in A^*\right),$$

with the set A^* defined as

$$A^* := \left\{\mathbf{v} \in \mathbb{E}_1 :\ \sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} v_i > x\right\}.$$

Since the measure H_1 is defined via the spectral decomposition of an exponent measure, we first need to transform the measure $P \circ \mathbf{V}^{-1}$ to the unit scale with the transformation

$$U : \mathbb{R} \to \mathbb{R},\ x \mapsto x^{\alpha_1}.$$

Consider also the spectral decomposition h defined in equation (3.11). Then

$$\begin{aligned}
P\left(\mathbf{V} \in A^*\right) &= \left(P \circ h \circ U\right)\left(\mathbf{V} \in (U^{-1} \circ h^{-1})(A^*)\right) \\
&= P\left(\left(R_1^{1/\alpha_1}, \Theta_1\right) \in \left\{(r, \boldsymbol{\omega}) : \sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} (r\omega_i)^{1/\alpha_1} > x\right\}\right) \\
&= d \int_{\mathcal{N}} \int_{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} (r\omega_i)^{1/\alpha_1} > x} 1 \;\; (F_1 \circ U^{-1})(dr) H_1(d\omega) \\
&= d \int_{\mathcal{N}} \int_{r > \left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{\alpha_1}} 1 \;\; (F_1 \circ U^{-1})(dr) H_1(d\omega),
\end{aligned}$$

where F_1 is defined as in Theorem 10. Corollary 1 gives

$$(F_1 \circ U^{-1})(x) = x^{-1} l_2(x)$$

for an asymptotically unique slowly varying function $l_2(x)$. This results in

$$\begin{aligned}
& d \int_{\mathcal{N}} \int_{r > \left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{\alpha_1}} 1 \;\; (F_1 \circ U^{-1})(dr) H_1(d\omega) \\
&= d \int_{\mathcal{N}} l_2 \left(\left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{\alpha_1} \right) \left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{-\alpha_1} H_1(d\omega) \\
&= 2d \; x^{-\alpha_1} \int_{\mathcal{N}} l_2 \left(\left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{\alpha_1} \right) \left(\sum_{i=1}^{d} c_i m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1} \right)^{\alpha_1} H_1(d\omega) \\
&= 2d \; x^{-\alpha_1} \int_{\mathcal{N}} l_2 \left(g_2(x, \mathbf{c}, \mathbf{m}^*, \boldsymbol{\omega}) \right) \left(\sum_{i=1}^{d} c_i m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1} \right)^{\alpha_1} H_1(d\omega),
\end{aligned}$$

where

$$g_2(x, \mathbf{c}, \mathbf{m}^*, \boldsymbol{\omega}) = \left(\frac{x}{\sum_{i=1}^{d} c_i 2^{1/\alpha_1} m_i^{*1/\alpha_1} \omega_i^{1/\alpha_1}} \right)^{\alpha_1}.$$

Term II is a straightforward calculation. Since

$$\begin{aligned}
P\left(\tilde{X}_i > m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha} x\right) &= l_1\left(m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha} x\right) \left(m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha}\right)^{-\alpha} \\
&= 2x^{-\alpha} l_1\left(m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha} x\right) m_i c_i^{\alpha} \\
&= 2x^{-\alpha} l_1\left(g_1(x, c_i, m_i)\right) m_i c_i^{\alpha}
\end{aligned}$$

with

$$g_1(x, c, m) = m^{-1/\alpha} c^{-1} 2^{-1/\alpha} x,$$

we can collect the terms

$$\begin{aligned} II &= \frac{1}{d}\sum_{i=1}^{d} P\left(\tilde{X}_i > m_i^{-1/\alpha} c_i^{-1} 2^{-1/\alpha} x\right) \\ &= \frac{2}{d}\sum_{i=1}^{d} x^{-\alpha} l_1\left(g_1(x, c_i, m_i)\right) m_i c_i^{\alpha} \end{aligned}$$

and get as a final result

$$\begin{aligned} \delta = {} & x^{-\alpha}\frac{1}{d}\sum_{i=1}^{d} l_1\left(g_1(x, c_i, m_i)\right) m_i c_i^{\alpha} \\ & + d\; x^{-\alpha_1}\int_{\mathcal{N}} l_2\left(g_2(x, \mathbf{c}, \mathbf{m}^*, \boldsymbol{\omega})\right)\left(\sum_{i=1}^{d} c_i {m_i^*}^{1/\alpha_1}\omega_i^{1/\alpha_1}\right)^{\alpha_1} H_1(d\omega), \end{aligned} \tag{4.5}$$

as we wanted to show. The proof is complete. □

We can easily show that equation (4.5) has always a solution; for any given portfolio weights $\mathbf{c}$ and scaling factors $\mathbf{m}, \mathbf{m}^*$, the sum terms and the integral are bounded, so there exist constants $C_1, C_2 > 0$ such that

$$C_1 x^{-\alpha} + C_2 x^{-\alpha_1} = \delta. \tag{4.6}$$

Since the left-hand side is a continuous and monotonically decreasing function with

$$\begin{aligned} \lim_{x\to 0} C_1 x^{-\alpha} + C_2 x^{-\alpha_1} &= \infty, \\ \lim_{x\to\infty} C_1 x^{-\alpha} + C_2 x^{-\alpha_1} &= 0, \end{aligned} \tag{4.7}$$

there exists one unique Value-at-Risk that solves problem (4.5) for a given set of parameters.

In practice, we not only try to estimate the losses that a portfolio invested in a specific market is expected to have, but are interested in how we can construct a portfolio to replicate a certain loss, with focus on replicating the portfolio with the smallest losses, therefore minimizing the risk the portfolio is exposed to when invested in a given market. We call this process of minimizing loss diversification. A natural question is whether there exists a possibly unique portfolio which minimizes the market risk. As discussed in [50], this question is closely related to the results of Fama and Miller (see [16]). In the case of a simple multivariate regular varying market, we can check that when $\alpha > 1$ we can find a uniquely determined risk-minimizing portfolio, so diversification decreases the cumulative risk of the portfolio; when $\alpha = 1$, diversification has no effect on the portfolio; and when case $\alpha < 1$, we increase the risk of the portfolio when we spread out the portfolio

over the market, so the optimal choice is therefore to compare the individual risks and fully invest into the asset with the lowest likelihood of being the most risky, which is called the probability of dominance, see [50]. In order to compare how the individual risk factors link to systemic risk, we can define an "extreme" situation for the system and consider the conditional probability that the i-th factor generates the highest loss, given the fact that the extreme loss occurs, so comparing the probabilities

$$p_i = P\left(X_i = \bigvee_{j=1}^{d} X_j \,\middle|\, \bigvee_{j=1}^{d} X_j > t\right)$$

of each asset in the portfolio for a large threshold t and considering the asset with the smallest of these probabilities. This is implied by

Proposition 14. *Let $\boldsymbol{X} = (X_1, \ldots, X_d) \in MRV_\alpha$ with $\alpha > 0$ and let $\boldsymbol{c}$ be an admissible vector of portfolio weights. Then the map*

$$c \mapsto \lim_{t\to\infty} \mathrm{VaR}_\delta\left(\sum_{i=1}^{d} c_i X_i\right) \tag{4.8}$$

mapping the portfolio weights to the Value-at-Risk of a corresponding portfolio is

- *convex for $\alpha > 1$,*
- *linear for $\alpha = 1$,*
- *concave for $\alpha < 1$.*

Proof. This can be shown by the Minkowski inequality. See [32], Theorem 2.4 for example. □

So we can expect for δ close to zero that we can decide through the tail index α if diversification can be of any benefit or if losses tend to be so heavy-tailed that adding new assets to the portfolio would not reduce the overall risk assessed by the Value-at-Risk. However, when we consider a market with hidden regular variation and the same preconditions as in Theorem 11, we can observe a refinement of this statement for $\alpha = 1$ and show that there is indeed a unique risk-minimizing portfolio under certain technical assumptions.

Corollary 3. *Assume that $\boldsymbol{X}$ is multivariate regularly varying with tail index $\alpha = 1$ and possesses hidden regular variation on a subcone $\mathbb{E}_1$ with a finite hidden spectral measure H_1 and hidden tail index $1 < \alpha_1 < 2$. Then $\boldsymbol{X}$ is tail-equivalent to a mixture model whose Value-at-Risk is described via an implicit function given by equation (4.5). If $l_1 \circ g_1$ is constant with respect to $\boldsymbol{c}$, then diversification is advantageous and there exists a unique, risk-minimizing portfolio.*

Proof. Since H_1 is finite on $\mathbb{E}_1$, we know from Theorem 10 that $\mathbf{X}$ is tail equivalent to a mixture surrogate model as described in the theorem. Define

$$F(c_1,\dots,c_d,x) = \frac{1}{dx}\sum_{i=1}^{d} l_1\left(g_1(x,c_i,m_i)\right)c_i m_i \tag{4.9}$$
$$+ x^{-\alpha_1}\, d\int_{\mathcal{N}} l_2\left(g_2(x,\mathbf{c},\mathbf{m}^*,\boldsymbol{\omega})\right)\left(\sum_{i=1}^{d} c_i {m_i^*}^{1/\alpha_1}\omega_i^{1/\alpha_1}\right)^{\alpha_1} H_1(d\omega) - \delta$$

as the level set function of the Value-at-Risk problem. We assume at this point that the functions l_1 and l_2 behave "well enough" in a sense that they are smooth enough, which can be ensured by choosing an asymptotically equivalent version, so we will assume that both functions are differentiable and bounded away from zero. We want to show that there exists an equation $f:\mathbb{R}^d\to\mathbb{R}$ such that

$$F(c_1,\dots,c_d,f(c_1,\dots,c_d)) = 0,$$

where f is the actual Value-at-Risk for a given portfolio vector $\mathbf{c}$. As we can see from equations (4.6) and (4.9), we can find an open subset U of $\mathbb{R}$ in which

$$\begin{aligned}\frac{\partial F}{\partial x} = {}& \frac{-1}{dx^2}\sum_{i=1}^{d} l_1\left(g_1(x,c_i,m_i)\right)c_i m_i + \frac{1}{dx}\sum_{i=1}^{d}\frac{\partial}{\partial x} l_1\left(g_1(x,c_i,m_i)\right)c_i m_i \\ & - x^{-\alpha_1-1}\, d\int_{\mathcal{N}} l_2\left(g_2(x,\mathbf{c},\mathbf{m}^*,\boldsymbol{\omega})\right)\left(\sum_{i=1}^{d} c_i {m_i^*}^{1/\alpha_1}\omega_i^{1/\alpha_1}\right)^{\alpha_1} H_1(d\omega) \\ & + x^{-\alpha_1}\, d\int_{\mathcal{N}} \frac{\partial}{\partial x} l_2\left(g_2(x,\mathbf{c},\mathbf{m}^*,\boldsymbol{\omega})\right)\left(\sum_{i=1}^{d} c_i {m_i^*}^{1/\alpha_1}\omega_i^{1/\alpha_1}\right)^{\alpha_1} H_1(d\omega)\end{aligned}$$

is not zero. The second and fourth terms are due to the product rule. We can therefore assume by the implicit function theorem that such a function f exists and we can proceed to calculate the implicit derivatives of f with respect to $c_1,\dots,c_d$.

Since $\alpha_1 > 1$, we know that the second implicit derivative of f exists and is not zero. Since by assumption $\frac{\partial}{\partial c}(l_1\circ g_1)$ is the zero vector, we can see from equation (4.9) that the Hessian of the first summand with respect to $\mathbf{c}$ is zero. But this implies that the Hessian of f is the same as the Hessian of the map given in (4.8) for $\alpha_1 > 1$, which is positive definite for small δ by Proposition 14. By equivalence this also holds for $\mathbf{X}$ therefore as well. This concludes the proof. □

This corollary demonstrates that although hidden regular effects are

asymptotically weak and therefore hidden under the major regular variational effects, they can still be strong enough to affect the risk of a portfolio and should therefore be considered in the choice of a risk-minimizing portfolio. When $\alpha < 1$ and $\alpha_1 > 1$, this no longer holds. We couldn't find general statements about this situation, but from numerical experiments we suspect that choosing not to diversify the portfolio is still optimal if hidden regular variation is present.

Chapter 5

Simulation study

We will now come to an illustrative simulation study. We will demonstrate how the approximation of the Value-at-Risk and the implicit calculation of the risk-minimizing portfolio weights of a financial portfolio

$$P = \sum_{i=1}^{d} c_i X_i$$

with d assets priced under linear factor models with heavy-tailed components can be improved using the theory of hidden regular variation. Although more general market models are possible, we restrict ourselves to linear factor models to provide a readily understandable analysis.

5.1 The Arbitrage Pricing Theory Model

Several fundamental market models have had an enormous impact on the financial landscape. The precursor of most of the commonly used models today, still used a lot in the industry, is the Capital Asset Pricing Model (CAPM). The CAPM is a single-factor model introduced by Jack Treynor in 1961, William F. Sharpe in 1964, John Lintner in 1965 and Jan Mossin in 1966 independently, building on the earlier work of Harry Markowitz on diversification and modern portfolio theory. It states that given certain assumptions on the market, the expected return X_i on an asset is

$$E(X_i) = \rho_0 + \beta_i(E(R) - \rho_0), \quad i = 1, \dots, d, \tag{5.1}$$

where ρ_0 represents the return of a riskless asset, $E(R)$ is the expected return of the market portfolio and β_i is known as "the beta of the asset". Traynor and Sharpe argued that the risk of holding an asset in a portfolio can be quantified by regressing it linearly against the market portfolio to obtain β, which measures systematic risk relative to the market portfolio as the sole determinant of return. Any additional variability caused by events peculiar to the individual asset can be diversified away: capital markets do not reward risks borne unnecessarily. The CAPM was the first major attempt to explain asset return behaviour and has been developed continuously. Despite its

fame, the CAPM has a lot of shortcomings. In response to the model's empirical failures, Stephen Ross introduced the Arbitrage Pricing Theory Model (APTM) in 1976 as an alternative to the CAPM. The APTM has the potential to overcome the weaknesses of the CAPM: it requires fewer and more realistic assumptions, which are generated by a simple arbitrage argument, and its explanatory power is potentially better, since it is a multi-factor model, whose factors can be practically arbitrarily chosen. A possible factor could be an index or a macroeconomic factor. However, the power and the generality of the APTM are its main strength and weakness: the APTM permits the researchers to choose whatever factors provide the best explanation for the data, but it cannot explain variation in asset returns in terms of a limited number of easily identifiable factors. In contrast, CAPM theory is intuitive and easy to apply, but is often too restrictive in practice. The Arbitrage Pricing Theory Model assumes that the assets follow a linear multi-factor model with n factors, given by

$$X_i = Y_i + \sum_{j=1}^{n} \beta_{ij} R_{ij}, \quad i = 1, \dots, d, \tag{5.2}$$

where we assume that the linear risk factors which make up the asset price follow heavy-tailed distributions. The $(\mathbf{R}_j)_j \in MRV_{\alpha_c}$ are i.i.d. random vectors with fully dependent components, which means that $R_{aj} \sim R_{bj}$ for all $1 \leqslant a, b \leqslant d$. They are furthermore pairwise independent of $\mathbf{Y} \in MRV_{\alpha_{id}}$, which is assumed to be a random vector with completely independent components. The components of $\mathbf{Y}$ are called the idiosyncratic risk factors and $(\mathbf{R}_i)_i$ are called systemic risk factors, shared by at least two assets.

We will now demonstrate how the Value-at-Risk of a portfolio is approximated by the classical setup of regular variation and how this approximation can be improved by taking the second-order effects of the dependence structure into consideration, as elaborated in Chapter 4. Since all individual risk factors are regularly varying and thus so too are linear combinations as shown in Proposition 3, we can apply Theorem 11 to approximate the Value-at-Risk at different levels. We assume that the risk factors are Pareto-distributed, and, to generate an asymptotic independent model, we assume that $\alpha_{id} < \alpha_c$. This means that the idiosyncratic risks are heavier-tailed, they dominate the components of $\mathbf{X}$ and thus define an asymptotic independent model. To focus on the effects through hidden regular variation we assume that all idiosyncratic factors have equal scale 1, so to a first-order approximation the portfolio consists of independent identically distributed assets, which means they are always equally-weighted. Since the second-order effects vanish in the limit, we expect the second-order approximation to converge towards the first-order approximation for $\delta \to 0$, thus converging

to an equally-weighted portfolio. We take this as the benchmark-portfolio for comparison. Asymmetric systemic market risk also means that there is one or more assets that have the largest exposure towards the systemic component, so the worst choice for a portfolio would therefore be to fully invest into this group. This is the second benchmark portfolio we will compare against.

To investigate the performance of the developed analytical approximation, we sample 2,000,000 points each from the original model factor models under different assumptions and calculate the Value-at-Risk for $\delta \in (0, 0.1]$ from the first- and second-order approximation and compare those to the actual empirical Value-at-Risk of the generated data. We will elaborate the details of how each parameter impacts the model behaviour and the approximation quality in several example cases. The main focus lies on the comparison of the performance gain over the equally-weighted portfolio (which is the first-order optimal portfolio) to show the approximation quality of the term described in Theorem 11. After that we compute the optimal portfolio weights for the Value-at-Risk-minimizing portfolio for $\delta \in (0, 0.1]$ implicitly from the samples and from the analytical approximations and compare them. The code is provided in the Appendix.

Simulation parameters						
Number of assets	2	2	2	3	3	3
α_{id}	2	2	2	2	2	2
α_c	3	2.25	3.75	3	3	3
β_1	1.3	1.3	1.3	1.3	1.3	1.3
β_2	0.8	0.8	0.8	0.8	0.8	0.8
β_3	-	-	-	1	1.4	0.7

TABLE 5.1: Parameter for study of the linear single-factor model

5.1.1 Linear single-factor model

The first model assumption will be $n = 1$ in equation 5.2, thus a single-factor model of the form

$$X_i = Y_i + \beta_i R, \quad i = 1, \ldots, d. \tag{5.3}$$

We designed a simulation study to capture how the different parameters of a single-factor model play a role in the quality of a regularly varying first- and second order approximation to the true market model. We limited the maximum number of assets to three. Table 5.1 lists the different combinations investigated, we will first consider bivariate portfolios to demonstrate the major effects by changing the hidden tail coefficient and demonstrate with trivariate portfolios the impact of changing the linear coefficients of the "hidden" systemic component.

Two assets

We start with a visual inspection of the first example, given in Figure 5.1 with parameters given by the first example in Table 5.1, showing the data generated by equation (5.3) on the left and the surrogate model (3.26) on the right, on which our analytical solution is based. The systemic factor is slightly shifted towards X_1, as it has the larger exposure to it with the chosen linear factors. The contribution of the common risk factor is visible in the scatter plot with 10,000 samples, as it generates samples further out in the cone. We can see the asymptotic behaviour of the idiosyncratic factors which concentrate on the axes and, more weakly, also on a single, slightly asymetrical axis in the middle of the cone, induced by the systemic risk factor. As we have shown in Example 2, the assets' idiosyncratic risk factors Y_i will eventually concentrate on the axes. Comparing the mixture simulations with the simulated data from the linear single-factor model, we see that, especially for small values, the quantile of a summed-up mixture portfolio will always be smaller than the true Value-at-Risk of a considered portfolio of assets. This is demonstrated in Figure 5.2. It shows the Value-at-Risk for a bivariate equally-weighted portfolio to demonstrate the behaviour of the

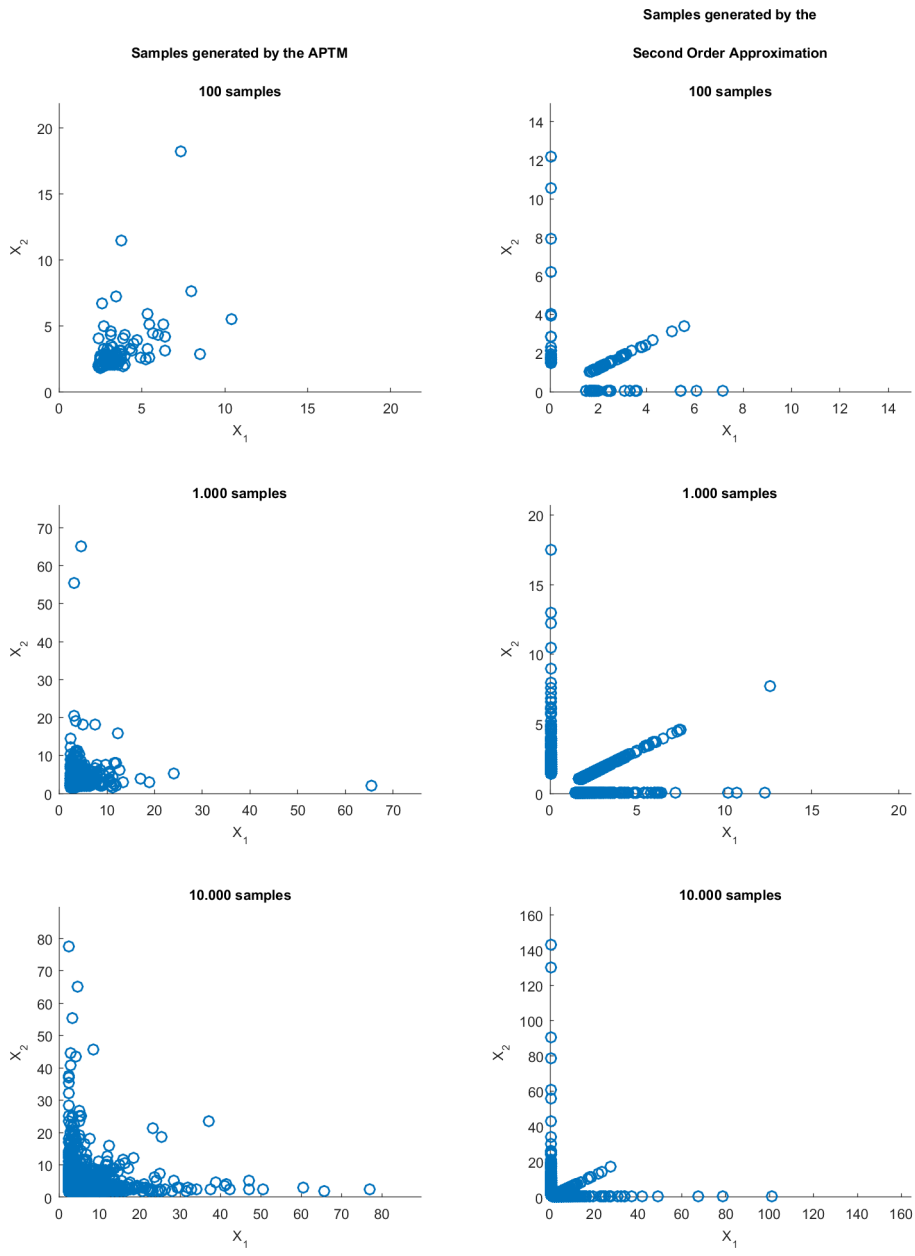

FIGURE 5.1: Samples generated by the true linear single-factor model its and second-order approximation with the mixture method

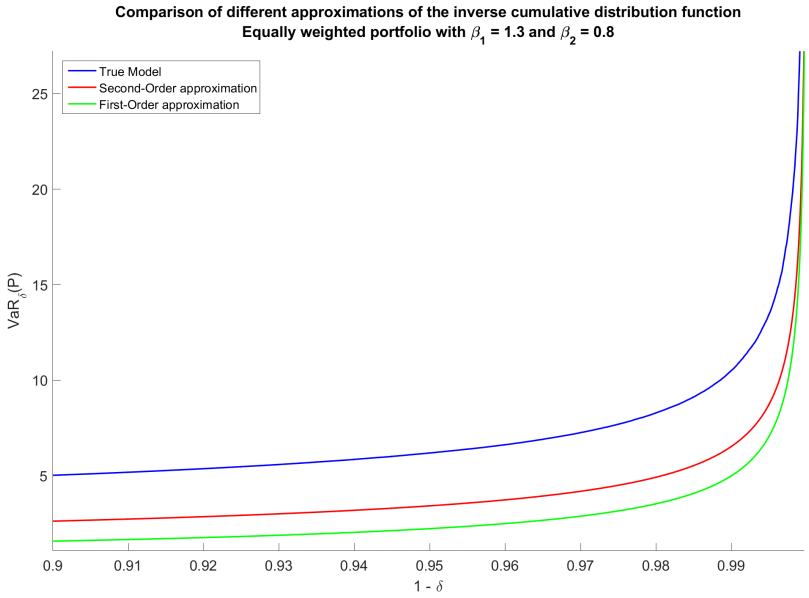

FIGURE 5.2: Approximations to the Value-at-Risk for an equally weighted bivariate portfolio. The true model is calculated by empirical quantile estimation, while the first- and second-order approximations are the analytical solutions of the model and assume an equally-weighted portfolio of two assets with the first configuration described in Table 5.1

three considered models compared to each other. The true model dominates both first- and second-order approximations. We also see a slight improvement through including second-order effects, giving a better approximation of the true Value-at-Risk over the first-order approximation. We will see in other examples that the difference in the two approximations is not always small, and find several conditions which influence the approximation quality of the mixture model positively or negatively.

As shown in Theorem 11, the approximations are tail-equivalent to the true model on $\mathbb{E}$ for the classical solution and additionally on $\mathbb{E}_1$ for the second-order approximation, so both distribution functions have asymptotically the same shape as $\delta \to 0$. Alternatively, we can use (4.9) with the intention of finding the optimal set of portfolio weights, such that the Value-at-Risk for given δ is minimized. Since the mixture model incorporates information about the systemic effects, we expect this to be reflected in the implied portfolio weights.

We cannot formulate an explicit formulation of the Value-at-Risk problem in equation (4.9) as a closed-form function of the portfolio weights but we can show that we can still find an implicit function f defined as the level curve of level 0 of the surface defined by equation (4.9) which expresses the Value-at-Risk for a given set of portfolio weights. We showed this in the proof of Corollary 3 by using the implicit function theorem. This ensures

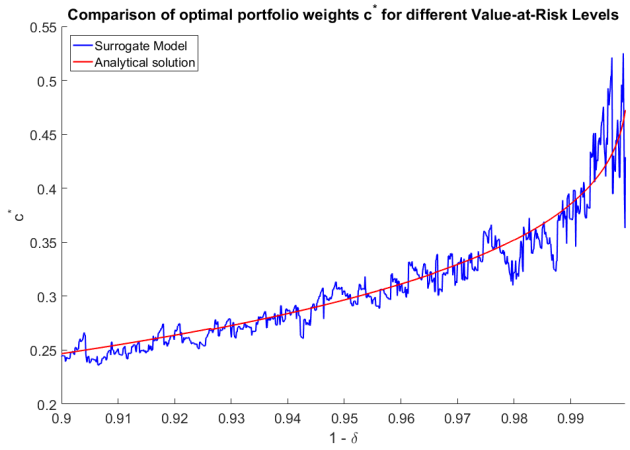

FIGURE 5.3: Comparison between analytical optimal weights and empirical optimal weights of the mixture surrogate model, assuming the first example set of parameters given in Table 5.1.

existence and well-definedness of f and makes it possible to use it as an implicit target function. As discussed in Section 4.1.2, since $1 \leqslant \alpha < \alpha_1$ we know that f must be convex; the following optimization problem therefore has a unique solution:

$$\begin{aligned} \min\ & f(c_1, \ldots, c_d) \\ \text{s.t. } & F(c_1, \ldots, c_d,\ f(c_1, \ldots, c_d)) = 0, \\ & f(c_1, \ldots, c_d) > 0, \\ & \min(c_1, \ldots, c_d) > 0, \\ & \sum_{i=1}^{d} c_i = 1. \end{aligned} \tag{5.4}$$

We solve this as a nested optimization problem by performing a constrained minimization based on finding the root in the first constraint, while ensuring the other three constraints. As we can see from equation 4.9, this problem is not well-conditioned, so with increasing dimension we expect the results to get worse. As a first trivial test, we compare the optimal portfolio weights calculated by (5.4) with the empirical result of the mixture surrogate model in Equation (3.26) and demonstrate a first problem. The sample size is 2,000,000. Figure 5.3 shows that the analytical solution replicates the empirical solution overall, but for Value-at-Risk levels going to $\delta \to 0$ we see the variance increase caused by the lack of data in the extreme tails. The APTM shows the same behaviour, as we will see below. To compare larger portfolios we would need to increase the sample size tremendously to ensure enough extreme values for the optimization algorithm to converge. Other ideas would

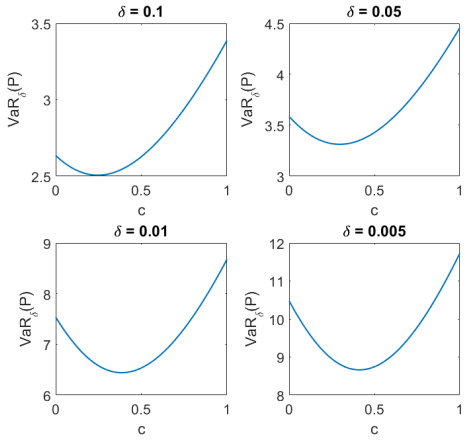

FIGURE 5.4: Portfolio loss function f described in optimization problem (5.4) for different values of δ, assuming the first example set of parameters.

be to include techniques like importance sampling or kernel smoothing, but we decided limit ourselves to two and three assets per portfolio, since they demonstrate the major characteristics of our model. Figure 5.4 shows the map $c \mapsto f(c, 1-c)$ for several levels. The shape of the target function closely resembles a parabola so we will therefure use a sequential quadratic programming algorithm (SQP) combined with a Newton algorithm to solve the nested optimization problem (5.4). We compare this analytical result with the empirical optimal solution and the first-order optimal solution.

Figure 5.5 compares the optimal portfolio weights. By assumption the idiosyncratic factors are equally-weighted, so the optimal case for the first-order model is the equally-weighted portfolio for all δ. The empirically optimal solution which minimizes $\mathrm{VaR}_\delta(P)$, is to invest almost solely into the second asset, which gradually changes into an equally-weighted portfolio as $\delta \to 0$. This is the optimal solution, since the common risk factor has more exposure to the first asset, and thus we expect higher losses with it. Since the idiosyncratic risk factors are still heavier-tailed than the systematic risk factors, their behaviour dominates more and more as $\delta \to 0$, so we asymptotically get the same behaviour as in an independent portfolio and observe the gradual change into the independent limit portfolio. The second-order approximation which takes hidden effects into account detects this behaviour and gives an optimal solution which weights the asset less exposed to market risk slightly more and also converges to independence. We also observe this behaviour in Figure 5.4, where the minimum of the loss function f moves more and more to $c = 0.5$ as $\delta \to 0$, which is the optimal solution for the independent model. An acceptable close approximation of the true optimal portfolio weights seems to be achievable for $\delta < 0.03$.

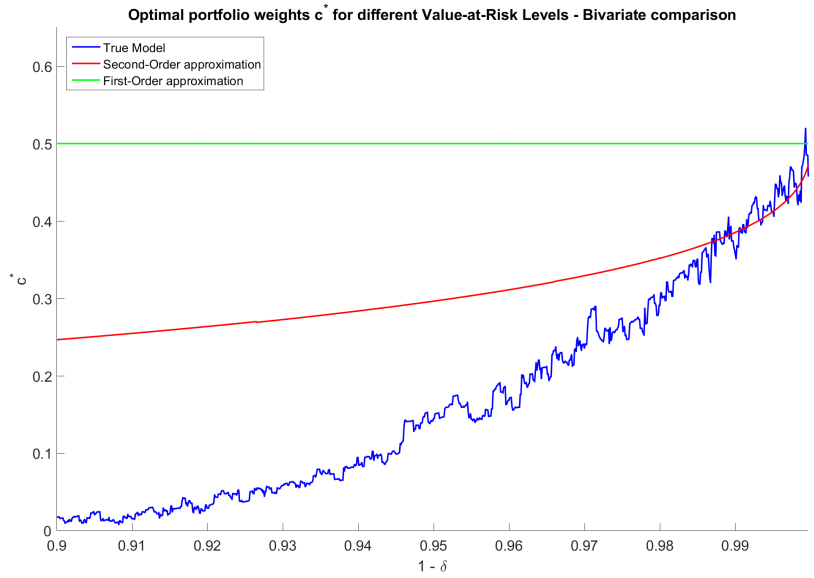

FIGURE 5.5: Optimal portfolio weights for the first example given in Table 5.1. The portfolio consists of two assets with a single common risk factor and we assume that the first asset has larger exposure to it.

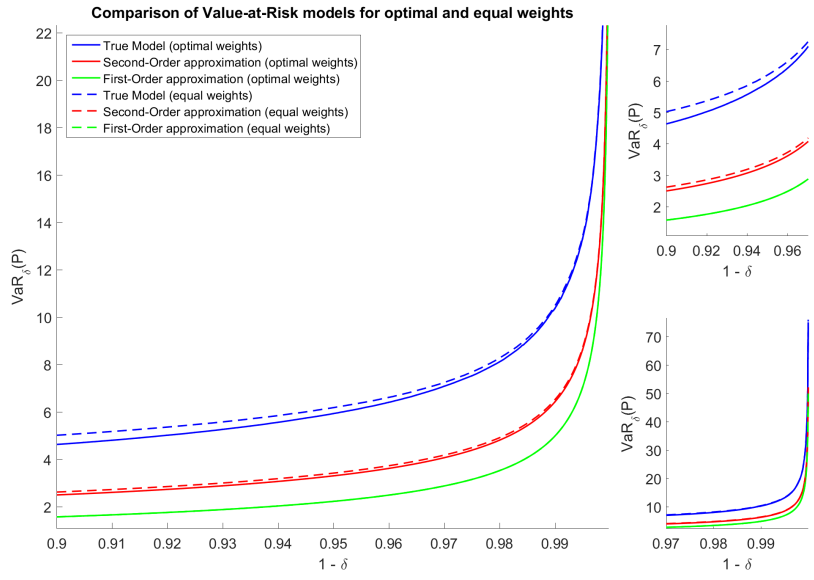

FIGURE 5.6: Value-at-Risk comparison for the first example. Compared are equally-weighted portfolio and VaR-optimal portfolio. The graphs show the approximation quality and the actual versus the predicted diversification potential over the equally weighted portfolio. The graphs on the right focus on the regions $\delta > 0.03$ (top) and $\delta < 0.03$ (bottom).

We now consider how the choice of optimal portfolio weights influences the analytical and the empirical Value-at-Risk for the portfolio, demonstrated in Figure 5.6. We see slight improvements in both the true model and the second-order approximation, since they actually have improvement potential. The first-order approximation is determined by the idiosyncratic factors alone and does not consider any systemic risk, so we don't observe any potential improvement. We observe that choosing optimal portfolio weights does not improve the direct approximation quality of the Value-at-Risk significantly and we therefore conclude that the mixture model does not seem to be a very good surrogate in this particular application.

However, we already observed in Figure 5.5 that the portfolio weights can be approximated adequately, so to see how close the actual optimal portfolio performance can be achieved with the calculated weights from the second-order approximation, we can recalculate the realized Value-at-Risk with first- and second-order optimal weights and compare how close they are to the empirical minimum. For a comparison, we also show how bad the worst actual portfolio performance is predicted by the second-order model by minimizing $-f$ in optimization problem (5.4). This portfolio consists of investing everything into a single asset which is exposed to systemic risk. This solution can also be confirmed by looking at Figure 5.4, where the maximum of f is, as discussed, at $c = 1$, so allocating everything into the first asset, which has the largest exposure to the systemic risk factor.

Considering Figure 5.7 we observe close to no difference between the empirically optimal and the second-order optimal portfolios for $\delta < 0.03$. Comparing the diversification effect over the equally-weighted portfolio through the second-order effect, we get a 7.41% diversification effect over the equally-weighted portfolio for $\delta = 0.1$, 5.70% for $\delta = 0.05$ and 2.57% for $\delta = 0.01$. We also compare how bad the loss can become in the portfolio, predicted by weights implied by the second-order approximation. Here we definitely see a strong improvement from worst-case to best-case scenario. We now turn to the question how the tail index influences the Value-at-Risk and implicitly calculate optimal portfolio weights to replicate the risk-minimizing portfolio. Beginning with the second simulation study in Table 5.1, we consider a bivariate model which has a common component that is just slightly lighter-tailed than the idiosyncratic risks. Although this is asymptotically independent, we expect the common effects to affect the Value-at-Risk and the choice of portfolio weights more than before.

Figure 5.8 shows that the heavier systemic component improves the

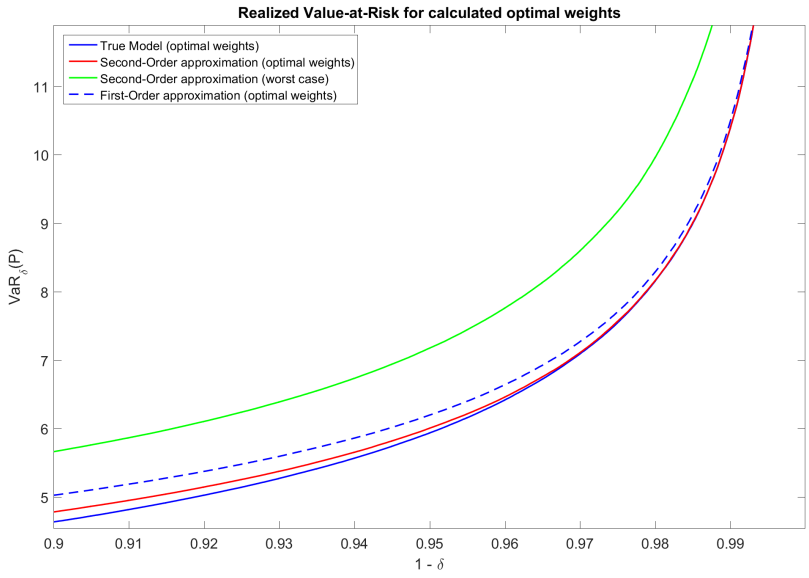

FIGURE 5.7: Realized Value-at-Risk comparison for the first example. Compared are equally-weighted portfolio against best- and worst-case portfolio when considering the Value-at-Risk as performance measure.

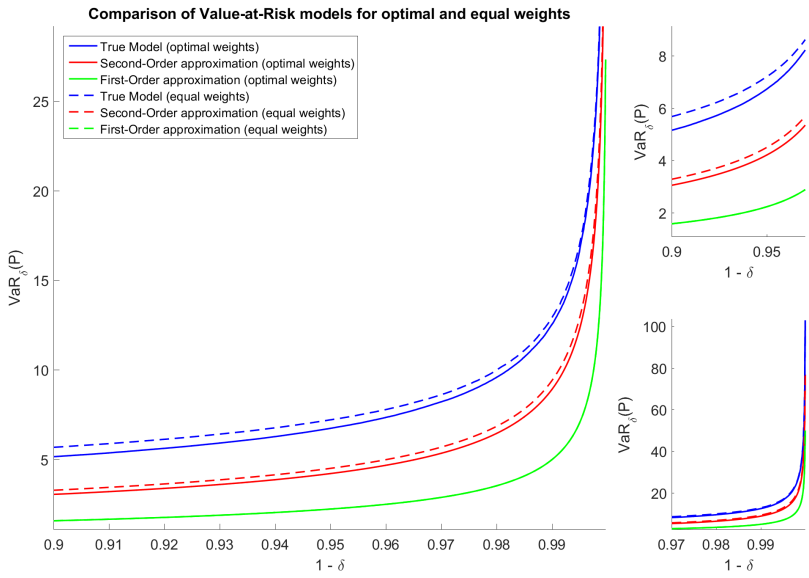

FIGURE 5.8: Value-at-Risk comparison for the second simulation example with heavier-tailed systemic component.

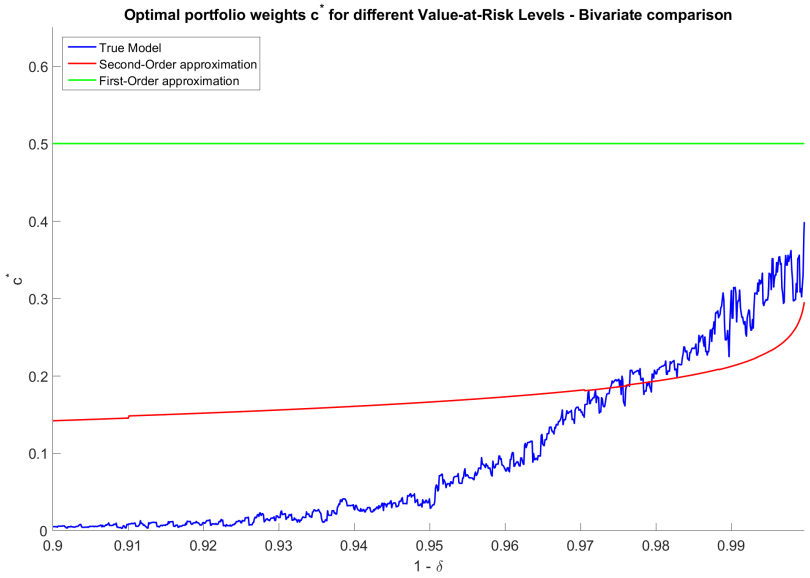

FIGURE 5.9: Optimal portfolio weights for the second example.

approximation quality of the surrogate model towards the real model given by portfolio weights shown in Figure 5.9. We can explain the improvement in approximation quality through the heavy-tailed systemic component: $\mathbb{E}_1$ contains more extreme values which concentrate on a spectral component described by H_1 at last, so the mixture surrogate is a better approximation in extreme tail regions. This is especially apparent by replicating a portfolio with the second-order optimal portfolio weights, as we can observe in Figure 5.10; the Value-at-Risk of the portfolio with weights from the analytical second-order model approximates the empirical solution very well in this case.

We turn now to the opposite case, the third example described in Table 5.1, a lighter-tailed hidden component close to independence. Since it is closer to the independent case, we observe weaker effects and less prevalent differences from the first-order optimal portfolio. This is demonstrated in in Figures 5.11 and 5.12. The portfolio weights converge less quickly to the independent limit case and we find a smaller improvement in the approximation of the true Value-at-Risk. Figure 5.12 shows that the analytical approximation of the Value-at-Risk calculated with equation (4.9) is a much worse approximation to the empirical solution than in the other two examples. The reasons are the same as in the second example; from a regular variation point of view, this example is very close to independence, so the mixture model is a bad approximation on the inner cone. Implying second-order optimal portfolio weights still gives a slight improvement over the equally-weighted portfolio, as we see in Figure 5.13, the effect however

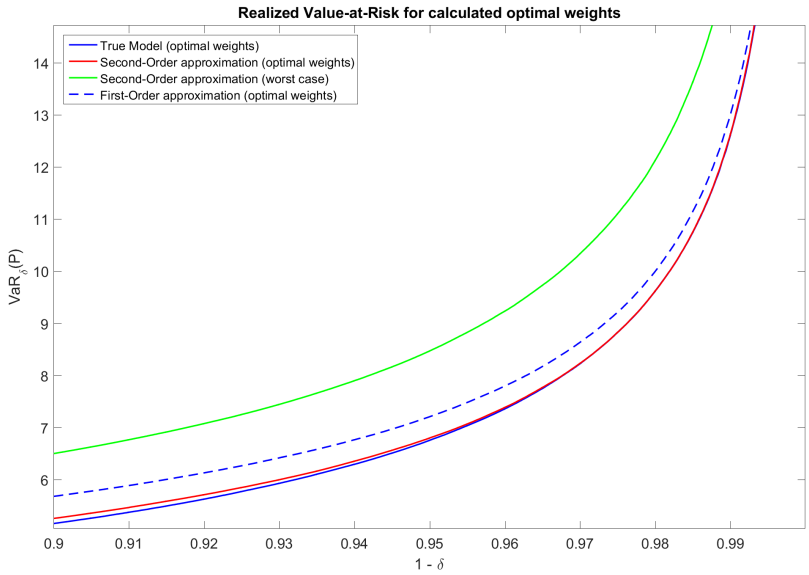

FIGURE 5.10: Realized Value-at-Risk for the second example with heavier-tailed systemic component.

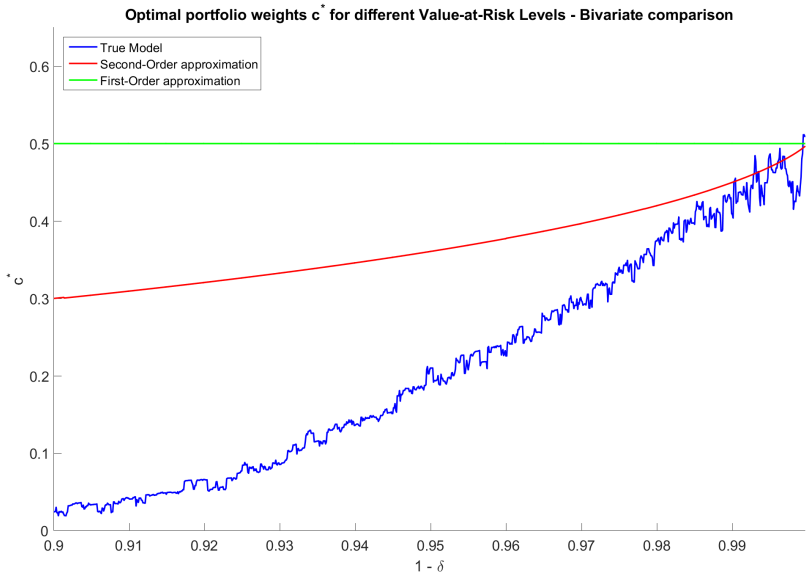

FIGURE 5.11: Optimal portfolio weights for the third example.

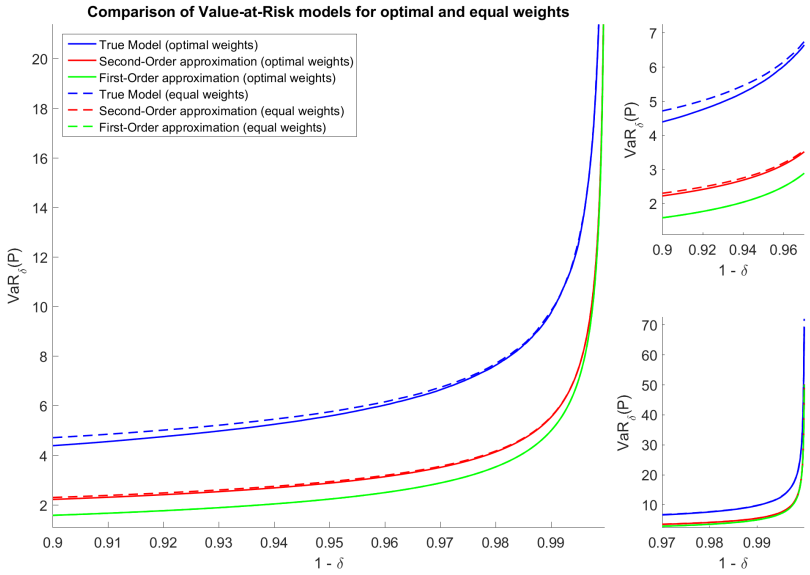

FIGURE 5.12: Results for the third simulation study with same linear coefficients, but lighter-tailed systemic component.

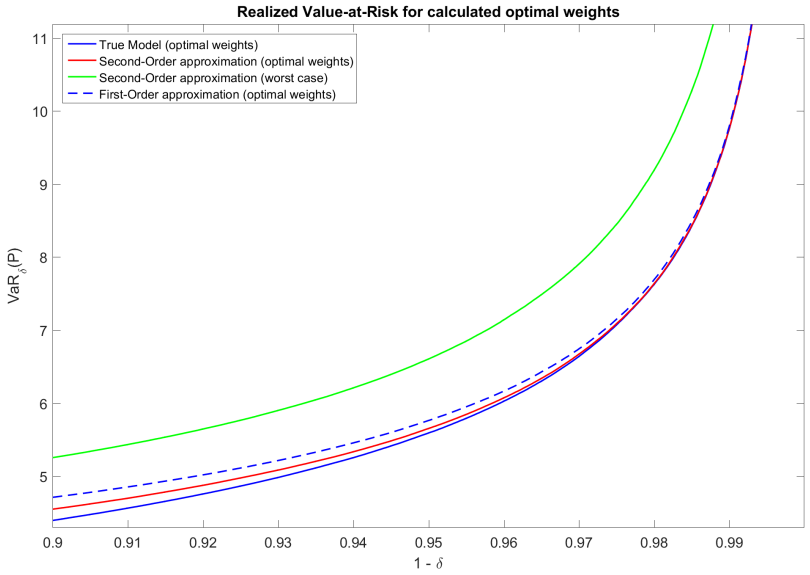

FIGURE 5.13: Realized Value-at-Risk comparison for the light-tailed systemic component.

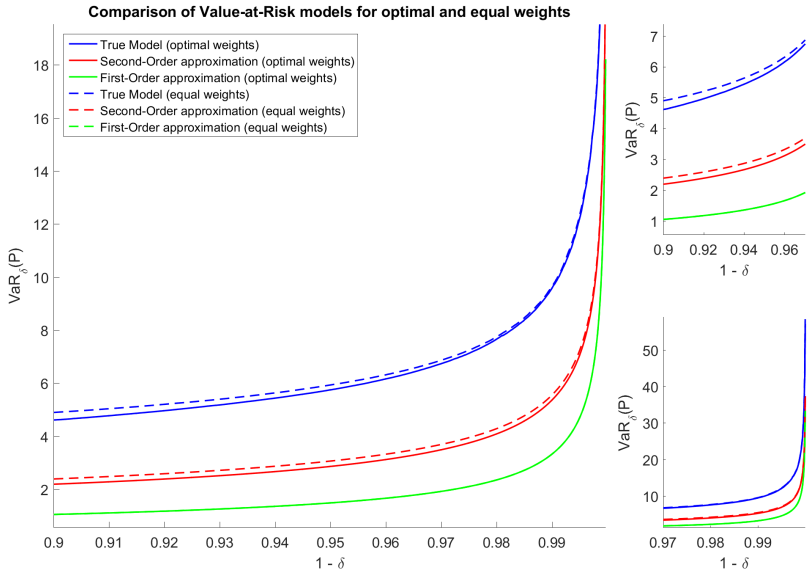

FIGURE 5.14: Value-at-Risk comparison for equally chosen portfolio and VaR-optimal portfolio assuming a three-asset portfolio with parameters given in the fourth example of Table 5.1.

is not too strong anymore, and closer to the first-order optimal portfolio.

Three assets

We will now test how the model performs with three assets. In Figure 5.14 we see basically the same effects as we did in the bivariate portfolio. Both first- and second-order models underestimate the true portfolio risk and as the idiosyncratic risk factors become more and more dominant, the first-order model is a better direct approximation to the true Value-at-Risk. Nevertheless, analogously to the two-asset case, we observe that the result of implying optimal portfolio weights from optimization problem (5.4) is close enough to the true model for $\delta < 0.03$ to form a sufficiently good approximation of the portfolio weights; see Figure 5.15. The optimal solution for the classical first-order approximation is, as before, equal weights, since we still assume equally-scaled idiosyncratic risks. We observe in Figure 5.15 that the approximation quality of the portfolio weights is as good as in the two-asset examples. Figure 5.16 shows the realized Value-at-Risk of the portfolio with first- and second-order optimal portfolio weights compared to the empirical minimum and the worst-case scenario. We see that we almost replicate the empirically optimal portfolio with the implied second-order optimal weights; when compared with Figure 5.7, we achieve slightly lower Value-at-Risk values for all three optimal portfolios (first-order, second-order and empirically). This underlines that we consider tail indices, under which diversification is beneficial, so adding assets to the portfolio and thus diversifying helps to

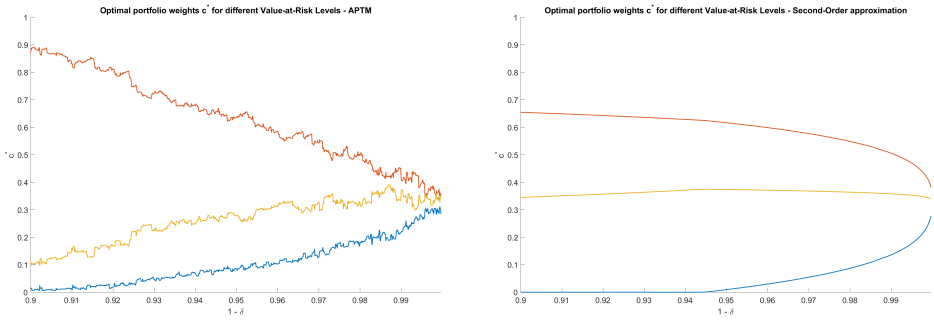

FIGURE 5.15: Optimal portfolio weights for a three-asset portfolio. The considered parameter are given by the fourth example in Table 5.1. The left graph shows the empirically optimal portfolio weights and the right graph the second-order optimal portfolio weights.

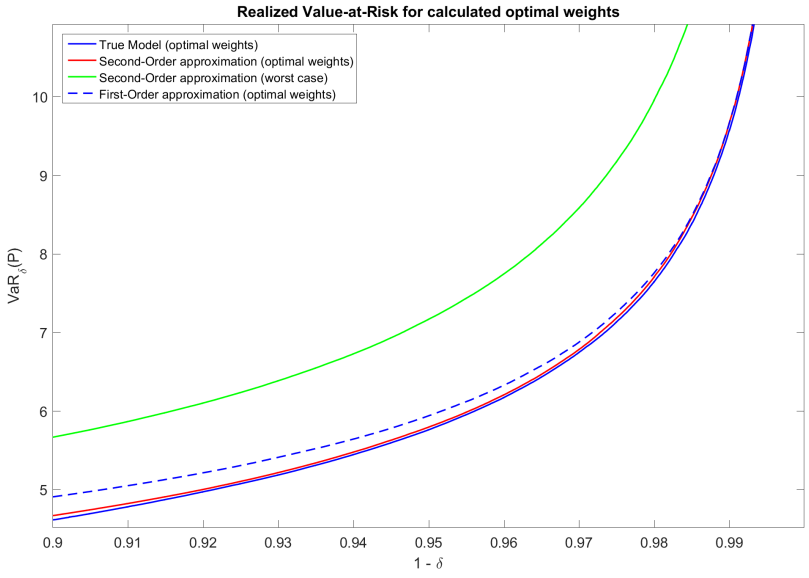

FIGURE 5.16: Value-at-Risk comparison for equally chosen portfolio and VaR-optimal portfolio for the fourth example.

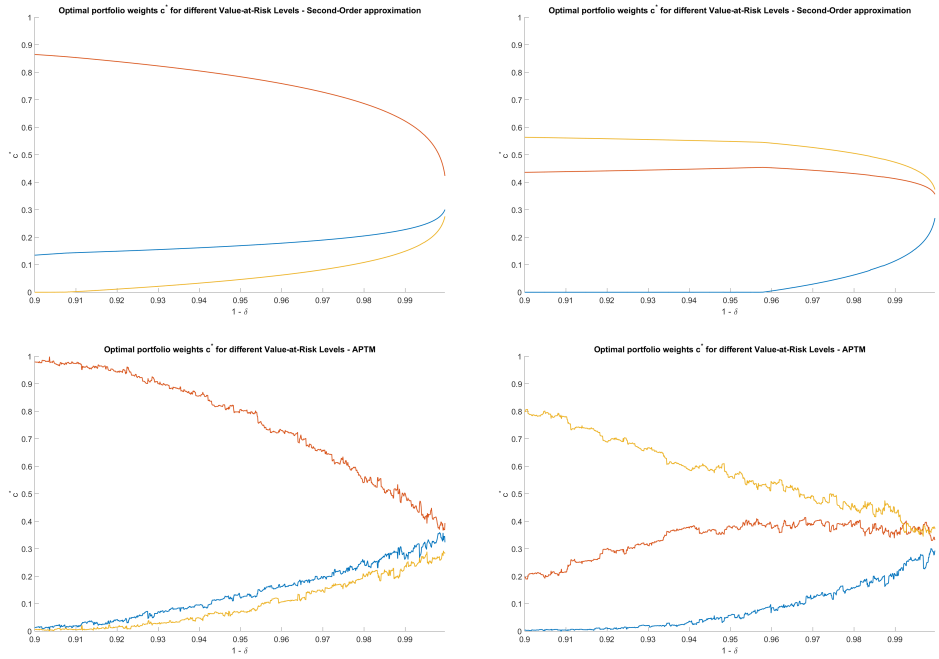

FIGURE 5.17: Portfolio weights of the fifth (left) and sixth (right) example. The top graphs are the second-order optimal portfolio weights and the bottom graphs sow the empirically optimal portfolio weights.

achieve overall lower portfolio risk.

Since the implied portfolio weight calculation seems to deliver good results, we look into sensitivity of the portfolio weight calculation and consider changing the exposure of one of the assets to the systemic risk component. This is represented in Figure 5.17, which gives the results of the last two simulation studies. The second-order-optimal portfolio weights are very similar to the empirically optimal weights and provide a good approximation for $\delta < 0.05$. This results in a near-perfect replication of the realized Value-at-Risk, as demonstrated in Figure 5.18. The left plot of Figure 5.18, which gives results from the last Example from Table 5.1, shows that the equal-weighted portfolio is worse than in the other examples. This is due to the fact that we have two assets which are disproportionally strong exposed to systemic risk with $\beta_2 = 1.3$ and $\beta_3 = 1.4$, resulting in a portfolio with a low proportion in these two more risky assets and a high proportion of the third asset. The empirically optimal portfolio is thus far away from the equal-weighted portfolio. This serves as an example that there can be quite large differences between the first- and second-order optimal portfolio, which can result in a substantially lower risk. The right plot of Figure 5.18 shows the opposite situation with two assets having low exposure to the systemic risk factor and one asset having larger exposure to it. In the corresponding optimal portfolio weights given in the right plots of Figure 5.17 we see that

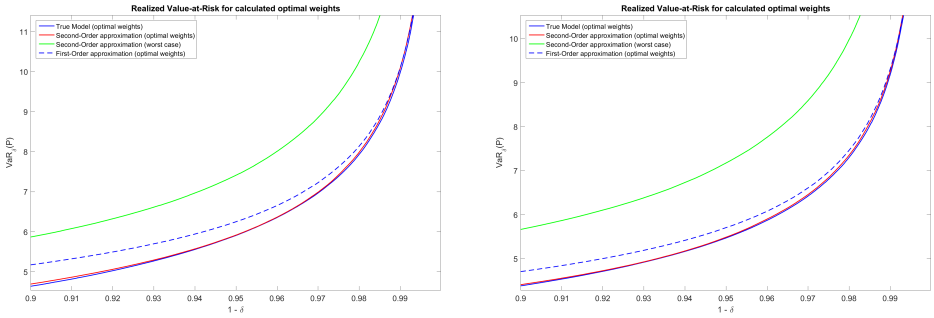

FIGURE 5.18: Sensitivity of realized Value-at-Risk to a change in risk factor exposure simulated via the fifth and sixth example.

it is optimal to invest in just a small proportion of the asset most exposed to the risk and to compose the rest of the portfolio out of similar propotions of the other two assets. The empirically optimal portfolio here is closer to an equally weighted portfolio, we therefore get a smaller diversification effect over the equally weighted portfolio.

This concludes the survey on the single-factor model.

5.1.2 Linear multi-factor model

We will now turn to multi-factor models and test how portfolios following a model with multiple linear factors behave. We therefore consider

$$X_i = Y_i + \sum_{j=1}^{n} \beta_{ij} R_j, \quad i = 1, \dots, d, \ n \geqslant 2.$$

As before, we assume asymptotic independence, meaning that $\alpha_{id} < \alpha_c$ for all systemic components and that there exists at least one $\beta_{ij} > 0$. Before we begin our analysis, we need to derive a way to represent the angular component as a scaled version of a permissible spectral measure, which is not trivial, since we do not much restrict the linear coefficients. We will mainly focus on simulation results for a two-factor model, but will briefly discuss the procedure for a general $n \geqslant 1$. It can easily be verified that a linear systemic risk represented by a hidden regular variation component on $\mathbb{E}_1$ has a spectral measure on $\mathbb{E}_1$ of the form

$$H_1(\boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\Theta_1 = \omega_i}(\omega_i), \tag{5.5}$$

for $\operatorname{supp}(\Theta_1) \subset \mathcal{N}$ bounded away from the axes. Because Dirac measures don't interact with each other, we can regard H_1 as a mixture of single Dirac measures, which results in (5.5) as a spectral measure for n linear independent components. Since we can calculate the limit measure on $\mathbb{E}_1$, we can find a tail-equivalent mixture with Theorem 10 and define the multivariate surrogate model for the hidden regular varying part on $\mathbb{E}_1$ as

$$Z_i = \sum_{j=i}^{n} \beta_{ij} R_j \mathbb{I}_{B_2 = j}, \quad i = 1, \dots, d, \tag{5.6}$$

with an equally distributed categorical random variable B_2 independent of all other random variables in the expression. Trivially, the spectral distribution of $\mathbf{Z} = (Z_1, \dots, Z_d)$ is of the form (5.5), which we seek to be a scaled version of an admissible measure H_1. To apply Theorem 11, we need to find out how to scale the factor model to such an admissible spectral measure H_1, meaning that we need to find $(m_i)_i$ and the support of Θ_1. Define

$$U_i = \frac{Z_i}{\sum_{k=1}^{d} Z_k}, \quad i = 1, \dots, d.$$

This random variable is integrable and we can define

$$m_i = E\left(U_i\right) = E\left(\frac{Z_i}{\sum_{k=1}^{d} Z_k}\right) = \frac{1}{n} \sum_{j=1}^{n} \frac{\beta_{ij}}{\sum_{k=1}^{d} \beta_{kj}}, \quad i = 1, \dots, d, \tag{5.7}$$

where the last equation used the fact that we defined Z_i as a mixture. We will now scale Z_i to points on the unit simplex with m_i and show that the components we get fulfil the moment condition in Theorem 9. Through rescaling and norming Z_i to the unit simplex, we can thus define

$$\Theta_{1_i} = \frac{Z_i/m_i}{\sum_{k=1}^{d} Z_k/m_k}, \quad i = 1, \ldots, d. \tag{5.8}$$

We will show that $\Theta_1 = (\Theta_{1_1}, \ldots, \Theta_{1_d})$ fulfils the moment condition. Consider

$$\begin{aligned}
E\left(\Theta_{1_i}\right) &= E\left(\frac{Z_i/m_i}{\sum_{k=1}^{d} Z_k/m_k}\right) \\
&= E\left(\frac{Z_i/(m_i \sum_{k=1}^{d} Z_k)}{\sum_{k=1}^{d} Z_k/(m_k \sum_{k=1}^{d} Z_k)}\right) \\
&= E\left(\frac{\frac{1}{m_i} Z_i/\sum_{k=1}^{d} Z_k}{\sum_{k=1}^{d} \frac{1}{m_k}(Z_k/\sum_{k=1}^{d} Z_k)}\right) \\
&= E\left(\frac{1}{m_i} U_i\right) \frac{1}{\sum_{k=1}^{d} \frac{1}{m_k}} \\
&= \frac{1}{\sum_{k=1}^{d} \frac{1}{m_k}} \\
&= \frac{1}{d}.
\end{aligned}$$

The last equality comes from the observation that $\frac{1}{\sum_{k=1}^{d} \frac{1}{m_k}}$ is independent of i and by construction

$$\sum_{i=1}^{d} \Theta_{1_i} = 1, \ \Theta_{1_i} \geqslant 0 \implies \sum_{i=1}^{d} E(\Theta_{1_i}) = 1.$$

Thus scaling by m_i leads to an admissible spectral measure H_1. We notice at this point that we did not need to assume that Z_i stems from a mixture assumption here, so the calculation that $E\left(\Theta_{1_i}\right) = \frac{1}{d}$ with Θ_{1_i} defined as in equation 5.8 is also valid for more complicated models.

We design a study with two linear factors and investigate their performance. Since we have two linear factors, we have more degrees of freedom to cover compared to the single-factor case. Many effects turned out to be comparable to what we already saw in the single-factor case, so we focus on the key points. We begin with the first model, which is slightly shifted towards the first asset, so we expect a weaker exposure towards it in the Value-at-Risk-minimizing portfolio. We will limit ourselves to the portfolio weights here, since the effects on the Value-at-Risk are quite similar to the single-factor case. The model behaves as expected: Figure 5.19 shows that

Simulation parameters				
Number of assets	2	2	2	2
α_{id}	2	2	2	2
α_c	3	3	2.25	3.75
β_{11}	1.3	1.1	1.3	1.3
β_{12}	1	1.5	1	1
β_{21}	0.8	1	0.8	0.8
β_{22}	1.1	1.1	1.1	1.1

TABLE 5.2: Parameter for study of the linear two-factor model

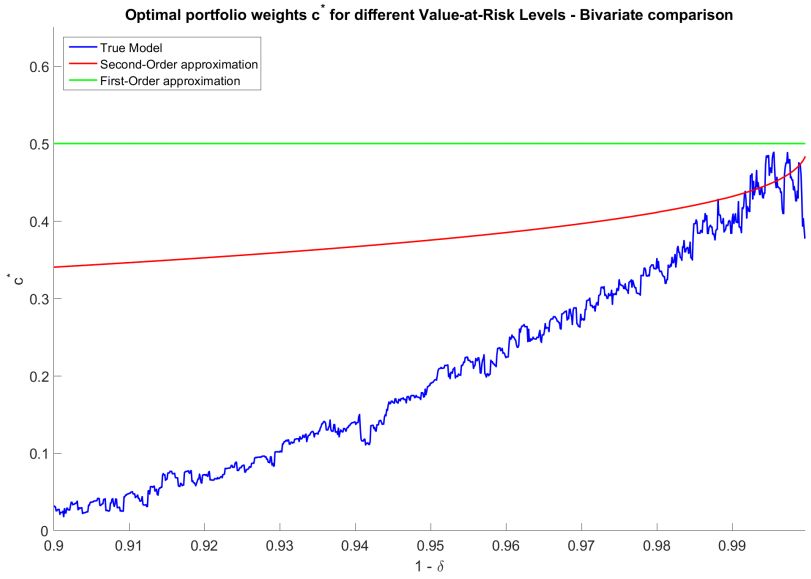

FIGURE 5.19: Results for the second simulation study with unit first-order coefficients given in Table 5.2 and two lighter-tailed systemic factors

it is better to allocate the portfolio more towards the second asset which is less exposed to the systemic factors. Two effects which are already slightly visible here, but will be more dominant in the following examples, are that the two-factor model tends to be more "stiff", so the map between the Value-at-Risk level and optimal portfolio weights tends to be more nearly linear for $\delta > 0.01$, and the quality of the second-order approximation gets slightly worse compared to the single-factor model. To see how shifting the factors influences the results, we consider the second example, in which we simulate an even larger exposure of both systemic factors towards the first component. Figure 5.20 shows an example where one of the assets has a strong exposure to one of the factors, while both assets are about equally exposed to the second factor. We see the same behaviour as in the first example shown in Figure 5.19, but slightly larger differences between the optimal second-order portfolio weights and the optimal empirical portfolio weights. As a rule of

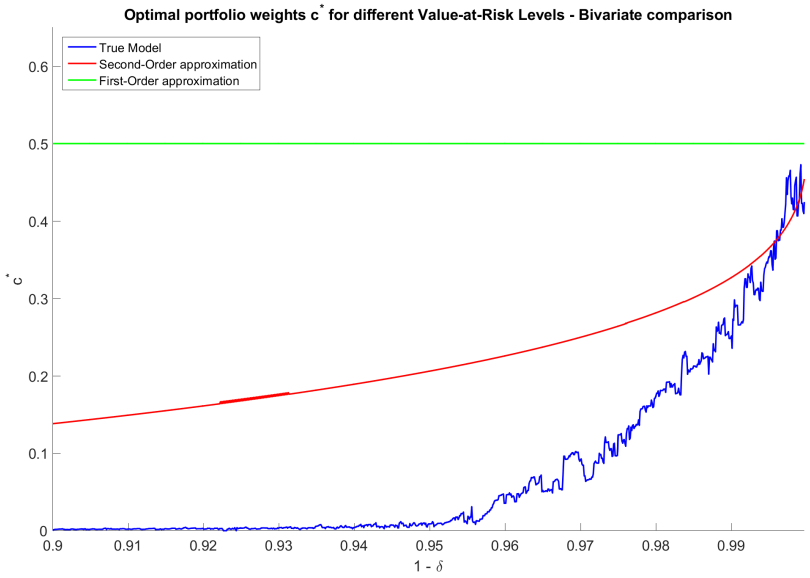

FIGURE 5.20: A two factor model with one neutral factor and one factor, to which the first asset is strongly exposed to.

thumb we can say that we get satisfying portfolio weights for $\delta < 0.02$, for the two-factor model while we see fairly good replication for $\delta < 0.03$.

Turning to the third example, with the same linear coefficients as in the first example and a smaller tail index of the systemic component, we see in Figure 5.21 more clearly the already observed effect of "stiffer" optimal portfolio weights, which tend to be nearly constant for $\delta > 0.05$. This example resembles a slightly better fit compared to the first example, giving acceptable portfolio weights for $\delta < 0.03$, and behaves similarly like the single-factor example with heavier-tailed systemic component.
We move now to the light-tailed, almost independent case, described in the fourth study. We observe again in Figure 5.22 that the quality of the result drops significantly if the systemic component is light-tailed. Comparing this with the single factor case, we get an even weaker result, getting an optimal portfolio which is almost equally-weighted, although the empirical optimal weight gets as small as $c^*_{emp} = 0.05$ for $\delta = 0.1$.

Portfolios with three portfolios behaved similarly. We saw the same effects already observed in the single-factor case; good approximation of the portfolio weights and a larger potential for diversification. We will therefore not present these here.

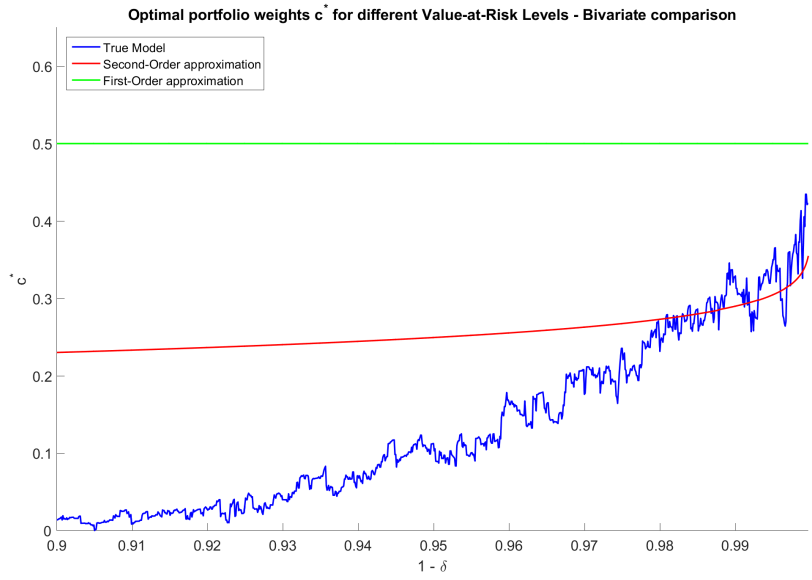

FIGURE 5.21: Results for the second simulation study with same linear coefficients as in the first example, but heavier-tailed systemic component

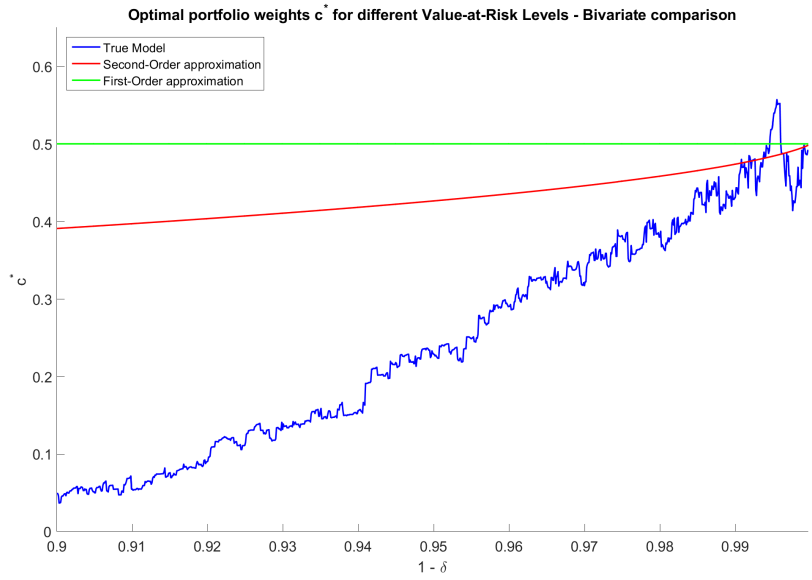

FIGURE 5.22: Results for the second simulation study with same linear coefficients as in the first example, but lighter-tailed systemic component

5.2 Discussion of the results

In Section 5.1 we successfully demonstrated the different characteristics of the mixture solution. The analytical second-order model seems to be a good approximation if used implicitly to calculate optimal portfolio weights for $\delta < 0.03$ in the single-factor model and $\delta < 0.02$ in the two-factor model. Interestingly, the single-factor model seemed to provide more pronounced portfolio weights, perhaps due to multiple factors being more "diluted" than one single factor, which models exposure to a clearer and more focussed systemic component. While we observe that the analytical second-order solution for the Value-at-Risk approximates the true Value-at-Risk for the model, we can see that the quality of the direct approximation is not very good in most cases we have tested, so for an explicit application, the use of a mixture surrogate model as basis for an analytical solution does not provide really good results in general, unless the modeller has reason to believe that there are hidden systemic components with a tail coefficient close to the tail coefficient of the idiosyncratic risks. The differences are visible when comparing Figures 5.6, 5.8 and 5.12. However we note that in the heavy-tailed example in Figure 5.8 the approximation is somehow acceptable and we conclude that the surrogate model is just suitable for a direct Value-at-Risk calculation if we have a systemic risk which has a tail index not much larger than the tail index of the idiosyncratic risk factors.

The rather bad approximation quality when using a mixture surrogate has been addressed by several authors and there have been several alternative proposals. One recent proposal is the sum-model of Cooley and Weller in [49], who show that we don't need to assume a mixture but that for a model with hidden regular variation we can find two independent additive components with support on $\mathbb{E}$ and $\mathbb{E}_1$, such that we have tail equivalence on both cones. Das and Resnick analysed the model further in [9] and gave sufficient conditions under which we have a finite spectral measure. Cooley and Weller show that their model can be used to define a subset of the subcone with hidden regular variation, such that the hidden measure is finite. This procedure was known, but an analytical solution would be much harder to calculate and would most likely be much less intuitive since it would mostly consist of quite complicated terms which have to be solved numerically, therefore we decided against it in this thesis. It might be an important step to take the sum-model and develop an analytical expression based on this surrogate to get a better approximation to the true result.

Besides this, using the analytical expression (4.9) to calculate actual portfolio weights which replicate a risk-minimizing portfolio seems to work

well overall and we can see why developing analytical expressions incorporating higher order effects is necessary for an effective risk estimation and portfolio optimization. We assumed that the idiosyncratic risk factors are equally weighted to ensure the comparability between the results. It is of course also possible to consider other combinations, but we limit ourselves to this case to demonstrate the impact of the hidden systemic factors, which can, as we see, still be quite large. We note by inspecting Figure 5.7 that the first-order optimal approach, ignoring the systemic risk and allocating the assets by just considering the (equal) scales of the idiosyncratic risk factors, already gives a fairly good result. This is both due to the fact that the proposed coefficients provide a quite centered systemic risk factor, so the optimal bivariate portfolio is already fairly close to equal weights, and simply a consequence of our assumption of asymptotical independence. So for $\delta \to 0$ the dependence structure converges to an independent market for which we can use the first-order approximation. We see for example in Figures 5.7 or 5.10 that for high thresholds the first-order optimal portfolio is a fairly good approximation to the minimal-loss portfolio, so if in doubt about the systemic effects in practice, the modeller can fall back to the first-order optimal portfolio when considering tail risk. We could show that for the first example there was an improvement of between about 2% and 7%, under other circumstances, for example for the third example, where we have the situation that we have a quite strong systemic factor, we could calculate improvements of up to 14% for a bivariate model. In practice it has to be decided, based on the individual market conditions, if it is worth considering this step. Another example of this is demonstrated by the fifth example in the single-factor model with portfolio weights given in Figure 5.17. When comparing with the realized losses for different portfolios (see Figure 5.18) we see that the equally-weighted portfolio is a worse choice for $\delta \in [0.04, 0.1]$ and that we find a larger difference in possible losses between the worst-case and the best-case portfolios. This also demonstrates that, under the assumption of $\alpha > 1$, adding more assets to the portfolio leads to a greater potential decrease of the possible losses.

We made another assumption that needs to be discussed from a practical point of view. To use Theorem 11, we need to assume that all marginal distributions have the same tail index, or that at least we concentrate on a subcone on which we have the same marginal tail indeces. This needs to be taken with a grain of salt. Consider for example three assets X_1, X_2, X_2 with tail indices $\alpha_1 = 2, \alpha_2 = 2, \alpha_3 = 2.2$. From extreme value theory we know that for large values the risk must come from X_1 and X_2. If we now seek to improve the estimate given by a portfolio with X_1 and X_2 via hidden regular variation and we find a systemic component $\mathbf{V}$ hidden under X_1 and

X_2 with tail index 3, then it would actually be more suitable for diversification to incorporate X_3 into the portfolio, since it has a larger effect in the tails and therefore a stronger impact on the portfolio Value-at-Risk. This phenomenon, known as non-standard regular variation, has also been mentioned by various authors, we refer here to [43] and [9]. While there exist transformations via individual marginal scaling (see [43]) to transform first-order regular varying marginals to a common tail index, we are not aware wether the effect of individual rescaling on hidden regular variation has been addressed so far, and how marginal rescaling affects systemic hidden components.

All these effects show potential for future research; there are many questions to be addressed. Overall we are very satisfied with the results and conclude that this approach can be very fruitful in application.

Chapter 6

Going further

During the course of Chapter 4 and 5, we developed an analytical model which acts as a surrogate for quantile problems in heavy-tailed portfolios. This was based on assumptions that might not hold in reality or can be improved. In this chapter we will briefly discuss how these assumptions can be improved and what has been done in the literature. There has been progress on the following assumptions that we made:

1. we search for hidden effects on cones as defined in Equation (3.29);
2. we assume asymptotic independence;
3. we assume a linear factor model;
4. our risk measure of interest is the Value-at-Risk.

We will give a short summary to what has been done on these topics and what are current topics of interest.

The first question one can raise, is: are the cones defined in (3.29) the only sub-cones of $\mathbb{E}$ we can find? We based the existence of the limit distribution in equation (3.5) on Theorem 8, which tells us that we can define max-stable distributions on cones shaped like $\mathbb{E}$. If we wanted more general cones in the positive orthant, we'd have to ensure that we can define regular variation on them. However, this appears to be problematic, since vague convergence introduces counter-intuitive geometrical properties. Consider the topology on $[\mathbf{0}, \infty]\backslash\{\mathbf{0}\}$ defined through the homeomorphism $\phi : [\mathbf{0}, \infty]\backslash\{\mathbf{0}\} \to [\mathbf{0}, \mathbf{1}]\backslash\{\mathbf{0}\}$ with

$$\phi(x, y) = \left(\frac{1}{1+x}, \frac{1}{1+y}\right).$$

Consider two parallel lines on $[\mathbf{0}, \infty]\backslash\{\mathbf{0}\}$ with the same positive and finite slope. These lines converge to the point (∞, ∞) so in the compactified space they are not bounded away from each other. Interestingly, this is not the case when every line is parallel to one of the base vectors. Therefore, to preserve vague convergence on a cone, we need the subcone considered to

be shaped like a half-open rectangle. There has been a very recent and promising development; central papers discussing this are [24], [8] and [31]. Hult and Lindskog reformulated the concept of regularly varying measures by only assuming them to be non-negative finite Borel measures, defined as

$$M_b(\mathbb{E}) = \{\mu : \ \mu \text{ is a non-negative finite Borel measure on } \mathbb{E}\},$$

for a cone $\mathbb{E}$ metricized by the norm $d(x, \mathbb{D}) = \inf_{y\in\mathbb{D}} d(x, y)$, where $\mathbb{D}$ is a cone such that $\mathbb{E}$ is defined as $\mathbb{E} = \mathbb{C}\backslash\mathbb{D}$ for another cone $\mathbb{C}$. They defined a complete and separable measure space on M_b, developed a theory analogously to Section 3.1.2 without the need of compactification, and showed that convergence on this measure space even implies vague convergence. A big plus for this formulation is that all measures are finite on the unit ball, although the unit ball itself does not need to be compact. The downside for inference is that the unit balls of subcones do not completely contain each other, as we assumed in this thesis. Instead, the unit ball of $\mathbb{E}$ is defined as $\{x \in \mathbb{E} : d(x, \mathbb{D}) = 1\}$. For $\mathbb{E}$ as defined in equation (3.7), the unit balls agree for both metrics, while the unit ball for $\mathbb{E}_1$ as defined in (3.8) would be a half-open rectangle bounded away from the axes with distance one. We are not aware if a sufficient analogy to Theorem 10 and Proposition 13 had been formulated, which was the main reason we decided to stick with the classical formulation via vague convergence.

When introducing a more general choice for cones one can also introduce hidden regular variation on those cones and can drop the assumption of asymptotic independence. This is a very recent development, see [10], which was published earlier this year. The authors describe in Example 3.1 a setup very similar to the simulation study in Chapter 5, which would be a potential starting point to generalize Theorem 11 without the assumption of asymptotic independence.

Linear factor models are indeed not the only possible model; in fact, Theorem 10 assumes any finite hidden spectral measure. We focussed on factor models here due to their popularity in practice. A possible continuous alternative would be to investigate the Dirichlet distribution on the unit simplex, which has the density function

$$f\left(x_1, \ldots, x_d; \alpha_1, \ldots, \alpha_d\right) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^{d} x_i^{\alpha_i - 1},$$

$$\alpha_1, \ldots, \alpha_d > 0, \quad 0 < x_i < 1, \quad \sum_{i=1}^{d} x_i = 1, \quad d \geqslant 2,$$

on the open positive $(d-1)$-dimensional simplex, where

$$\mathrm{B}(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{d} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{d} \alpha_i\right)}, \qquad \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d).$$

This distribution is very flexible and provides broad possibilities for modelling extreme value dependence. We can also assume non-linear models, which are in general more complex to handle. We decided to concentrate on linear models, since Proposition 3 ensures that the sum of two regularly varying functions is regularly varying. As stated in [25], Chapter 5.6., non-zero continuous maps preserve regular variation, so one could formulate a study for polynomial models for example.

The Value-at-Risk is of course not the only risk measure to consider, we could also think of the expected shortfall (ES_α)

$$\mathrm{ES}_\alpha = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\gamma(X) d\gamma, \quad 0 < \alpha < 1,$$

which has received more attention lately. This requires a slightly different approach, but since we know from Karamata's theorem how the integral of a regularly varying function behaves, we could reformulate Theorem 11 for ES_α. One has to consider though that we need to assume that $\alpha \geqslant 1$ for the expected shortfall to exist.

In practice, the Value-at-Risk has its flaws. Since its introduction around 1994, Value-at-Risk has been criticized by numerous academics as well as practitioners for its weaknesses as the benchmark (see [26]) for the calculation of regulatory capital in banking. Concerns about it are well-founded and can be heard in some of the recent political discussions concerning banking regulation and the financial crisis. The reason why it's important to use, and why we decided to analyse, the Value-at-Risk instead of other measures in this paper is twofold: on the one hand, the Value-at-Risk is very easily accessible for regularly varying distributions, but the most important reason is that it is the current standard risk measure in Basel II and Basel III. However, for the new framework Basel III.5, regulators try to replace the Value-at-Risk with the expected shortfall, which is expected to play a more dominant role, so it might be important for future applicability to reformulate our results in terms of the expected shortfall. A recently published discussion about the applicability and the benefits and drawbacks this new standard would bring is in [15].

Many improvements should, have to and can, be made. In the wake of new regulations which asset managers and banks are expected to fulfil, the

theory developed in this thesis opens the door to ways to better model large losses and thus to provide better and more responsible services in today's financial landscape.

Appendix A

Code

We list here the MATLAB code that was used to create the results in chapter 5. It is complete and reproduces the results accordingly.

Listing A.1: Main script for analysis in chapter 5

```
%% LaTeX2e file `./Appendix/Code/main.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%

% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Main computation script for data study
% Produces:
%  - Scatter plots for APTM and mixture surrogate
%  - Inverse cumulative distribution functions
%  - Optimal portfolio weights for different VaR Levels
%  - Comparison between ICDFs for optimal and equal weight
%
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Set parameters

clear all
close all
clc
addpath(pwd)
plotPath = [pwd,'\Figures and Results'];

numberSamples        = 2000000;
burnIn               = 0.02;
leftEnd              = 0.9;
leftEndSimulation    = leftEnd - burnIn;
stepsOptim           = 1000;
stepsOptimBurnIn     = round(stepsOptim * (1 - leftEndSimulation)/(1 -
    leftEnd));
alpha                = [2,3];
beta1                = [1; 1];
beta2                = [1.3; 0.8];
scale1               = beta1;
scale2               = beta2.^alpha(2);
numberAssets         = length(beta2);
scaleFactors         = [scale1, scale2];
app      = '2standard';
secondOrderConfigs   = 'APTMsingle';

%% Sample data
```

```
% We generate data samples from the actual
% APTM model and from a mixture representation of the
% general first order approximation (MRV on the full cone)
% and a second order approximation (MRV on the full cone
% plus HRV on the inner cone).

addpath(pwd)

tic
realData             = APTM(alpha, beta2, numberSamples, numberAssets, size
    (beta2, 2));
disp(['Time to sample APTM model: ', num2str(toc)]);
tic
secondOrderModel     = APTMmixture(alpha, beta2, numberSamples,
    numberAssets, size(beta2, 2));
disp(['Time to sample surrogate mixture model: ', num2str(toc)]);
firstOrderModel      = APTMmixture([alpha(1), Inf], beta2, numberSamples,
    numberAssets, size(beta2, 2));

%% Calculate portfolio weights
% We use the generated data samples to find optimal
% weights for a bivariate portfolio in all three used
% models and save them in matrices to be used later.

disp('Starting search for optimal portfolio weights.')
tic
val     = zeros(4*numberAssets, stepsOptimBurnIn);
[sortBeta ,index] = sort(beta2, 'descend');
eval(['empiricalTest = ',testdata,';'])

lb_rD      = zeros(numberAssets,1);
ub_rD      = ones(numberAssets,1);
lb_fO      = zeros(numberAssets,1);
ub_fO      = ones(numberAssets,1);
lb_sO      = zeros(numberAssets,1);
ub_sO      = ones(numberAssets,1);
varStartsO = 4;
varStartfO = 2;

w0_rD   = beta2(index).^alpha(2)/sum(beta2.^alpha(2));
w0_fO   = ones(numberAssets, 1)/numberAssets;
w0_sO   = (w0_fO + w0_rD)./2;

opts    = optimset('display', 'off', 'LargeScale', 'off', 'UseParallel',
    true,  'TolX', 0.0001, 'DiffMinChange',10e-10);


for i = 1:(stepsOptimBurnIn - 1)

    val(:,i) =  [
        fmincon(@(w) EMPportfolioVaR(empiricalTest, leftEndSimulation + (1
            - leftEndSimulation)/stepsOptimBurnIn*i, w), w0_rD, [], [],
            ones(1, numberAssets), 1, lb_rD, ub_rD, [], opts);
        fmincon(@(w) IMPLportfolioVaR({'secondOrder', secondOrderConfigs},
            scaleFactors , alpha, leftEndSimulation + (1 -
            leftEndSimulation)/stepsOptimBurnIn*i, w, varStartsO), w0_sO,
            [], [], ones(1, numberAssets), 1, lb_sO, ub_sO, [], opts);
```

```
90          fmincon(@(w) IMPLportfolioVaR({'firstOrder'}, scaleFactors , alpha,
                 leftEndSimulation + (1 - leftEndSimulation)/stepsOptimBurnIn*i
                , w, varStartfO), w0_fO, [], [], ones(1, numberAssets), 1,
                lb_fO, ub_fO, [], opts);
91          fmincon(@(w) - IMPLportfolioVaR({'secondOrder', secondOrderConfigs
                },scaleFactors , alpha, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, w, varStartsO), ones(
                numberAssets, 1)/numberAssets, [], [], ones(1, numberAssets),
                1, lb_sO, ub_sO, [], opts);
92          ];
93
94
95      varStartsO  = IMPLportfolioVaR({'secondOrder', 'APTMsingle'},
            scaleFactors , alpha, leftEndSimulation + (1 - leftEndSimulation)/
            stepsOptimBurnIn*i, val(numberAssets+1:2*numberAssets,i),
            varStartsO);
96      varStartsO2 = IMPLportfolioVaR({'secondOrder', 'APTMsingle'},
            scaleFactors , alpha, leftEndSimulation + (1 - leftEndSimulation)/
            stepsOptimBurnIn*i, val(3*numberAssets+1:4*numberAssets,i),
            varStartsO);
97      varStartfO  = IMPLportfolioVaR({'firstOrder', 'APTMsingle'},
            scaleFactors , alpha, leftEndSimulation + (1 - leftEndSimulation)/
            stepsOptimBurnIn*i, val(2*numberAssets+1:3*numberAssets,i),
            varStartfO);
98
99      w0_rD   = val(1:numberAssets,i);
100     w0_sO   = val(numberAssets + 1:2*numberAssets,i);
101     w0_fO   = val(2*numberAssets + 1:3*numberAssets,i);
102
103     lb_rD   = max(w0_rD - 0.05, zeros(numberAssets,1));
104     ub_rD   = min(w0_rD + 0.05, ones(numberAssets,1));
105
106     if ~mod(i + 1, 50)
107         disp(['Step ',num2str(i + 1), ' in config ',app,', estimated
                remaining time: ', ...
108             num2str(toc * (stepsOptimBurnIn - i)/50)])
109         tic
110     end
111 end
112
113
114 disp('Calculate Value at Risk with optimal portfolio weights, compare to
        equal portfolio.')
115 varVal      = NaN(stepsOptimBurnIn, 10);
116 varStartsO  = 5;
117 varStartfO  = 3;
118
119 for i = 1:(stepsOptimBurnIn - 1)
120     varVal(i,:) =   [
121         EMPportfolioVaR(empiricalTest, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, val(1:numberAssets,i)),
                ...
122         IMPLportfolioVaR({'secondOrder', secondOrderConfigs},scaleFactors ,
                 alpha, leftEndSimulation + (1 - leftEndSimulation)/
                stepsOptimBurnIn*i, val(numberAssets+1:2*numberAssets,i),
                varStartsO), ...
123         IMPLportfolioVaR({'firstOrder'},scaleFactors , alpha,
                leftEndSimulation + (1 - leftEndSimulation)/stepsOptimBurnIn*i,
                 val(2*numberAssets+1:3*numberAssets,i), varStartfO), ...
```

```
124         EMPportfolioVaR(empiricalTest, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, ones(numberAssets, 1)/
                numberAssets), ...
125         IMPLportfolioVaR({'secondOrder', secondOrderConfigs},scaleFactors ,
                 alpha, leftEndSimulation + (1 - leftEndSimulation)/
                stepsOptimBurnIn*i, ones(numberAssets, 1)/numberAssets,
                varStartsO), ...
126         IMPLportfolioVaR({'firstOrder'},scaleFactors , alpha,
                leftEndSimulation + (1 - leftEndSimulation)/stepsOptimBurnIn*i,
                 ones(numberAssets, 1)/numberAssets, varStartfO), ...
127         EMPportfolioVaR(empiricalTest, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, val(1*numberAssets+1:2*
                numberAssets,i)), ...
128         EMPportfolioVaR(empiricalTest, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, val(2*numberAssets+1:3*
                numberAssets,i)),...
129         EMPportfolioVaR(empiricalTest, leftEndSimulation + (1 -
                leftEndSimulation)/stepsOptimBurnIn*i, val(3*numberAssets+1:4*
                numberAssets,i)),...
130         IMPLportfolioVaR({'secondOrder', secondOrderConfigs},scaleFactors ,
                 alpha, leftEndSimulation + (1 - leftEndSimulation)/
                stepsOptimBurnIn*i, val(3*numberAssets+1:4*numberAssets,i),
                varStartsO), ...
131         ];
132
133     varStartsO = varVal(i,5);
134     varStartfO = varVal(i,6);
135
136 end
137
138
139 val(:,1 : stepsOptimBurnIn - stepsOptim)    = [];
140 varVal(1 : stepsOptimBurnIn - stepsOptim, :) = [];
141
142 %% Plot and save results
143 % We plot the results and save the data.
144
145 dataStruct.leftEnd              = leftEnd;
146 dataStruct.stepsOptimization    = stepsOptim;
147 dataStruct.varVal               = varVal;
148 dataStruct.val                  = val;
149 dataStruct.beta                 = beta2;
150 dataStruct.realData             = realData;
151 dataStruct.firstOrderModel      = firstOrderModel;
152 dataStruct.secondOrderModel     = secondOrderModel;
153 dataStruct.secondOrderConfigs   = secondOrderConfigs;
154 dataStruct.numberSamples        = numberSamples;
155 dataStruct.numberAssets         = numberAssets;
156 dataStruct.scaleFactors         = scaleFactors;
157 dataStruct.alpha                = alpha;
158 dataStruct.app                  = app;
159
160
161 cd(plotPath)
162 mkdir(app), cd(app)
163 createPlots(dataStruct)
164 save(['simulationData_',date,'_',num2str(numberSamples),'_',num2str(
        numberAssets),'_', app])
165 cd .., cd ..
166 close all
```

Listing A.2: APTM code

```
%% LaTeX2e file `./Appendix/Code/APTM.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%

function data = APTM(alpha, beta, sampleSize, numberAssets)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This generates samples via the heavy-tailed APTM model.
%
% INPUT:
%  alpha        - [2x1]          double extremal coefficients
%  beta         - [2x1]          double betas for APTM
%  sampleSize   - integer        size of data
% OUTPUT:
%  data         - [2xsampleSize] produced model data
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

alphaI    = alpha(1);
alphaC    = alpha(2);

individual   = 1./(rand(numberAssets, sampleSize)).^(1/alphaI);
common       = zeros(numberAssets, sampleSize);

for i = 1:numberFactors
    common  = common + repmat(1./(rand(1, sampleSize).^(1/alphaC)),
        numberAssets, 1)...
          .*repmat(beta(:,i),1,sampleSize);
end

data = individual + common;
end
```

LISTING A.3: APTM (surrogate model) code

```
%% LaTeX2e file `./Appendix/Code/APTMmixture.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%

function data = APTMmixture(alpha, beta, sampleSize, numberAssets)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This generates samples via the mixture surrogate model for
% the heavy-tailed APTM model.
%
% INPUT:
%  alpha  - [2x1] double extremal coefficients
%  beta  - [2x1] double betas for APTM
%  sampleSize - integer   size of data
% OUTPUT:
%  data  - [2xsampleSize] produced model data
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

alphaI  = alpha(1);
alphaC1  = alpha(2);

B       = repmat(round(rand(1, sampleSize)), numberAssets, 1);
B_fO    = mnrnd(1, ones(1,numberAssets)/numberAssets, sampleSize)';
B_sO    = mnrnd(1, ones(1, numberFactors(1))/numberFactors(1), sampleSize)
    ';
R       = 1./(rand(numberAssets, sampleSize).^(1/alphaI));
R0      = zeros(numberAssets, sampleSize);

if ~isinf(alphaC1)
    for i = 1:numberFactors(1)
        R0  = R0 + repmat(1./(rand(1, sampleSize).^(1/alphaC1)),
            numberAssets, 1)...
            .*repmat(beta(:,i),1,sampleSize) .* repmat(B_sO(i,:),
                numberAssets, 1);
    end
end

data    = 2^(1/alphaI) * R.*B_fO.*B + 2^(1/alphaC1)*(R0).*(1 - B);

end
```

LISTING A.4: Empirical Portfolio Value at Risk

```
%% LaTeX2e file `./Appendix/Code/EMPportfolioVaR.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%

function t = EMPportfolioVaR(data, level, weight)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This function calculates the empirical cumulative Value at
% Risk of a portfolio.
%
%  INPUT:
%  data  -  [nxm] double  returns of n assets
%  level - integer   the quantile level
%  weight - [nx1] double vector of portfolio weights
%
% OUTPUT:
%  t  - double   Value at Risk
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

portfolio = data'*weight;
t = quantile(portfolio, level);

end
```

LISTING A.5: Analytical Portfolio Value-at-Risk via mixture surrogate

```
%% LaTeX2e file `./Appendix/Code/IMPLportfolioVaR.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%
function t = IMPLportfolioVaR(config, scaleFactors, alpha, level, weight,
    initVaR)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This function calculates the cumulative Value at Risk of
% a portfolio implicitly.
%
%  INPUT:
%  config  -  [2x1] cell  model choice
%  scaleFactors -  [nx2] double  scale factors
%  alpha -   [2x1] tail indices
%  level - integer   the quantile level
%  weight - [nx1] double vector of portfolio weights
%  initVaR  -  Initial value
%
% OUTPUT:
%  t  - double   Value at Risk
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

warning('off')
fOterm = @(t) (1/t)^(alpha(1))* 1/size(scaleFactors, 1) * weight'.^alpha(1)
    *scaleFactors(:,1);

```

```
if strcmp(config{1}, 'firstOrder')

    sOterm = @(t) (1/t)^(alpha(1))* (size(scaleFactors, 1) - 1)/size(
        scaleFactors, 1) * weight'.^alpha(1)*scaleFactors(:,1);

elseif strcmp(config{1}, 'secondOrder')

    if strcmp(config{2}, 'APTMsingle')
        sOterm = @(t) (1/t)^(alpha(2)) * (weight'*(scaleFactors(:,2)).^(1/
            alpha(2)))^alpha(2);
    elseif strcmp(config{2}, 'APTMdual')
        sOterm = @(t) (1/t)^(alpha(2)) * ((weight' * scaleFactors(:,2).^(1/
            alpha(2)))^alpha(2) + (weight'* scaleFactors(:,3).^(1/alpha(2))
            )^alpha(2));
    else
        error('Wrong second order model.')
    end

else
    error('Wrong choice of order config.')
end

t = fzero(@(t) fOterm(t) + sOterm(t) - 1 + level, initVaR);

end
```

Listing A.6: Script to generate plots used in chapter 5

```
%% LaTeX2e file `./Appendix/Code/createPlots.m'
%% generated by the `filecontents' environment
%% from source `thesis' on 2016/10/11.
%%
function createPlots(dataStruct)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% This function creates all plots and saves the output to files.
%
%  INPUT:
%   struct  - [1x1] struct   contains all needed parameters
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


leftEnd             = dataStruct.leftEnd;
stepsOptimization   = dataStruct.stepsOptimization;
varVal              = dataStruct.varVal;
val                 = dataStruct.val;
beta                = dataStruct.beta;
realData            = dataStruct.realData;
firstOrderModel     = dataStruct.firstOrderModel;
secondOrderModel    = dataStruct.secondOrderModel;
numberSamples       = dataStruct.numberSamples;
numberAssets        = dataStruct.numberAssets;
app                 = dataStruct.app;

scaleVector         = leftEnd:((1 - leftEnd)/(stepsOptimization - 1)):1;

if numberAssets == 2
    figure('Position', [0 0 350 1050], 'PaperPositionMode', 'auto')
```

```
33      set(gca,'fontsize',15)
34      hold on
35      subplot(3,1,1)
36      set(gca,'fontsize',15)
37      scatter(realData(1,1:100), realData(2,1:100))
38      title({'Samples generated by the APTM','', '100 samples'})
39      xlabel('X_1')
40      ylabel('X_2')
41      limit = ceil(min(max(realData(2,1:100))))*1.15;
42      xlim([0, limit])
43      ylim([0, limit])
44      subplot(3,1,2)
45      set(gca,'fontsize',15)
46      scatter(realData(1,1:1000), realData(2,1:1000))
47      title('1.000 samples')
48      xlabel('X_1')
49      ylabel('X_2')
50      limit = ceil(min(max(realData(2,1:1000))))*1.15;
51      xlim([0, limit])
52      ylim([0, limit])
53      subplot(3,1,3)
54      set(gca,'fontsize',15)
55      scatter(realData(1,1:10000), realData(2,1:10000))
56      title('10.000 samples')
57      xlabel('X_1')
58      ylabel('X_2')
59      limit = ceil(min(max(realData(2,1:10000))))*1.15;
60      xlim([0, limit])
61      ylim([0, limit])
62      hold off
63      saveas(gcf,['APTMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
64
65
66      figure('Position', [0 0 350 1050], 'PaperPositionMode', 'auto')
67      set(gca,'fontsize',15)
68      hold on
69      subplot(3,1,1)
70      set(gca,'fontsize',15)
71      scatter(firstOrderModel(1,1:100), firstOrderModel(2,1:100))
72      title({'Samples generated by the','','First Order Approximation','', '
            100 samples'})
73      xlabel('X_1')
74      ylabel('X_2')
75      limit = ceil(min(max(firstOrderModel(2,1:100))))*1.15;
76      xlim([0, limit])
77      ylim([0, limit])
78      subplot(3,1,2)
79      set(gca,'fontsize',15)
80      scatter(firstOrderModel(1,1:1000), firstOrderModel(2,1:1000))
81      title('1.000 samples')
82      xlabel('X_1')
83      ylabel('X_2')
84      limit = ceil(min(max(firstOrderModel(2,1:1000))))*1.15;
85      xlim([0, limit])
86      ylim([0, limit])
87      subplot(3,1,3)
88      set(gca,'fontsize',15)
89      scatter(firstOrderModel(1,1:10000), firstOrderModel(2,1:10000))
90      title('10.000 samples')
```

```
91      xlabel('X_1')
92      ylabel('X_2')
93      limit = ceil(min(max(firstOrderModel(2,1:10000))))*1.15;
94      xlim([0, limit])
95      ylim([0, limit])
96      hold off
97      saveas(gcf,['FOMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
98
99
100     figure('Position', [0 0 350 1050], 'PaperPositionMode', 'auto')
101     set(gca,'fontsize',15)
102     hold on
103     subplot(3,1,1)
104     set(gca,'fontsize',15)
105     scatter(secondOrderModel(1,1:100), secondOrderModel(2,1:100))
106     title({'Samples generated by the','','Second Order Approximation','', '
            100 samples'})
107     xlabel('X_1')
108     ylabel('X_2')
109     limit = ceil(min(max(secondOrderModel(2,1:100))))*1.15;
110     xlim([0, limit])
111     ylim([0, limit])
112     subplot(3,1,2)
113     set(gca,'fontsize',15)
114     scatter(secondOrderModel(1,1:1000), secondOrderModel(2,1:1000))
115     title('1.000 samples')
116     xlabel('X_1')
117     ylabel('X_2')
118     limit = ceil(min(max(secondOrderModel(2,1:1000))))*1.15;
119     xlim([0, limit])
120     ylim([0, limit])
121     subplot(3,1,3)
122     set(gca,'fontsize',15)
123     scatter(secondOrderModel(1,1:10000), secondOrderModel(2,1:10000))
124     title('10.000 samples')
125     xlabel('X_1')
126     ylabel('X_2')
127     limit = ceil(min(max(secondOrderModel(2,1:10000))))*1.15;
128     xlim([0, limit])
129     ylim([0, limit])
130     hold off
131     saveas(gcf,['SOMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
132
133 elseif numberAssets == 3
134
135     figure('Position', [0 0 900 900], 'PaperPositionMode', 'auto')
136     set(gca,'fontsize',15)
137     scatter3(realData(1,1:10000), realData(2,1:10000), realData(3,1:10000))
            ; view(105,20)
138     limit = ceil(min(min(max(realData(2,1:100)))))*1.15;
139     xlim([0, limit])
140     ylim([0, limit])
141     zlim([0, limit])
142     title({'Samples generated by the APTM','', '10.000 samples'})
143     xlabel('X_1')
144     ylabel('X_2')
145     zlabel('X_3')
```

```
146     saveas(gcf,['APTMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
147
148
149     figure('Position', [0 0 900 900], 'PaperPositionMode', 'auto')
150     set(gca,'fontsize',15)
151     scatter3(firstOrderModel(1,1:10000), firstOrderModel(2,1:10000),
            firstOrderModel(3,1:10000)); view(105,20)
152     limit = ceil(min(min(max(firstOrderModel(2,1:100)))))*1.15;
153     xlim([0, limit])
154     ylim([0, limit])
155     zlim([0, limit])
156     title({'Samples generated by the','','First Order Approximation','','
            10.000 samples'})
157     xlabel('X_1')
158     ylabel('X_2')
159     zlabel('X_3')
160     saveas(gcf,['SOMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
161
162
163     figure('Position', [0 0 900 900], 'PaperPositionMode', 'auto')
164     set(gca,'fontsize',15)
165     scatter3(secondOrderModel(1,1:10000), secondOrderModel(2,1:10000),
            secondOrderModel(3,1:10000)); view(105,20)
166     limit = ceil(min(min(max(secondOrderModel(2,1:100)))))*1.15;
167     xlim([0, limit])
168     ylim([0, limit])
169     zlim([0, limit])
170     title({'Samples generated by the','','Second Order Approximation','', '
            10.000 samples'})
171     xlabel('X_1')
172     ylabel('X_2')
173     zlabel('X_3')
174     saveas(gcf,['FOMdata_',date,'_',num2str(numberSamples),'_',num2str(
            numberAssets),'_',app,'.png'])
175
176 end
177
178 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
179 set(gca,'fontsize',15)
180 hold on
181 plot(scaleVector, varVal(:,4), 'Color','blue')
182 plot(scaleVector, varVal(:,5), 'Color','red')
183 plot(scaleVector, varVal(:,6), 'Color','green')
184 xlim([leftEnd 0.99965])
185 ylim([0,max(varVal(end-1, :)) + 5])
186 title({['Comparison of different approximations of the inverse cumulative'
        ...
187     ' distribution function'], ['Equally weighted portfolio with' ...
188     ' \beta_1 = ',num2str(beta(1)),' and \beta_2 = ',num2str(beta(2))]})
189 legend('True Model', 'Second-Order approximation', 'First-Order
        approximation')
190 xlabel('\delta')
191 ylabel('VaR_\delta(P)')
192 hold off
193 saveas(gcf,['comparisonICDF_',date,'_',num2str(numberSamples),'_',num2str(
        numberAssets),'_',app,'.png'])
194
195
```

```
196 if numberAssets == 2
197     figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
198     set(gca,'fontsize',15)
199     hold on
200     plot(scaleVector, val(1,:), 'Color','blue')
201     plot(scaleVector, val(3,:), 'Color','red')
202     plot(scaleVector, val(5,:), 'Color','green')
203     xlim([leftEnd 0.99965])
204     ylim([0 1])
205     title('Optimal portfolio weights c^* for different Value-at-Risk Levels
            - Bivariate comparison')
206     legend('True Model', 'Second-Order approximation', 'First-Order
          approximation')
207     ylabel('c^*')
208     xlabel('\delta')
209     hold off
210     saveas(gcf,['comparisonPortfolioWeightsBiva_',date,'_',num2str(
          numberSamples),'_',num2str(numberAssets),'_',app,'.png'])
211 end
212
213
214 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
215 set(gca,'fontsize',15)
216 hold on
217 for i = 1:numberAssets
218     plot(scaleVector, val(i,:))
219 end
220 xlim([leftEnd 0.99965])
221 ylim([0 1])
222 title('Optimal portfolio weights c^* for different Value-at-Risk Levels -
      APTM')
223 ylabel('c^*')
224 xlabel('\delta')
225 hold off
226 saveas(gcf,['comparisonPortfolioWeightsAPTM_',date,'_',num2str(
      numberSamples),'_',num2str(numberAssets),'_',app,'.png'])
227
228
229 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
230 set(gca,'fontsize',15)
231 hold on
232 for i = (numberAssets + 1):2*numberAssets
233     plot(scaleVector, val(i,:))
234 end
235 xlim([leftEnd 0.99965])
236 ylim([0 1])
237 title('Optimal portfolio weights c^* for different Value-at-Risk Levels -
      Second-Order approximation')
238 ylabel('c^*')
239 xlabel('\delta')
240 hold off
241 saveas(gcf,['comparisonPortfolioWeightsSO_',date,'_',num2str(numberSamples)
      ,'_',num2str(numberAssets),'_',app,'.png'])
242
243
244 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
245 set(gca,'fontsize',15)
246 hold on
247 for i = (2*numberAssets + 1):3*numberAssets
248     plot(scaleVector, val(i,:))
```

```
249 end
250 xlim([leftEnd 0.99965])
251 ylim([0 1])
252 title('Optimal portfolio weights c^* for different Value-at-Risk Levels -
        First-Order approximation')
253 ylabel('c^*')
254 xlabel('\delta')
255 hold off
256 saveas(gcf,['comparisonPortfolioWeightsFO_',date,'_',num2str(numberSamples)
        ,'_',num2str(numberAssets),'_',app,'.png'])
257
258
259 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
260 set(gca,'fontsize',15)
261 subplot(2,4,[1:3 5:7 ]);
262 set(gca,'fontsize',15)
263 hold on
264 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),1), 'Color','blue')
265 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),2), 'Color','red')
266 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),3), 'Color','green')
267 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),4),'--', 'Color','blue')
268 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),5),'--', 'Color','red')
269 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),6),'--', 'Color','green')
270 xlim([leftEnd 0.99965])
271 ylim([min(varVal(1,[1:6])) - 0.5 ,max(varVal(end-15, [1:6])) + 0.5])
272 title('Comparison of Value-at-Risk models for optimal and equal weights')
273 legend('True Model (optimal weights)', 'Second-Order approximation (optimal
        weights)',...
274     'First-Order approximation (optimal weights)', 'True Model (equal
            weights)',...
275     'Second-Order approximation (equal weights)', 'First-Order
            approximation (equal weights)')
276 xlabel('\delta')
277 ylabel('VaR_\delta(P)')
278 hold off
279 subplot(2,4,4);
280 set(gca,'fontsize',15)
281 hold on
282 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),1), 'Color','blue')
283 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),2), 'Color','red')
284 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),3), 'Color','green')
285 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),4),'--', 'Color','blue')
286 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),5),'--', 'Color','red')
287 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),6),'--', 'Color','green')
288 xlim([leftEnd 0.97])
289 ylim([min(varVal(1,[1:6])) - 0.5 ,max(varVal(find(scaleVector > 0.97, 1),
        [1:6])) + 0.5 ])
290 xlabel('\delta')
291 ylabel('VaR_\delta(P)')
292 subplot(2,4,8);
293 set(gca,'fontsize',15)
294 hold on
295 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),1), 'Color','blue')
296 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),2), 'Color','red')
297 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),3), 'Color','green')
298 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),4),'--', 'Color','blue')
299 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),5),'--', 'Color','red')
300 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),6),'--', 'Color','green')
301 xlim([0.97 scaleVector(end - 1)])
```

```
302 ylim([min(varVal(find(scaleVector > 0.97, 1), [1:6])) - 0.5 ,max(varVal(end
        - 1, [1:6])) + 0.5 ])
303 xlabel('\delta')
304 ylabel('VaR_\delta(P)')
305 saveas(gcf,['comparisonVaRPortfolios_',date,'_',num2str(numberSamples),'_',
        num2str(numberAssets),'_',app,'.png'])
306
307
308 figure('Position', [0 0 1100 700], 'PaperPositionMode', 'auto')
309 set(gca,'fontsize',15)
310 subplot(2,4,[1:3 5:7 ]);
311 set(gca,'fontsize',15)
312 semilogy(scaleVector(1:(end - 1)), varVal(1:(end - 1),1), 'Color','blue')
313 hold on
314 semilogy(scaleVector(1:(end - 1)), varVal(1:(end - 1),7), 'Color','red')
315 semilogy(scaleVector(1:(end - 1)), varVal(1:(end - 1),8), 'Color','green')
316 semilogy(scaleVector(1:(end - 1)), varVal(1:(end - 1),4),'--', 'Color','
        blue')
317 xlim([leftEnd scaleVector(end - 1)])
318 ylim([min(varVal(1,[1 4 7 8])) - 0.1 ,max(varVal(end-200, [1 4 7 8])) +
        0.5])
319 title('Realized Value-at-Risk for calculated optimal weights')
320 legend('True Model (optimal weights)', 'Second-Order approximation (optimal
         weights)',...
321     'First-Order approximation (optimal weights)', 'Equally weighted
            portfolio')
322 xlabel('\delta')
323 ylabel('VaR_\delta(P)')
324 hold off
325 subplot(2,4,4);
326 set(gca,'fontsize',15)
327 hold on
328 plot(scaleVector(1:(end - 1)), (varVal(1:(end - 1),1)), 'Color','blue')
329 plot(scaleVector(1:(end - 1)), (varVal(1:(end - 1),7)), 'Color','red')
330 plot(scaleVector(1:(end - 1)), (varVal(1:(end - 1),8)), 'Color','green')
331 plot(scaleVector(1:(end - 1)), (varVal(1:(end - 1),4)),'--', 'Color','blue'
        )
332 xlim([leftEnd 0.97])
333 ylim([min(varVal(1,[1 4 7 8])) - 0.5 ,max(varVal(find(scaleVector > 0.97,
        1), [1 4 7 8])) + 0.5 ])
334 xlabel('\delta')
335 ylabel('VaR_\delta(P)')
336 subplot(2,4,8);
337 set(gca,'fontsize',15)
338 hold on
339 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),1), 'Color','blue')
340 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),7), 'Color','red')
341 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),8), 'Color','green')
342 plot(scaleVector(1:(end - 1)), varVal(1:(end - 1),4),'--', 'Color','blue')
343 xlim([0.97 scaleVector(end - 1)])
344 ylim([min(varVal(find(scaleVector > 0.97, 1), [1 4 7 8])) - 0.5 ,max(varVal
        (end - 1, [1 4 7 8])) + 0.5 ])
345 xlabel('\delta')
346 ylabel('VaR_\delta(P)')
347
348 end
```

Bibliography

[1] August A. Balkema and Sidney I. Resnick. Max-infinite divisibility. *Journal of Applied Probability*, 14:309–319, 1977.

[2] Heinz Bauer. *Measure and integration theory*, volume 26. Walter de Gruyter, 2001.

[3] Patrick Billingsley. *Convergence of probability measures*. Wiley, 2013.

[4] Nicholas H Bingham, Charles M Goldie, and Jef L Teugels. *Regular variation*, volume 27. Cambridge University Press, 1989.

[5] Leo Breiman. Probability. *Classics in Applied Mathematics*, 7, 1992.

[6] Kai Lai Chung. *A course in probability theory*. Academic Press, 2001.

[7] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer, 2007.

[8] Bikramjit Das, Abhimanyu Mitra, and Sidney Resnick. Living on the multidimensional edge: seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(01):139–163, 2013.

[9] Bikramjit Das and Sidney I. Resnick. Models with hidden regular variation: Generation and detection. *Stochastic Systems*, 5(2):195–238, 2015.

[10] Bikramjit Das and Sidney I Resnick. Hidden regular variation under full and strong asymptotic dependence. *arXiv preprint arXiv:1602.01262*, 2016.

[11] Laurens de Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer, 2007.

[12] Laurens de Haan and Sidney I Resnick. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337, 1977.

[13] Denis Denisov, Serguei Foss, and Dmitry Korshunov. On lower limits and equivalences for distribution tails of randomly stopped sums. *Bernoulli*, 14(2):391–404, 2008.

[14] Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[15] Paul Embrechts, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. An academic response to basel 3.5. *Risks*, 2(1):25–48, 2014.

[16] Eugene F Fama and Merton H Miller. *The theory of finance*, volume 3. Dryden Press Hinsdale, IL, 1972.

[17] Herbert Federer. *Geometric measure theory*. Springer, 2014.

[18] Raya Feldman and Murad Taqqu. *A practical guide to heavy tails: statistical techniques and applications*. Springer, 1998.

[19] Gerald B Folland. *Real analysis: modern techniques and their applications*. Wiley, 2013.

[20] Sergey Foss, Dmitry Korshunov, and Stan Zachary. *An introduction to heavy-tailed and subexponential distributions*. Springer, 2011.

[21] Jakob Levinus Geluk and LFM de Haan. Regular variation, extensions and tauberian theorems. *CWI tracts*, 40:1–131, 1987.

[22] Christian Gourieroux, Jean-Paul Laurent, and Olivier Scaillet. Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3):225–245, 2000.

[23] Janet E Heffernan and Jonathan A Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546, 2004.

[24] Henrik Hult and Filip Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math.(Beograd)(NS)*, 80(94):121–140, 2006.

[25] Anders Hedegaard Jessen and Thomas Mikosch. Regularly varying functions. *Publications de l'Institut Mathématique. Nouvelle Série*, 80(94):171–192, 2006.

[26] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. MacGraw-Hill international editions: Finance series. McGraw-Hill, 2001.

[27] Achim Klenke. *Probability theory: a comprehensive course*. Springer, 2013.

[28] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer, 1983.

[29] Anthony W Ledford and Jonathan A Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

[30] Anthony W Ledford and Jonathan A Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499, 1997.

[31] Filip Lindskog, Sidney I Resnick, Joyjit Roy, et al. Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probability Surveys*, 11:270–314, 2014.

[32] Georg Mainik and Paul Embrechts. Diversification in heavy-tailed portfolios: properties and pitfalls. *Annals of Actuarial Science*, 7(01):26–45, 2013.

[33] Georg Mainik, Georgi Mitov, and Ludger Rüschendorf. Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *Journal of Empirical Finance*, 32:115–134, 2015.

[34] Benoit B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[35] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[36] Krishanu Maulik and Sidney Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2004.

[37] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press, 2015.

[38] Thomas Mikosch. *Regular variation, subexponentiality and their applications in probability theory*. Eurandom, 1999.

[39] Abhimanyu Mitra and Sidney I Resnick. Hidden regular variation: Detection and estimation. *arXiv preprint arXiv:1001.5058*, 2010.

[40] Ser-Huang Poon, Michael Rockinger, and Jonathan A. Tawn. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of financial studies*, 17(2):581–610, 2004.

[41] Sidney I Resnick. Tail equivalence and its applications. *Journal of Applied Probability*, 8(1):136–156, 1971.

[42] Sidney I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336, 2002.

[43] Sidney I Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer, 2007.

[44] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.

[45] Laurent Schwartz. *Radon measures on arbitrary topological spaces and cylindrical measures*. Oxford University Press, 1973.

[46] Masaaki Sibuya. Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210, 1959.

[47] Charles Swartz. *Measure, integration and function spaces*. World Scientific, 1994.

[48] Michael Eugene Taylor. *Measure theory and integration*. American Mathematical Soc., 2006.

[49] Grant B Weller and Daniel Cooley. A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika*, 101(1):17–36, 2014.

[50] Chen Zhou. Dependence structure of risk factors and diversification effects. *Insurance: Mathematics and Economics*, 46(3):531–540, 2010.